Characterising transcriptomic and immune heterogeneity in aggressive invasive lobular breast cancers

Dr Sarah Nash

This thesis is submitted to the University of London/The Institute of Cancer Research for the Degree of Doctor of Philosophy

September 2023

Breast Cancer Now Research Centre

The Institute of Cancer Research

237 Fulham Road

London

Declaration

This thesis was completed under the supervision of Dr. Rachael Natrajan. The work described here was carried out at the Breast Cancer Now Research Centre, Institute of Cancer Research, 237 Fulham Road, London SW3 6JB.

I, Dr Sarah Nash, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, or where work has been performed with the assistance of others, I confirm that this has been indicated in the thesis.

Date: 12.09.2023

Signed: SNash

Publications during my time in research

Khalique S, **Nash S***, Mansfield D, Wampfler J, Attygale A, Vroobel K, Kemp H, Buus R, Cottom H, Roxanis I, Jones T, von Loga K, Begum D, Guppy N, Ramagiri P, Fenwick K, Matthews N, Hubank MJF, Lord CJ, Haider S, Melcher A, Banerjee S, Natrajan R. Quantitative Assessment and Prognostic Associations of the Immune Landscape in Ovarian Clear Cell Carcinoma. Cancers (Basel). 2021 Jul 30;13(15):3854. doi: 10.3390/cancers13153854. PMID: 34359755; PMCID: PMC8345766.

Khalique S, **Nash S***, Natrajan R. Definitive study shows no association between ARID1A mutation status and clinical outcome in endometriosis-related ovarian cancers‡. J Pathol. 2022 Sep;258(1):1-3. doi: 10.1002/path.5973. Epub 2022 Jun 29. PMID: 35647895; PMCID: PMC9540905.

Bland P, Saville H, Wai P, Curnow L, Muirhead G, Nieminuszczy J, Ravindran N, Beatrix John M, Hedayat S, Barker H, Wright J, Yu L, Mavrommatis I, Read A, Peck B, Allen M, Gazinska P, Pemberton HN, Gulati A, **Nash S**, Noor F, Guppy N, Roxanis I, Pratt G, Oldreive C, Stankovic T, Barlow S, Kalirai H, Coupland SE, Broderick R, Alsafadi S, Houy A, Stern MH, Pettit S, Choudhary JS, Haider S, Niedzwiedz W, Lord CJ & Natrajan R. *SF3B1* hotspot mutations confer sensitivity to PARP inhibition by eliciting a defective replication stress response. *Nat Genet* **55**, 1311–1323 (2023). https://doi.org/10.1038/s41588-023-01460-5

*joint first author

Abstract

Invasive Lobular Carcinoma (ILC) is the second most common histological subtype of breast cancer (BC) accounting for 10-15% of all BCs. It is a unique disease entity with distinct histological appearances, molecular alterations and clinicopathologic features. It also has a unique tumour immune microenvironment. A subgroup of ILC patients have clinically aggressive disease with metastases occurring early (< 3 years) after primary diagnosis. These patients have limited treatment options. There is thus a need to better understand the molecular basis and transcriptional drivers of aggressive ILC, as well of the immune landscape to help identify potential new drug targets to improve patient outcomes.

The project firstly investigated molecular drivers of clinically aggressive ILC at the genomic and transcriptomic level. This identified higher rates of *TP53*, *FAT1* and *HER2* alterations and an association between FGFR1 alterations and relapse in the aggressive pleomorphic ILC histological subtype. Using RNA sequencing of pleomorphic ILC (n = 47), survival analysis and the use of a random forest model enabled the generation of prognostic risk scores which further validated as predictors of overall survival in independent validation cohorts.

Characterisation of the immune landscape at the histological level through the quantification of stromal TILs in 163 ILCs showed that the majority of ILCs are characterised by a low level of stromal TILs but pleomorphic cases contain higher levels, although stromal TILs were not associated with clinical outcome in both pleomorphic and non-pleomorphic ILC. Further immune cell characterisation using NanoString Digital Spatial Profiling (DSP) technology in a cohort of 20 pleomorphic ILCs identified an association between high levels of CD68+ cells (macrophages) and early relapse. Validation using dual IHC identified the M2-like/M1-like macrophage ratio is associated with poor clinical outcome. Co-culture experiments further showed enhanced growth of ILC cell-line tumour cells when grown with M2-like over M1-like macrophages. NanoString Whole Transcriptome (WTA) analysis of 10 pleomorphic ILCs revealed differences between ‘immune-hot’ and ‘immune-cold’ tumour cells and cancer-associated fibroblasts (CAFs) and identified higher expression of *HOXB13* in ‘immune-cold’ tumour cells which was associated with poor outcome in independent ILC cohorts. A further experiment studying matched primary and brain metastases from an in vivo ILC model identified enrichment of pre-existing minor subclones that were enriched in brain metastatic lesions defined by high levels of the EMT associated gene *GCNT2*.

Overall, the project revealed new insights into the transcriptomic and immune heterogeneity in clinically aggressive ILCs and identified new patient stratification biomarkers for further validation.

Statement of COVID impact

The first year of my thesis was interrupted by the impact of COVID-19. The ICR labs were shut for six months, and a shift pattern of work came into effect following this for several months resulting in limited access to the lab. We also had to observe social distancing rules. This had a significant impact on my ability to develop new skills and knowledge specifically in the lab at the start of my PhD as one-to-one teaching of new techniques was limited by strict social distancing and limited staff members were present in the lab at any one time to receive teaching and supervision from. In addition certain critical experiments and projects had to be prioritised during this time for publication purposes. All of these factors combined meant that I had limited lab experience during the start of my PhD to start experiments and learn practical skills.

Date; 12.09.2023
Signed (student): SLWash
Signed (supervisor):……..……

Acknowledgements

Firstly I would like to thank my primary supervisor Dr. Rachael Natrajan firstly for the opportunity to complete my PhD in her lab and secondly for her endless support, guidance and stimulating discussions which I have enjoyed throughout my research. I am extremely grateful for the time she has dedicated to guiding me through the PhD and for her insights and advice which have contributed to my learning and development during my time as a clinical fellow. Equally I would like to thank my clinical supervisor Professor Elinor Sawyer for her encouragement and support throughout the PhD and for always providing a clinical perspective.

I am grateful to the members of the Functional Genomics Lab team past and present who have made my PhD a pleasant and enjoyable experience and offered practical guidance and assistance throughout, including Harry Saville, Phil Bland, Ioanna Mavrommatis, Kate Moore, Shefali Thakur, Somaieh Hedayat and Flora Johnson. I am also grateful to Renee Flaherty and Flavia Hughes in our neighbouring lab for their assistance in tissue culture and their friendship. I must also thank Professor Elinor Sawyer's team at Kings, in particular Vandna Shah for teaching me practical lab-based skills nearer the start of the PhD and for her continued support and advice. I am also grateful to Maria Roman-Escorza for helping me with genomic analysis, Jasmine Timbres and Anca Mera for helping me work with clinical data and Ahmed Ahmed for his practical help in the lab. In addition, I'm grateful for the hard work of James Rosekilly from the KHP tissue bank for preparing large numbers of tissue sections for various elements of the project.

I am grateful to the Institute of Cancer Research and Breast Cancer Now for providing the equipment and the core facilities which have been invaluable throughout the PhD. I am extremely thankful for the help and assistance that I have received from Syed Haider and Shaun Tan from the BCR bioinformatics team and for their endless hard work and patience in navigating me through the complexities of bioinformatic analysis in the various areas of the project. Furthermore I cannot thank Dr Ioannis Roxanis from the BCN Histopathology facility enough, for the hours of time he has dedicated to reviewing histological images with me and for the informal teaching he has provided throughout which has helped me to develop as a histopathology trainee. I have thoroughly enjoyed our thought-provoking discussions over the past few years. Equally I wish to thank Chris Starling, who has kindly dedicated time to helping me get to grips with digital pathology approaches which I have enjoyed learning throughout the project.

Finally I am extremely thankful for the support of all my family and friends that I have received throughout the best and hardest parts of the PhD. In particular I want to thank my wonderfully kind and encouraging parents Jayne and Tony Nash, my lovely siblings Laura, Tom and Lizzie and my best friend and fiancé Jeremy, for their endless support and love for which I will be forever grateful

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Invasive Lobular Carcinoma: Overview

Invasive lobular Carcinoma (ILC) is the second most common histological subtype of breast cancer accounting for 10 - 15% of cases [1-3]. It differs from invasive breast cancer of no special type (IC-NST), the most common subtype, in its epidemiology, molecular alterations, clinicopathologic features and responsiveness to therapy [1, 4-6]. Since the 1980s the incidence of ILC has increased relative to IC-NST, attributed to an increased use of hormone-replacement therapy (HRT) and increased consumption of alcohol [7-30]. Despite the fact that ILC represents a unique disease entity, with clear biological and clinical differences from IC-NST, its clinical management has long been the same as that of IC-NST [31]. ILCs tend to be oestrogen (ER) and progesterone (PR) positive (ER/PR+), HER2 negative (HER2-) and have a low Ki67 index which are features of the luminal A intrinsic subtype [1, 5, 32]. A minority of cases are classed as luminal B, showing an increased Ki-67 index, and rarely HER2-enriched or triple-negative/basal like. Whilst a luminal A intrinsic subtype typically predicts a favourable prognosis, evidence suggests that a subset of patients with ILC have poorer responses to endocrine therapy compared to those with IC-NST and similar biomarkers [33] as well as lower pathologic complete response rates after neoadjuvant chemotherapy [34]. When overall survival (OS) is considered, it has become clear that the long-term prognosis of ILC is worse than ER+ IC-NST [4, 6] (Figure 1.1).

The molecular basis of disease relapse and drug resistance is poorly understood in ILC and biomarkers with the ability to identify early resistance, or to predict the likelihood of successful treatment at outset are yet to be identified. Whilst the majority of ILC patients with disease recurrence relapse several years after primary diagnosis, there is a subgroup of women who relapse early, within 3 years of primary diagnosis and this subgroup have limited treatment options and represent a clinically unmet need. Pleomorphic ILC is a rare histological ILC subtype which tends to follow a more clinically aggressive course and many pleomorphic ILC patients fall into this ‘early-relapse’ ILC group. Therefore, a comprehensive characterisation of aggressive early-relapsing ILCs including pleomorphic cases is warranted to better understand the molecular basis and immunological landscape of these tumours and to identify drivers of clinically aggressive disease and potential new drug targets, to ultimately improve the poor clinical outcomes for these patients.

1.2 Incidence and Risk Factors for ILC

1.2.1 Incidence

In the UK, breast cancer is the commonest cancer type and accounts for 15% of all new cancer cases. It is the commonest cancer among women, accounting for 30% of new female cancer cases [35]. Among men it is uncommon, accounting for less than 1% of new male cancer cases. The vast majority (99%) of UK breast cancer cases occur in women [35].

Approximately 55,500 new breast cancer cases occur per year in the UK and as the most common special breast cancer subtype, ILC accounts for a reported 10 - 15% of all cases [35, 36]. The incidence rates of breast cancer have increased by 5% among women in the UK over the last decade [35]. Projections predict rises in the number of new cases among women each year from approximately 61,800 cases in 2023 - 2025 to 69,900 in 2038 - 2040 [35].

The change in incidence rate of ILC was compared to IC-NST using data from the 'Studies based on Surveillance, Epidemiology, and End Result' (SEER) database and showed that the ILC rates increased from 9.6% in 1987 to 15.6% in 1999 whilst the IC-NST rate remained stable during this period [2, 3]. The increase in ILC rates was attributed to an increased use of combined HRT (oestrogen and progestin). After a subsequent decrease in the use of HRT a non-significant reduction in the incidence of ILC occurred from 2002 to 2006 [37].

From a demographic perspective, the incidence of the disease appears to be highest in white women, yet an increased mortality rate of ILC is observed in black women. A 2017 study assessed clinical outcomes in 18,295 ILC patients from the SEER database, including white (n = 15,936), black (n = 1,451), and other race (e.g. Asian/Pacific/American Indian, n = 908) patients [38]. This demonstrated that 5-year OS and breast cancer-specific survival (BCSS) were worst in black women (76.0% and 84.4%, respectively) vs white (85.5% and 87.7%, $p < 0.01$) and other race women (91.1%, and 91.6%, $p < 0.01$) [38].

1.2.2 Risk Factors

The use of HRT is an established risk factor for the development of ILC. There have been 25 observational studies (10 cohort and 15 case control) examining the associations between the use of HRT and breast cancer development according to histological subtype [7-30]. Most of these have demonstrated that HRT use is more strongly associated with ILC risk compared to IC-NST risk. These studies also examined the associations separately for users of combined HRT (CHT) and oestrogen only HRT (EHT). Among current CHT users, the relative risk (RR) of IC-NST development was mostly lower than 2.0 (overall RR 1.5, 95% CI: 1.5 - 1.6) whereas it surpassed 2.0 for ILC in most of the studies (overall RR 2.0, 95% CI: 1.9 - 2.1). For current users of EHT the risks were less marked with the relative risk of IC-NST development ranging from 0.7 - 2.0 for IC-NST (overall RR 1.1, 95% CI: 1.0 - 1.1) and from 1.0 - 2.1 for ILC (overall RR 1.4, 95% CI: 1.3 - 1.5) [39].

Additional risks factors for the development of ILC are the same as those for IC-NST. Factors which increase the exposure to endogenous ovarian oestrogens are established breast cancer risk factors, such as early menarche, late menopause, increased age at first birth and low parity [39]. Interestingly a significantly stronger association between the age at menarche and risk of breast cancer development was identified for ILC compared to IC-NST in a meta-analysis of 85 studies of individual patient data [40]. For each year younger at menarche the relative risk of breast cancer development was 1.073 for ILC vs 1.035 for IC-NST ($p = 0.0001$). Similarly, increased age of menopause was more strongly associated with ILC risk compared to IC-NST risk, with a 3.6% increase in risk of ILC development for each year older at menopause compared to 2.6% increased risk in IC-NST ($p = 0.006$) [40].

The association between the use of the oral contraceptive pill and breast cancer risk has been widely studied with initial concerns raised about its carcinogenic potential. An analysis assessing outcomes in over 150,000 women who had been included in 54 epidemiologic studies demonstrated that patients who had ever used oral contraception had a mild (7%) increase in the relative risk of breast cancer when compared to never users of oral contraceptives. Current users appeared to carry a 24% increase in risk, which did not appear to rise with increased duration of use. The risk decreased after cessation of use and there was no increased risk evident after 10 years of stopping use [41]. The UK Royal College of General Practitioners' Oral Contraception Study recruited 46,022 women in 1968 and 1969 and followed them up for 44 years [42]. Standardised rates of different types of cancer were calculated within patients who had ever used combined oral contraception vs patients who

had never used it. The data was standardised for a range of demographic and lifestyle factors e.g. smoking, age, parity. Incidence rate ratios between previous users and never users were calculated and the effects of increasing time since previous oral contraceptive use was assessed. An increased breast cancer risk was observed in current and recent users (used within 5 years) however this increased risk was lost after 5 years of stopping combined oral contraception [42].

In addition to hormonal factors, lifestyle factors are also associated with risk of breast cancer development. Alcohol is a key modifiable risk factor and breast cancer is the most common cancer caused by alcohol among women globally. The World Health Organisation (WHO) reports that alcohol consumption is responsible for 7 of 100 new breast cancers in Europe and in 2020, 40,000 new breast cancers in Europe were caused by alcohol consumption according to the International Agency for Research on Cancer (IARC) [43].

An increased risk of the development of breast cancer is also associated with being overweight or obese in postmenopausal women with an increase of 5 kg/m^2 in body mass index (BMI) being associated with an 8% increase in breast cancer risk [44]. In contrast a decreased breast cancer risk has been observed in overweight or obese premenopausal women [44].

The associations between alcohol and excessive weight and breast cancer risk have been attributed to increased circulating oestrogen levels which these factors cause. Extra-glandular secretion of oestrogen by adipose tissue occurs in postmenopausal women whereas among premenopausal women, excessive weight is associated with anovulatory cycles and a reduction in circulating oestrogen which has been associated with the apparent reduced breast cancer risk in this group [45].

1.3 Clinical Features of ILC

ILC differs from IC-NST in its clinical presentation as summarised in Table 1 [4, 6, 46-48]. Studies consistently show that ILC patients tend to present at an older age than IC-NST [47, 49-51]. A large retrospective study assessed ILC and IC-NST patients from the SEER tumour registry database. When the clinicopathological features of 30,190 ILC patients and 288,216 IC-NST patients were compared, a significant increase in age at presentation was observed in the ILC group with a median presentation age of 63 years compared to 59 years in the IC-NST group ($p < 0.0001$) [47]. Most women present with a poorly defined palpable breast mass which can be difficult to detect on physical examination due to the diffusely infiltrative nature of the tumour. For this reason, they tend to go clinically undetected for a longer period, and patients therefore present with larger overall tumour sizes and with more advanced-stage disease compared to IC-NST. SEER tumour registry data was used to study 263,408 patients diagnosed with either ILC or IC-NST. ILC patients were more likely to have tumour sizes of > 2cm (43.1% vs 32.6%; $p < 0.001$) and positive lymph nodes at diagnosis (36.8% vs 34.4%; $p < 0.001$) [52]

Multicentricity in breast cancer describes the presence of two or more foci of cancer which have formed separately within different quadrants of the same breast [46]. Bilateral breast cancer describes the presence of a second tumour in the contralateral breast. Bilateral breast cancers can be synchronous or metachronous depending on the time interval between diagnosis of the first and second lesions with synchronous tumours being diagnosed at the same time or within three months of the index tumour whereas metachronous tumours are diagnosed after 3-6 months of the index tumour [46, 47]. Studies consistently report higher rates of multicentricity, multifocality and bilaterality in ILC when compared to IC-NST [48, 49]. Bilateral involvement is reported in 20 - 29% of ILCs and in a large cohort comparing the clinical features of ILC and IC-NST, the rate of cancer involvement in the contralateral breast in ILC was almost double that observed in IC-NST [49].

Radiologically, ILCs can be challenging to diagnose due to the limited contrast difference between the tumour and the normal adjacent fibro-glandular breast parenchyma. The most common mammographic finding is a spiculated mass with architectural distortion [50]. The pattern of subtle invasion of ILC means that the tumour size often exceeds the mammographic findings. MRI is more helpful in diagnosing ILCs than traditional mammography and is therefore used in the setting of a diagnosis of ILC on core needle biopsy [51]. It is particularly useful in detecting multifocal lesions although can lead to false positives and an overestimation

of tumour size [51]. Ultrasound is also superior to traditional mammography in identifying multifocal lesions and allows for a more accurate evaluation of tumour size [50]. Tumours may also be subclinical by palpation and imaging modalities and instead present with metastases, known as *de novo* metastatic disease. Overall, the elusive nature of ILC on clinical and radiological assessment means that this breast cancer subtype tends to present at a later stage compared to IC-NST [52].

An area of particular interest is the somewhat distinct pattern of metastatic spread of ILC, which shows higher rates of spread to skin and bone compared to IC-NST as well as spread to non-traditional sites such as the gastrointestinal and gynaecological tracts, peritoneum and leptomeninges, reflecting the unique biology of ILC [53]. In contrast, extension to the lung is more frequently observed in IC-NST [53]. Similar to primary ILC of the breast, metastases are usually diffusely infiltrative and hypometabolic (having a low metabolic rate), instead of mass-forming and hypermetabolic (having an increased metabolic rate) [54]. This results in them being challenging to detect using conventional and metabolic based imaging modalities. Consequently, ILC metastases often present late with associated complications. Patients for example with intra-abdominal metastases often present with the clinical complications of extensive metastatic disease e.g. bowel obstruction, liver failure, ascites and hydronephrosis [54].

In contrast to IC-NST, ILCs show a strong tendency to spread to the peritoneum and retroperitoneum with two autopsy series of ILC patients identifying peritoneal metastatic deposits in 60 - 90% of cases in contrast to 15% of IC-NST [55, 56]. Consistent with the histological features of the primary disease, the peritoneal metastases in ILC were diffusely infiltrating, in contrast to more solid nodular deposits observed in IC-NST [55, 56]. Early peritoneal and retroperitoneal ILC metastases are clinically silent and present as stranding on imaging. When the indistinct lesions of peritoneal carcinomatosis become confluent they result in ‘omental caking’ which describes the radiological appearance of a thickened solid omentum and is a feature of advanced stage disease [57]. Peritoneal disease is associated with endocrine resistance and the prognosis following the diagnosis of peritoneal metastases is poor, with one study reporting a median survival time of 19 ± 9 months [58]. Metastatic retroperitoneal disease is also associated with advanced stage and poor outcome and can cause retroperitoneal fibrosis which can result in ureteral obstruction and hydronephrosis (hydrostatic dilation of the renal calyces and renal pelvis resulting from obstruction) [59].

When peritoneal abnormalities and ovarian lesions are present in a patient with a history of ILC, it can prove difficult to differentiate between metastatic breast cancer and metastatic

ovarian cancer [57]. ILCs show a relatively stronger tendency to spread to the gynaecological tract especially compared to IC-NST with reports of up to 13% of metastatic ILC patients showing ovarian lesions [53, 55, 60]. The presence of bilateral solid ovarian tumours known clinically as Krukenberg tumours, favours a diagnosis of metastatic breast cancer spread to the ovaries since primary ovarian cancers tend to have an admixture of both cystic and solid components [61, 62]. On the other hand, a subset of breast cancer patients; those harbouring *BRCA1* and *BRCA2* mutations, carry a higher risk for the development of both primary ovarian and breast tumours. *BRCA1* germline mutations are uncommon in ILC whilst *BRCA2* mutations show a higher frequency [39]. When the histology of the primary breast tumours of 6,893 *BRCA1* and *BRCA*2 mutant patients was assessed, only 2.2% of *BRCA1*-associated tumours were ILCs whereas the proportion of ILCs in the *BRCA2* mutant population was 8.4% [63]. It is therefore important to distinguish between metastatic ILC vs primary breast and ovarian disease in *BRCA* mutant ILC patients.

	ILC	**IC-NST**
Patient age	Older (median age 63)	Younger (median age 59)
Tumour size at presentation	Larger	Smaller
Tumour stage at presentation	Higher proportion of stage 3 and 4 cases	Lower proportion of stage 3 and 4 cases
Tumour grade	Predominantly grades 1 and 2	Higher proportion of grade 3 tumours
Tumour site	More frequently multifocal with bilateral lesions	Lower frequency of multifocal and bilateral lesions
Siites of metastatic spread	Unique sites: gastrointestinal tract, gynaecological tract e.g.ovaries, peritoneum, leptomeninges, ophthalmic region	Traditional metastatic sites: lung, liver, bone, brain
ER / PR / HER2 status	Higher rates of ER and PR positivity, lower rates of HER2 positivity	Lower rates of ER and PR positivity, higher rates of HER2 positivity
Surgery	Higher mastectomy rates	Higher rates of breast conserving surgery
Chemosensitivity	Poor responses to chemotherapy	Greater chemosensitivity
10 year Survival Outcome	Worse OS and DFS	Improved OS and DFS

Table 1.1: Clinical Differences between ILC and IC-NST [4, 6, 58, 64]

1.4 Survival Outcomes in ILC

The OS and disease-specific survival (DSS) for patients with ILC are poorer than IC-NST after 5 years (Figure 1.4) [6]. An explanation for the higher rates of late recurrence and associated worse outcomes in ILC is the presence of disseminated cancer cells which are distributed away from the primary tumour prior to surgery existing as minimal residual disease following primary surgery. These cells have the ability to rest dormant for extensive periods of time and upon reactivation may result in aggressive tumour growth and overt metastases at distant sites [65]. Metastatic ILC therefore often presents several years after remission with patients sometimes failing to declare a previous ILC diagnosis due to such an extensive period between ILC treatment and metastatic presentation [65].

A study comparing luminal A ILC and luminal A IC-NST at the gene expression level using data from TCGA and METABRIC showed that pathways that were less enriched in ILC were associated with protein translation and metabolism which are established characteristics of tumour dormancy [66]. A further study quantified circulating tumour cells (CTCs) in patients with ILC vs IC-NST and demonstrated that whilst ILC patients showed increased numbers of CTCs, the prognostic significance of CTCs was less strong in ILC compared to IC-NST implying a more dormant nature of these cells [67]. Overall, dormant disseminated cancer cells are recognised to play a key role in late disease recurrence in ILC and targeting this dormant niche may play a key role in delaying or preventing subsequent disease recurrence. Developing a better understanding of the dormant cell population and the specific molecular or microenvironmental factors which press the switch to reawaken the dormant cells, such that they become aggressive and grow into overt macro-metastases, will be critical for developing more effective and novel therapeutic strategies in ILC.

Survival outcomes among the ILC patient population also differ based on a range of clinical features such as tumour size, ILC histological subtype, tumour grade, lymph node involvement and hormone receptor status. For example, whilst ER+ disease accounts for the majority of ILCs, triple-negative (TN) ILCs are associated with a worse prognosis. A study compared outcomes in 38 TN-ILCs, 76 ER+ ILCs and 76 TN-IC-NST cases [68]. Whilst there was no significant difference in survival outcomes between TN-ILC and TN-IC-NST, when TN-ILCs were compared to ER+ ILCs, distant metastasis-free survival (DMFS) and OS were significantly worse in TN-ILC patients ($p = 0.039$, $p = 0.047$, respectively) [68]. Pleomorphic ILC, a rare histological subtype has also been shown in numerous studies to be associated with higher recurrence rates and in some studies, worse OS [69-72].

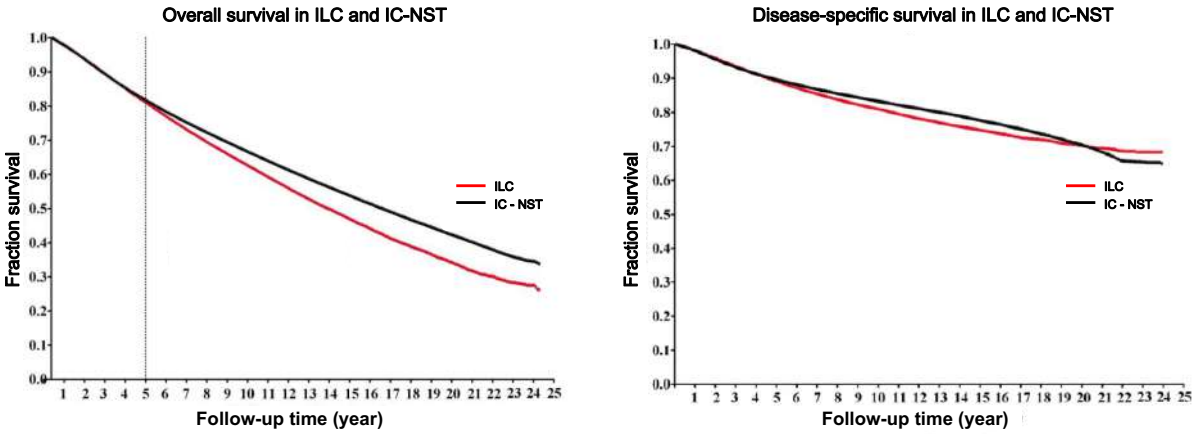


Figure 1.1: Comparison of overall survival (OS) and disease-specific survival (DSS) rates of IC-NST (n = 711,287) and ILC (n = 85,048): The OS curve shows an early survival advantage for the ILC patients before 5 years, (ILC vs IC-NST hazard ratio (HR): 1.118, $p < 0.0001$) but after 5 years an advantage for the IC-NST patients is seen (ILC vs IC-NST, HR: 0.775, $p < 0.0001$). The DSS curve shows that the IC-NST patients had improved survival over the ILC patients, both early and long-term (ILC vs IC-NST, HR: 0.809, $p < 0.0001$) Taken from [6]

1.5 Histology of Lobular Neoplasia

1.5.1 *In situ* Lobular Neoplasia

In terms of evolution, ILCs develop from a group of non-obligate precursor lesions know as lobular carcinoma *in situ* (LCIS) and atypical lobular hyperplasia (ALH) both of which are characterised by *in situ* proliferation of neoplastic lobular cells within the terminal duct lobular unit (TDLU) [73]. LCIS is considered both a risk factor and a non-obligate precursor of ILC with a relative risk of development of ILC following a diagnosis of LCIS being 9 - 10 times greater than that of the general population [74].

The most common form of LCIS is classic LCIS (cLCIS). In both cLCIS and ALH the cells are discohesive and can be polygonal, cuboidal, or round with light/clear cytoplasm [73]. They have high nucleus/cytoplasm ratios, with round/oval nuclei which are small to moderately sized with occasional nucleoli and indentations from intracytoplasmic mucin-containing vacuoles. The cells are highly uniform and monotonous and pagetoid growth can occasionally occur whereby the proliferating neoplastic cells grow above the basement membrane undermining the normal epithelial cell lining [73].

The degree of distention within the lobular units distinguishes LCIS and ALH. In LCIS greater than half of the acini within the TDLU are distended by the proliferation of neoplastic cells with at least 8 cells across each acinus. In contrast ALH is characterised by minimal distension with neoplastic cells filling less than half of the acini [73]. An important variant of LCIS is pleomorphic lobular carcinoma *in situ* (pLCIS). Here the architecture of the lesion and cellular discohesion is the same as that observed in cLCIS yet the cells themselves are larger, with high-grade (grade 3) pleomorphic nuclei as well as more abundant cytoplasm [73]. In addition, areas of central necrosis and calcification are frequently observed. Mitoses including atypical forms can also be seen, yet these are very rare in cLCIS [73].

PLCIS is less frequently ER+ compared to cLCIS and expresses HER2 more frequently [73]. In the past differentiating pLCIS from high-grade ductal carcinoma *in situ* (DCIS) has been challenging in view of their morphological similarities yet advances in immunohistochemistry (IHC) have made this achievable, in particular the use of E-cadherin IHC, as E-cadherin expression is normally absent in pLCIS and present in DCIS [73].

There is limited data assessing the clinical behaviour of pLCIS, yet it is widely considered to represent a more aggressive variant of LCIS, with a higher risk of progression to ILC [73]. Based upon this, the UK NHS Breast Screening Programme together with the European Society of Medical Oncology recommend that pLCIS should be managed the same as DCIS instead of cLCIS, given its molecular and biological profile [75]. Florid LCIS (FLCIS) is no longer recognised as an LCIS subtype but previously was defined as an architectural variant of LCIS that has the histological and radiographic appearance of solid DCIS, but lacking E-cadherin expression [76].

1.5.2 Invasive Lobular Carcinoma

The majority of ILCs are classed histologically as classic ILC but other subtypes have been identified based on varying architectural patterns (solid, alveolar, tubulo-lobular) and cytological features (pleomorphic ILC) [1, 5, 73, 77]. A 'mixed-group' defines cases containing an admixture of the classic type with one or more additional subtypes [73, 78]. Together the classic and mixed subtypes account for the majority (75%) of all ILCs. In addition, distinct from the mixed lobular group, 3 - 5% of all invasive breast cancers show mixed features of IC-NST and lobular differentiation and are termed 'mixed ductal-lobular tumours' [73].

In classic ILC, tumour cells are small, uniform and lack cohesion, with round or notched oval nuclei. Intra-cytoplasmic lumina, containing mucin inclusions may also be present. Classic ILC shares the same cytological features as cLCIS with which it is associated in 58 - 98% of cases [73]. The characteristic pattern of growth of classic ILC involves the infiltration of single files of tumour cells or individually dispersed single cells throughout the stroma, with minimal disturbance of the normal tissue architecture, or desmoplastic reaction [73, 78]. This lack of desmoplastic stromal response enables ILCs do go clinically undetected such that patients present at a later stage than IC-NST. It also makes them challenging to diagnose radiologically using standard mammography, and difficult to obtain clear resection margins upon surgical excision [79]. Another well-recognised growth pattern of classic ILC is a 'targetoid pattern' whereby invading tumour cells arrange in a concentric manner around normal ducts. Mitoses are infrequent in classic ILC and lymphovascular invasion is uncommon [73].

The same characteristic cells of classic ILC are seen in the solid, alveolar, and tubulo-lobular variants, but these tumours differ in their architectural patterns. Solid ILC is characterised by sheets of tumour cells and a higher rate of mitosis than classic ILC and has also been associated with significantly worse survival outcomes than classic ILC [73, 80]. Alveolar ILC

is characterised by globular aggregates of at least 20 cells as opposed to single files and shows the highest rate of ER positivity [73, 78]. The tubulo-lobular variant is a rare subtype showing an admixture of a tubular growth pattern and small uniform cells arranged in a linear pattern and has an excellent prognosis [73, 78, 81].

Pleomorphic ILC is defined by WHO as a subtype that retains the distinctive growth pattern of classic ILC but shows greater cellularity and a greater degree of cellular atypia and pleomorphism than the classic form [73, 78]. There is nuclear enlargement, marked variability in the size and shape of the nuclei, hyperchromasia (darker staining) and prominent nucleoli may also be seen. Moreover pleomorphic ILC is characterised by the presence of nuclei > 4 times the size of a lymphocyte or equivalent to that of high-grade DCIS [73, 78]. Apocrine or histiocytoid differentiation and signet-ring cells may also be seen and it is also associated with more frequent mitoses. This pattern is also frequently associated with pLCIS [73, 78]. A summary of the histological appearances of the various ILC subtypes is provided in Figure 1.2:

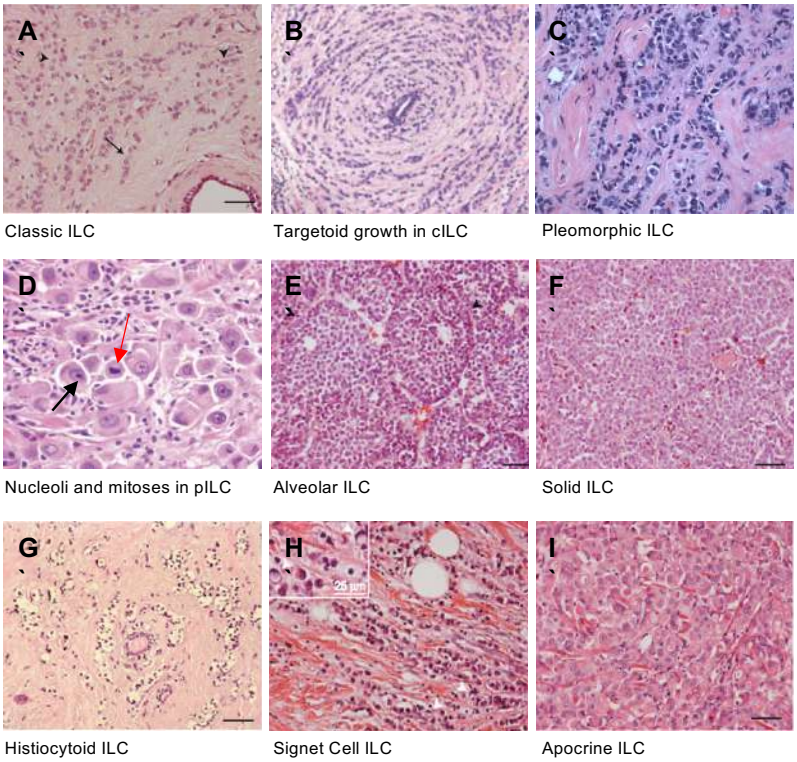


Figure 1.2: ILC Subtype Histology: Representative H&E sections showing the range of different histological appearances of ILC: A) Classic ILC, with arrows depicting the characteristic small, round uniform tumour cells B) Concentric, targetoid growth pattern of classic ILC around a vessel C) Pleomorphic ILC showing the single file growth pattern with greater nuclear pleomorphism D) Black arrow indicates a prominent nucleolus and red arrow shows an atypical mitosis and nuclear pleomorphism observed in pleomorphic ILC E) Alveolar ILC F) Solid ILC G) Histiocytoid ILC H) Signet cell ILC I) Apocrine ILC - adapted from [82].

1.5.3 Immunohistochemistry (IHC) in ILC

Loss of E-cadherin expression is an important diagnostic feature of ILC and is also observed in LCIS. The mis-localisation of p120-catenin from the membrane to the cytoplasm is an additional positive IHC marker for ILC [83-86]. This combination can be used to aid in the diagnosis of cases where it is difficult to distinguish between ILC and IC-NST on H&E (Table 2, Figure 1.3).

Tumour Type	E-cadherin	p-120 catenin
ILC	absence of staining	cytoplasmic
IC-NST	membranous	membranous

Table 1.2: Differences in IHC staining patterns between IC-NST and ILC [83, 84]

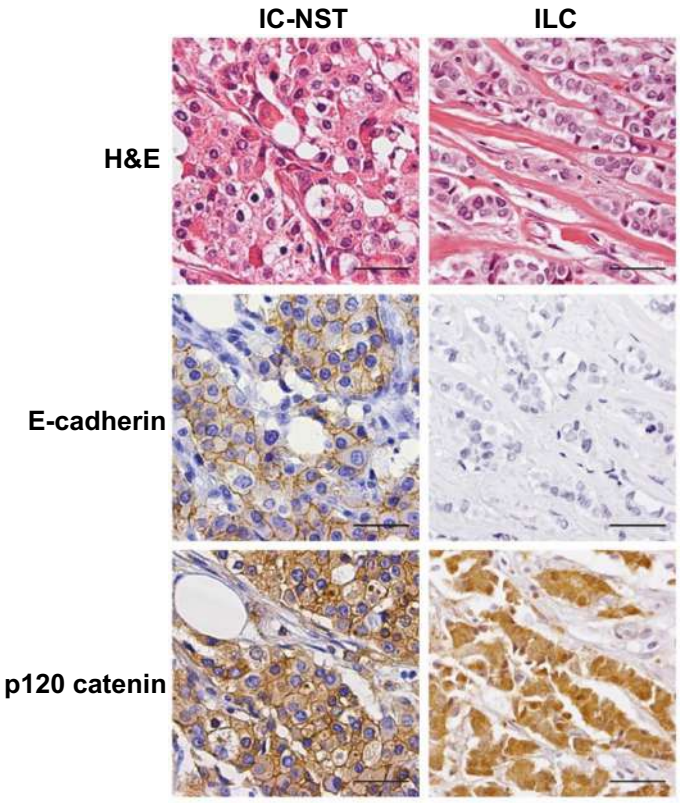


Figure 1.3: Differential IHC staining patterns in IC-NST and ILC: Left panel shows IC-NST with positive membranous staining for both E-cadherin and p120-catenin and right panel shows ILC with loss of E-cadherin expression and cytoplasmic staining for p120-catenin [87]

IHC findings must be interpreted with caution since a reported 10 - 16% of ILCs (and most ductal-lobular tumours) retain or show 'aberrant' E-cadherin expression, characterised by reduced membranous staining intensity, or partial, granular, fragmented membranous staining or cytoplasmic staining [86, 88]. In these cases, it is hypothesised that the E-cadherin binding complex is dysfunctional due to underlying *CDH1* alterations or aberrant loss of catenin binding proteins [88, 89].

1.6 Molecular subtyping, hormone and HER2 status in ILC

Based upon hormone receptor status and the level of Ki-67 expression (a proliferation marker assessed using IHC) breast cancers including ILC can be grouped into molecular subtypes, including luminal A (ER/PR+, HER2-, Ki-67 < 14%); luminal B (ER/PR+, HER2+/-, Ki-67 > 14%); HER2-enriched (ER-, PR-, HER2+) and TNBCs (ER-, PR-, HER2-) [90]. The basal-like subtype refers to TNBCs showing expression of basal markers (e.g. CK5/6) [91]. The vast majority of ILCs fall into the luminal A molecular subtype [49, 92].

ER+ breast cancer accounts for 70% of breast cancer and is defined by oestrogen receptor positivity as assessed through IHC [93]. ER+ breast cancers may also be PR+ or PR-. Oestrogen and progesterone are steroid hormones, secreted by the ovaries which play essential roles in the female reproductive system and whose levels change within the breast tissue throughout puberty, pregnancy, lactation and during the menopause [94]. In ER+ disease, the epithelial tumour cells retain the oestrogen receptor which is ordinarily present in the normal breast epithelium forming the mammary glands. The oestrogen receptor itself is a transcription factor which upon activation through binding of its corresponding ligand (oestrogen) translocates to the nucleus. Here it binds to DNA, regulating gene expression [94]. Ultimately the interaction between oestrogen and its receptor leads to signalling which drives the transcription and proliferation of the epithelial cells. Approximately 90% of ILCs are ER+ and 60 - 70% are PR+ [49, 92]. Rates of ER positivity are therefore higher than in IC-NST which are ER+ in 70% of cases [93].

Human epidermal growth factor receptor 2 (HER2) is amplified and the protein overexpressed in a reported 20 - 25% of breast cancers, and this confers an aggressive tumour behaviour [95, 96]. HER2+ tumours have been associated with higher rates of recurrence and increased mortality [97, 98], although in an era of effective HER2 targeted therapies, survival outcomes are approaching those of luminal disease [99]. In classic ILC *HER2* mutations and amplifications are rare, yet reported rates are higher in pleomorphic ILC, with a mutation rate of 32% based on the TCGA dataset and reported amplification/overexpression rates of 19 - 33% [100-103]. The *HER*2 gene is situated on the long arm of chromosome 17 [104, 105]. The encoded protein is a 185-kDa transmembrane tyrosine kinase receptor and a member of the epidermal growth factor receptor family along with 3 other receptors (EGFR (HER1), HER3 and HER4) [104-106]. The HER2 extracellular domain does not have a known ligand unlike other epidermal growth factor receptors and its activation results from the formation of homo and heterodimers [97, 105]. These dimers lead to the phosphorylation of tyrosine kinase residues present within the cytoplasmic domain of the receptor which provide docking sites

for proteins which in turn activate the mitogen activated protein kinase (MAPK) and phosphatidylinositol triphosphate kinase (PI3K) signalling pathways [97, 105]. This leads to cell-cycle progression and proliferation [97, 105]. In clinical practice patients are tested for HER2 status since this has implications for clinical management, identifying candidates for anti-HER2 therapy. IHC assessing HER2 expression at the protein level is commonly used and fluorescence *in situ* hybridization (FISH), can also be used which assesses for gene amplification. In many centres IHC analysis is used in the first instance, with FISH testing reserved for cases with equivocal results on IHC [107, 108].

Tumours which lack the expression of ER, PR and HER2 are known as triple-negative breast cancers (TNBCs). TNBC It is a highly heterogeneous disease which accounts for 10 - 15% of all breast cancers [109]. TNBCs in general are more common in younger patients and those with *BRCA* mutations. They are usually high grade, biologically aggressive and associated with poor outcomes [110, 111]. In terms of molecular subtype, the majority (80%) show a basal-like transcriptomic profile [112-114]. TN-ILCs are rare, reported to account for 1 - 1.4% of TNBCs based on two recent analyses using SEER and NCDB databases and for 1 - 2% of ILCs [115, 116].

A recent study assessed a series of primary triple-negative (TN) ILCs (n = 38), comparing clinical, histological, and molecular features of these tumours to TN-IC-NST (n = 76) and ER+ ILCs (n = 76). TN-ILC patients were significantly older than ER+ ILC ($p = 0.002$) and TN-IC-NST patients ($p < 0.001$) [68]. At the histological level, TN-ILCs were more frequently pleomorphic, with pleomorphic ILCs accounting for 50% of cases ($p = 0.003$). They were also of higher nuclear grade than ER+ ILCs ($p = 0.009$). IHC showed that TN-ILCs expressed basal markers (e.g. EGFR, SOX10 CK5/6) less frequently than TN-IC-NST ($p < 0.001$) [68]. In contrast positivity for the androgen receptor (AR) was more frequent in TN-ILC ($p < 0.001$). At a molecular level, TN-ILCs also showed distinct transcriptomic profiles, with increased AR signalling, as well as alterations in the phosphatidylinositol-3-kinase (PI3K) signalling pathway and an enrichment for *HER2* mutations (present in 26.9% of cases). Survival outcomes were also significantly poorer in TN-ILCs compared to ER+ ILCs [68]. Overall these rare, clinically aggressive tumours predominantly form part of the luminal androgen receptor subtype, which accounts for 10 - 15% of TNBCs and is characterised by AR protein expression [117], and they harbour genomic alterations which are relevant for targeted therapeutic approaches.

1.7 Pleomorphic ILC: An aggressive ILC variant

Pleomorphic ILC is a rare histological form of breast cancer and an uncommon subtype of ILC. It accounts for approximately 15% of ILCs and less than 1% of all invasive breast cancers [77, 118]. Whilst classic ILCs are generally ER/PR+, HER2- and have low Ki-67 scores, [1, 5, 77], reported rates of ER and PR positivity are lower in pleomorphic ILC, HER2 amplification is found in 19 - 33% of cases [100-103] and Ki-67 scores are generally higher than classic ILC [1, 69, 77, 119]. In addition, 12.5- 22.7% of pleomorphic cases are negative for ER, PR and HER2 (i.e. TNBC) [77, 120-122]. These less favourable biomarkers, together with various molecular alterations observed at higher frequency in pleomorphic ILC (e.g. *TP53* and *HER2* mutations) [101, 102, 123, 124], contribute to the more aggressive clinical nature of these tumours compared to classic ILC. They present at a more advanced stage, with larger, higher-grade tumours and higher rates of lymphovascular invasion, lymph node involvement and distant metastases [69, 77, 125]. Relatively higher numbers of patients require a mastectomy as opposed to more localised breast-conserving surgery (BCS) and patients show higher rates of recurrence [69, 125].

Earlier studies using small pleomorphic ILC cohorts conclude that pleomorphic ILC is associated with worse OS and disease-free survival (DFS) than classic ILC and this has been supported by some larger more recent studies [69-72]. However, others, when adjusting for a range of other prognostic factors have found that pleomorphism alone in not an independent predictor of worse outcome [126-129]. Nevertheless, nuclear pleomorphism is a key component of histological tumour-grade which itself has been significantly associated with BCSS and DFS [128, 130, 131]. The recognition of this histological subtype and the development of our existing knowledge of this rare tumour type is therefore still highly relevant.

1.8 Therapeutic Strategies in ILC

The clinical management of ILC varies depending on the stage at diagnosis and on the tumour biology. Despite representing a distinct disease entity, ILC is generally managed the same as IC-NST.

1.8.1 Role of Surgery

For surgical resection of the tumour in primary ILC, many patients undergo mastectomy which removes the entirety of the affected breast. The alternative option is BCS in the form of a wide-local excision, which involves removing only the tumour with a margin of normal tissue, in combination with postoperative radiotherapy. Mastectomy has been considered preferable for larger and multicentric/multifocal tumours whilst BCS is an alternative option in patients with early-stage unifocal tumours. As previously discussed, ILCs are generally larger with higher rates of multicentricity/multifocality at diagnosis compared to IC-NST. Moreover, the reduced fibrotic reaction in ILC and diffusely infiltrative growth pattern creates difficultly for surgeons in establishing the gross extent of disease at surgery and creates concerns about the potential for positive margins following BCS. For these reasons studies have demonstrated that mastectomy rates tend to be higher in ILC compared to IC-NST. For example, when the clinical management and outcomes of 4140 ILC patients were compared to 45,169 IC-NST patient, mastectomy rates appeared to be significantly higher in the ILC group ($p < 0.0001$).

Interestingly, studies show that mastectomy and BCS result in comparable outcomes in terms of locoregional and DMFS, and BCSS in ILC patients [132]. Even in patients presenting with larger tumours (≥ 4 cm), a recent cross-sectional multivariate analysis was performed in a cohort of 180 ILC patients who were treated with either mastectomy (n = 150) or BCS (n = 30) and showed no significant difference in recurrence free survival at 5 or 10 years in patients undergoing BCS vs mastectomy ($p = 0.88$, $p = 0.65$ respectively) [133]. Despite this, ILC patients who receive BCS often require re-excision of breast tissue in order to obtain clear surgical margins as positive surgical margins following BCS are reported in up to 60% of cases [134]. In this situation the options are a re-excision lumpectomy or completion mastectomy.

1.8.2 Radiotherapy

Radiotherapy is a treatment approach using X-rays to kill cancer cells by inducing DNA damage, in particular double strand DNA breaks which result from the high energy of x-rays. This leads to cell death as the tumour cells lack the ability to quickly repair the DNA damage [135, 136]. In breast cancer radiotherapy is commonly used in the adjuvant setting following surgical removal of the primary to reduce the likelihood of local recurrence. It can also be used in the metastatic setting in order to enable the effective palliation of metastases which are symptomatic [137].

Radiotherapy is the standard of care for early-stage breast cancers. As previously described mastectomy rates have been reported to be higher in ILC compared to IC-NST, often due to concerns about the likelihood of obtaining clear surgical margins, potential tumour recurrence in the same breast and the higher risk of contralateral disease. A retrospective study investigated the rates of local-regional recurrence, contralateral breast cancers and DSS in ILC vs IC-NST patients who were treated with BCS followed by radiotherapy and showed that there were no significant differences in the rates of these outcomes in ILC vs IC-NST [138]. This suggests that this treatment strategy is appropriate in ILC patients with early-stage disease. In addition, post-mastectomy radiation among ILC patients has been shown to be just as effective as in IC-NST leading to improvements in local recurrence rates and survival outcomes [139].

ILC is more frequently multifocal compared to IC-NST and therefore ILC histology has been an exclusion criterion for trials assessing partial breast irradiation (PBI) and this approach is not recommended for ILC patients [140]. A retrospective review conducted at a single institution assessing the use of PBI found higher rates of ipsilateral breast cancer recurrence among ILC patients [141]. There have been two large hypofractionated radiotherapy trials which enrolled early-stage breast cancer patients, ‘FAST-Forward’ (NCT04148586) [142] and ‘The UK Standardisation of Breast Radiotherapy (START)’ trial [143], however rates of local recurrence have not been reported according to histological breast cancer subtype.

1.8.3 Chemotherapy

Chemotherapies have been used since the 1940's in oncology with the primary goal of inhibiting tumour cell proliferation, invasion, and metastasis [144]. Tumours which are highly proliferative show greater responses to chemotherapy compared to tumours with low proliferation rates (such as ILCs) [145]. However additional factors can contribute to chemotherapy induced tumour cell death such as a high degree of genomic instability, DNA repair defects and mitochondrial priming [146-148]. Moreover, chemotherapy-induced cell death can activate and drive an immune response promoting anti-tumour immunogenicity [149, 150].

Chemotherapy encompasses a range of different drugs with different mechanisms of action [144]. It may be used in the neoadjuvant, adjuvant, combined and metastatic settings. Neoadjuvant therapy refers to treatment given before surgery with the aim of downstaging and reducing the size of large tumours. This enables BCS as opposed to mastectomy [144, 151]. It also provides the unique opportunity to measure early *in-vivo* responses to systemic therapy. Neoadjuvant chemotherapy is a therapeutic option for the management of locally-advanced operable and primarily non-operable breast tumours.

When clinical outcomes following the use of neoadjuvant chemotherapy have been assessed in multiple retrospective studies, it is evident that ILC patients derive less benefit from neoadjuvant chemotherapy than IC-NST patients [34, 152-155]. For example, a large retrospective study examined the surgical benefit and chemotherapy sensitivity of 1895 patients who had presented with stage I–III, ER+ breast cancer and who had been treated with neoadjuvant systemic chemotherapy [152]. The study included 177 ILC patients and the most frequently used treatment regimen was a combination of a taxane and anthracycline, which was used at comparable rates in ILC and IC-NST patients. Following systemic therapy, a significant downstaging compared to baseline was observed in both histological types ($p < 0.001$) however significantly lower numbers of ILC patients had a reduction in tumour T stage (reflecting tumour size) compared to IC-NST patients (41% vs 64% respectively, $p < 0.0001$) [152]. Positive or near (≤ 2 mm) tumour surgical resection margins were significantly more prevalent in the ILC group (19% vs 11%, $p < 0.001$). As a result, rates of BCS were lower in ILC compared to IC-NST patients (19% vs 34%; $p < 0.001$) and breast histology was an independent predictor of the use of mastectomy in multivariate analysis ($p = 0.01$). When pathological complete response (pCR) rate was assessed, the ILC group showed significantly lower pCR rates compared to IC-NST (3.5% vs 14%, $p < 0.001$) in univariate analysis, yet in multivariate analysis, statistical significance was lost [152].

A further study assessed pCR rates following neoadjuvant chemotherapy in ILC vs IC-NST and concluded that significantly lower pCR rates were observed in ILC [34]. The study included 1,034 stage II/III patients who participated in 6 clinical trials. It included 122 (12%) ILC patients and 912 (88%) IC-NST patients, who were all treated with anthracycline-based neoadjuvant chemotherapy, and an additional 33.5% of patients also received taxane-based treatment. The ILC group showed lower rates of pCR (3% vs 15%; $p < 0.001$) and showed increased numbers of residual involved axillary lymph nodes following therapy (41% vs 26% > 3 involved nodes; $p = 0.001$) [34]. A further study assessed the impact of chemotherapy sequence in relation to surgery in node-positive ILC patients and showed that chemotherapy administered in the neoadjuvant setting was associated with worse survival outcomes compared to adjuvant therapy [156].

Adjuvant therapy is treatment administered following surgery, with the intention of eliminating or at least suppressing the growth of occult tumour cells or micro-metastases [144]. Synergistic strategies are commonly employed, using combinations of chemotherapeutic agents. In breast cancer, combinations can typically include fluorouracil, cyclophosphamide, doxorubicin, paclitaxel and docetaxel [157]. Combinations of drugs with differing mechanisms of action as well as non-overlapping toxicities are chosen to reduce the likelihood of the development of resistant clones and reduce toxicities [144].

A large study evaluated the use of adjuvant chemotherapy in non-metastatic ILC (n = 3685) and IC-NST (n = 19,609) [158]. Patients were treated either with adjuvant endocrine therapy alone, or with a combination of adjuvant hormonal therapy and chemotherapy. Whilst in the IC-NST patients the 10-year OS rate was higher in the patient group that received additional chemotherapy (74% vs 69% $p < 0.0001$), this was not the case in ILC, with comparable OS rates in those patients who received the chemotherapy and hormone therapy combination and those who were treated with endocrine therapy alone (66% vs 68% respectively, $p = 0.45$) [158]. This therefore suggests that the use of adjuvant chemotherapy provides no additional benefit for primary ILC patients receiving endocrine therapy, in contrast to patients with IC-NST.

The low chemosensitivity of ILC is accounted for by its hallmark biological characteristics including its low histological grade, oestrogen receptor positivity and low proliferation rate as assessed through Ki-67 IHC. Lack of chemosensitivity contributes to residual tumour volumes in ILC which present a risk for disease recurrence. Prognostic biomarkers are therefore important to accurately identify which ILC patients are most likely to derive long-term clinical

benefit from these toxic agents, given that chemotherapy is associated with a range of unpleasant side-effects and long-term complications [159, 160]. In the metastatic setting, it is generally restricted for patients who have become resistant to endocrine therapy or among patients with rapidly progressing visceral involvement and extensive symptomatic disease [31, 161-163].

1.8.4 Endocrine Therapy

Endocrine therapies are a class of targeted treatment which aim to block the signalling of the oestrogen receptor [164]. The use of endocrine therapy is a pivotal strategy in the treatment of both early and advanced-stage ER+ breast cancers, such as ILC. There are three main approved classes of treatment which target the oestrogen signalling pathway: selective oestrogen receptor modulators (SERMs), selective oestrogen receptor degraders (SERDs) and aromatase inhibitors (AIs) [164]. These may be administered with or without ovarian suppression.

SERMs are oestrogen receptor ligands which competitively bind to the oestrogen receptor, displaying antagonist or agonist effects depending on the target tissue. They are used in the treatment of both pre- and postmenopausal ER+ breast cancer patients [165]. In the breast they have an anti-oestrogenic antagonist effect, thus inhibiting oestrogen-dependent proliferation. The SERM-bound oestrogen receptor binds to oestrogen response elements and downregulates transcriptional activity by associating with co-repressors [165]. In contrast, SERMs show agonist activity in bone, and tamoxifen, the most used SERM, shows agonist activity in the uterus, increasing the risk of endometrial cancer. Its anti-oestrogen effects on platelets increase the risk of arterial and venous thrombosis [166].

SERDs have dual effects and act by both antagonising the oestrogen receptor transcriptional activity and promoting its degradation [167]. It acts as a pure antagonist of the oestrogen receptor and has been shown to have a greater affinity for the receptor compared to tamoxifen [168]. Fulvestrant was the first SERD developed and approved for use in ER+ breast cancer. for metastatic patients who had progressed on prior endocrine therapy with tamoxifen, or an AI and it has demonstrated efficacy in these patients. Importantly SERDs are more clinically effective than alternative endocrine therapies in the context of oestrogen receptor 1 (*ESR1*) mutation related endocrine therapy resistance [169]. Fulvestrant presents a challenge relating to the lack of oral bioavailability of the drug, meaning it must be administered by slow intramuscular injection [169]. However, more recently oral SERDs have been developed, and the FDA has given approval for an oral SERD elacestrant to be used in ER+, HER2- metastatic

breast cancer patients with *ESR1* mutations who have already received at least one line of prior endocrine therapy [169]. This approval was based upon results of the phase III EMERALD study (NCT03778931) [170].

AIs such as anastrozole, letrozole, and exemestane, are used in the treatment of postmenopausal ER+ breast cancer patients, including ILCs. They work by blocking the conversion of androgens to oestrogens in non-ovarian tissues reducing the levels of systemic oestrogen [171]. In ER+ breast cancer adjuvant AIs have been shown to significantly decrease the risk of disease recurrence following curative treatment, and they are also used as standard first-line treatment among ER+ metastatic patients. There has been some evidence that they are superior to tamoxifen in terms of long-term efficacy and safety among postmenopausal women with early-stage disease based on the results of the ATAC trial (Arimidex, Tamoxifen, Alone or in Combination) which assessed the use of Arimidex (anastrozole), tamoxifen and the combination among this patient group [172]. The development of resistance to AIs is frequently associated with *ESR1* mutations which result in ligand-independent activation of the oestrogen receptor [173]. Among premenopausal and perimenopausal patients, ovarian suppression using a luteinizing hormone releasing hormone agonist is used to decrease circulating levels of oestrogen. Ovarian suppression facilitates AI use among high-risk younger patients as it blocks intrinsic oestrogen production, enabling AI use where this is considered superior to tamoxifen treatment for further risk reduction [174].

Given that ILC is predominantly ER+, the use of neoadjuvant endocrine therapy has been considered. A small retrospective study assessing the use of the AI letrozole in 61 ER+ ILC patients with large tumours (T4) or locally advanced (N2) disease showed that neoadjuvant endocrine therapy resulted in a mean reduction in tumour size of 66% at a 3-month time point, resulting in a rate of successful breast conservation of 81% [175]. This suggests that neoadjuvant endocrine therapy may be more effective than chemotherapy in ILC patients with large tumours who desire BCS [175]. Indeed an ongoing randomized phase III clinical trial is assessing the use of neoadjuvant endocrine therapy in ER+ HER2- postmenopausal breast cancer patients with stage II and III disease including ILC patients. The ALTERNATE trial (NCT01953588) is assessing the use of fulvestrant, the combination of fulvestrant and anastrozole, with anastrozole monotherapy for reduction of tumour shrinkage [176] .

In the adjuvant setting, studies have demonstrated that treatment with adjuvant endocrine therapy in early-stage ER+ breast cancers (including ILC) results in significant improvements in DFS and OS [177, 178]. The standard approach involves the use of SERMs e.g. tamoxifen

among premenopausal patients and AIs in postmenopausal patients [179]. The use of AIs over tamoxifen results in superior DFS in postmenopausal patients and a group of high-risk premenopausal patients when given together with ovarian suppression [180].

Equally, adjuvant endocrine therapy represents the pillar in the treatment of metastatic ER+ ILC patients, often in combination with a CDK4/6 inhibitor. Previous work has suggested that the differential benefit of AIs over tamoxifen is greater in ILC compared to IC-NST [33]. However, a meta-analysis was recently presented at the San Antonio Breast Cancer Symposium 2021 which included a total of 7415 patients who were part of the TEAM, BIG 1-98 and the 'Arimidex, Tamoxifen, Alone or in Combination' (ATAC) trials, and the differential efficacy of AIs over tamoxifen in ILC compared to IC-NST was not confirmed. Although AIs appear to be superior to tamoxifen in the treatment of ILC, the meta-analysis failed to support a more pronounced effect among ILC patients compared to IC-NST [181].

With regards to premenopausal ER+ breast cancer patients, the 'Suppression of Ovarian Function Trial' (NCT00066690) [182] as well as the 'Tamoxifen and Exemestane Trial' (NCT00066703) [183] demonstrated that among patients with high-risk tumours, ovarian suppression can be advantageous, resulting in significant improvements in DFS and OS [180]. The use of ovarian suppression specifically in premenopausal ILC patients has not been studied, possibly due to the low incidence of ILC among younger women.

1.8.5 Targeted therapy: CDK4/6 Inhibition

Targeted therapies are pharmacological agents designed to inhibit tumour growth, cause tumour cell death, and ultimately limit the metastatic potential of a tumour. However unlike traditional chemotherapy, targeted therapies are designed to interfere with very specific proteins which are driving tumorigenesis [184]. The targeting of specific molecular alterations which are unique to a particular cancer may be a superior strategy than broad-based treatments and advances have enabled the molecular analysis of individual patients' tumours and the tailoring of treatments based upon this. Targeted cancer therapies encompass three main groups: small molecule inhibitors, monoclonal antibodies, and immunotoxins [184].

CDK4/6 inhibitors are small molecule inhibitors which represent a key group of targeted therapies for ER+ patients and they include palbociclib, ribociclib and abemaciclib. In breast cancer as well as other cancer types, deregulation of various components of the cyclin D1/CDK4/6/Retinoblastoma (Rb) signalling pathway results in unchecked cellular proliferation and tumour growth [185, 186]. CDK4 and CDK6 are cyclin D-specific kinases which

phosphorylate the retinoblastoma protein. CDK4/6 inhibitors target the cell-cycle machinery and interrupt intracellular and mitogenic hormone signalling pathways which drive the growth and proliferation of tumour cells [185, 186]. They work by blocking the G1/S cell-cycle checkpoint where CDK4 and CDK6 are activated by D-type cyclins, resulting in cell-cycle arrest [185, 186]. They have also been shown to cause DNA damage and replication stress resulting in long-term cell-cycle withdrawal [187].

CDK4/6 inhibitors have revolutionised the treatment of metastatic ER+ breast cancer with proven efficacy in both the endocrine-sensitive [188-190] and endocrine-resistant [191-195] disease settings. Significant improvements in OS have been observed in both premenopausal and postmenopausal patients and these improvements appear to be independent of the endocrine therapy partner [193, 196-199]. As a result, the use of CDK4/6 inhibitors in combination with endocrine therapy has become standard of care for metastatic ER+ patients, including ILCs. More recent evidence suggests a significant survival benefit of CDK4/6 inhibitors in ER+ HER2+ patients [200, 201].

In the context of early-stage disease, based on the results of the monarchE phase III trial [202], the FDA recently approved the use of abemaciclib in combination with endocrine therapy for adjuvant treatment of patients with ER+ HER2- early-stage node-positive disease at high risk of recurrence. The differential benefit in ILC and IC-NST was not assessed. Interestingly, two further trials; PENELOPE-B [203] and the 'Palbociclib Collaborative Adjuvant Study' (PALLAS) [204] assessed the addition of the CDK4/6 inhibitor palbociclib to adjuvant endocrine therapy in ER+, HER2- early breast cancers (high-risk patients in PENELOPE-B) and failed to demonstrate significant clinical benefit from Palbociclib [203, 204].

PALOMA-2 was the only trial which has reported outcomes specifically in ILC patients demonstrating a significant improvement in progression-free survival (PFS) in metastatic ILC patients receiving the CDK4/6 inhibitor palbociclib and letrozole (AI) vs letrozole alone [188]. A pooled analysis of the trials assessing CDK4/6 use in metastatic ER+ disease, also showed that the addition of CDK4/6 inhibitors resulted in significant improvements in PFS among ILC patients [205]. A further pooled analysis of 3 trials which included ILC patients, assessed the use of CDK4/6 inhibitors in combination with fulvestrant and showed a survival benefit from the addition of CDK4/6 inhibitors [206]. A separate study pooled data from 5 phase III trials of CDK4/6 inhibitors with either an AI in the 1st-line or fulvestrant in the 2nd-line setting and showed that the addition of a CDK4/6 inhibitor resulted in a similar benefit in ILC patients as IC-NST [205]. Overall CDK4/6 inhibitors represent a powerful additional class of agent in the

context of ER+ breast cancers including ILC, although the development of treatment resistance poses a therapeutic challenge.

1.8.6 HER2 Targeted therapy

One of the more studied targets in breast cancer is the receptor tyrosine kinase, HER2. In the late 1990's trastuzumab (Herceptin), a human monoclonal antibody was approved as the first HER2 targeted therapy [207]. It binds to the extracellular domain of the HER2 receptor with high affinity and specificity, thus preventing HER2-mediated signalling [208]. It also facilitates antibody-dependent cell-mediated cytotoxicity which is driven by natural killer (CD56+) immune cells, resulting in the death of HER2+ cells [209]. Several other HER2 targeted therapies have since been approved. Furthermore, antibody-drug conjugates such as ado-trastuzumab emtansine have been developed. These are novel treatments which combine the benefits of specifically targeting the tumour cell surface antigens through the antibody component, with the additional cell killing from the high-potency cytotoxic chemotherapy component, thus reducing peripheral side-effects of chemotherapy [210].

Trastuzumab emtansine (T-DM1) was the first antibody-drug conjugate targeting HER2 which received approval for HER2+ patients with advanced disease and more recently it has been approved in high-risk HER2+ patients with early-stage disease who have residual disease following neoadjuvant therapy [210, 211]. Trastuzumab deruxtecan (T-DXd) was subsequently approved for metastatic HER2+ patients who have previously received at least two lines of previous anti-HER2 treatments [210]. Interestingly a recent clinical trial (NCT03734029) showed that trastuzumab deruxtecan led to significantly longer PFS and OS compared to chemotherapy among HER2-low metastatic patients (defined as 1+ on IHC or 2+ on IHC with negative FISH results) who had received one or two lines of previous chemotherapy [212]. It has therefore been approved for this use [212].

Small molecule tyrosine kinase inhibitors also present a therapeutic option for HER2+ patients [213]. For example, neratinib, lapatinib, tucatinib and pyrotinib have shown promising results in phase III trials demonstrating efficacy either as monotherapy or in combination with other anti-HER2 treatments or chemotherapy in early-stage disease and the metastatic setting [214-219]. Evidence suggests that patients who progress whilst being treated with trastuzumab may derive some benefit from a HER2-targeted tyrosine kinase inhibitor with or without trastuzumab [217, 220] suggesting a separate mechanism of action of these drugs [221, 222].

HER2 mutations and amplifications are extremely rare in classic ILC but show higher frequency in pleomorphic ILC, occurring in a reported 32% and 19 - 33% of cases respectively [100-103]. HER2 positivity has been identified as an independent prognostic factor that is associated with poor survival outcomes in ILC [80]. Despite this, there are few studies assessing the level of benefit derived from anti-HER2 targeted therapies specifically in ILC.

The degree of benefit derived from therapy with adjuvant trastuzumab in ILC was studied in a retrospective analysis which used results from the international randomised phase III Herceptin Adjuvant (HERA) trial [223-225]. The study compared treatment with 1 or 2 years of trastuzumab to observations (following conventional chemotherapy) among women with HER2+ early-stage breast cancer. Results showed that therapy with adjuvant trastuzumab for 1 year following chemotherapy was associated with a significant improvement in DFS compared to observation and 2 years of treatment offered no additional benefit [225]. The study findings resulted in the drug being used as standard of care in HER2+ patients with early disease in the adjuvant setting [224]. ILCs accounted for 5.5% of patients included in the study and of these, 97 received a year of trastuzumab therapy whilst the remaining 90 were assigned to the observation group. Among the IC-NST subgroup, 1611 patients received a year of trastuzumab therapy whilst 1602 patients were assigned to the observation group. There were no differences in the pattern of disease recurrence or in the degree of benefit derived from trastuzumab therapy between the two histological patient groups [223]. Whilst the study did not further categorize ILCs into the various histological ILC subgroups, one would hypothesise that a large proportion of these tumours showed pleomorphic ILC histology given the higher rate of HER2 alterations in this subtype. Importantly the study highlights that ILC as a histological diagnosis is not synonymous with HER2- disease and therefore it is important that HER2 testing takes place in ILC. Treatment with a year of adjuvant trastuzumab offers benefit to patients diagnosed with early-stage, HER2+ ILC and the level of benefit from the treatment is the same as for patients with IC-NST [223-225].

1.8.7 Additional Targeted therapies

Additional pathways which are frequently activated in breast cancer that are being studied for the development of targeted drug therapies, especially in the context of endocrine and anti-HER2 therapy resistance. For example, the PIK3/AKT signalling pathway is one of the most frequently altered pathways in breast cancer with genomic alterations affecting the pathway observed in approximately 70% of cases [226]. This signalling cascade plays a major role in driving tumour cell growth, survival and motility and has been associated with resistance to

endocrine therapy, anti-HER2 therapy and chemotherapy [227, 228]. *PIK3CA* mutations are found in around 40% of all ER+, HER2- patients [229] and the PI3Kα-specific inhibitor alpelisib has been developed, with early trials showing effective anti-tumour activity in heavily pre-treated endocrine therapy resistant patients with advanced disease [230]. In addition, the AKT inhibitor, capivasertib is in clinical trials and has also shown clinically meaningful benefit in combination with fulvestrant in ER+ HER2- locally advanced or metastatic disease in patients who have progressed on endocrine therapy alone [231]. Whilst these trials did not report outcomes specifically in ILC patients, genomic alterations affecting this pathway are higher in ILC compared to IC-NST [32], and capivasertib has been shown to induce a dose-dependent inhibition of growth and survival of human and mouse ILC cell lines [232]. ILC patients would therefore theoretically appear to be ideal candidates for these targeted treatments.

Inhibitors of mTOR are also being used to treat a range of cancers, including subsets of breast cancer patients. Everolimus targets the mTORC1 complex, which is frequently dysregulated in endocrine therapy resistant breast cancers. When used in combination with the AI exemestane, it has shown substantial clinical benefit among postmenopausal women with ER+/HER2- metastatic disease who have progressed on AIs alone based upon the results of the phase III BOLERO-2 trial [233]. When ILC patients (n = 144) were studied as a subgroup within the trial, patients still demonstrated a significant benefit from the addition of everolimus, with outcomes similar to those of the IC-NST patients [234].

1.8.8 Immunotherapies

One of the hallmarks of cancer is its ability to evade an effective anti-tumour host immune response [235]. Cancer immunotherapy refers to approaches which modify the host's immune system and stimulate components of the immune system to generate an effective anti-tumour immune response [236]. The advent of immunotherapies, in particular immune-checkpoint inhibitors (ICIs) has revolutionised the clinical management of a range of solid malignancies.

ICIs block immunosuppressive receptors e.g. cytotoxic T-lymphocyte–associated protein 4 (CTLA-4), programmed cell death 1 (PD-1) and its corresponding ligand, programmed death-ligand 1 (PD-L1) [237]. PD-1 is expressed on the surface of a range immune cell subsets, including T and B lymphocytes, natural killer cells, macrophages, and subsets of dendritic cells [238, 239]. The expression of PD-1 on naïve T cells is induced upon activation of the T-cell receptor (TCR) [240]. PD-1 expression therefore diminishes in the absence of TCR signalling however it is maintained upon chronic activation from a specific persistent epitope target, for example tumour neoantigen, and in the context of chronic viral infection [241]. PD-L1, is

expressed on tumour cells as well as immune cells [238]. The PD-1/PD-L1 pathway together with a subset of immunosuppressive FOXP3+ CD4+ T cells called regulatory T cells, are both of critical importance in terminating immune responses. They help defend against self-reactive effector T cells which could be pathogenic and thus are key in preventing autoimmunity, ensuring that the threshold for T-cell activation is high enough to prevent this [238]. However, this pathway is hijacked by tumour cells and the binding of PD-L1 expressed on tumour cells or infiltrating immune cells to PD-1 on T cells impairs TCR signalling as well as CD28 co-stimulation of T-cell activation, resulting in dampening down of an appropriate anti-tumour immune response and tumour escape [238, 242, 243].

By blocking immunosuppressive receptors and ligands ICIs help prevent these interactions thus preventing tumour-mediated immune inhibition, promoting an inflammatory immune microenvironment, and maintaining anti-tumour T-cell driven immune response. ICIs which have been approved include the anti-PD-1 antibodies such as pembrolizumab and nivolumab, anti-PD-L1 antibodies for example avelumab and atezolizumab and the anti–CTLA-4 antibodies ipilimumab and tremelimumab [238]. In breast cancer, the strongest data to support the use of ICIs exists for TNBC patients, who account for just 1 – 2% of ILCs and are often pleomorphic ILCs [115, 116].

Early trials (KEYNOTE-012 and KEYNOTE-086) assessed the use of pembrolizumab given as monotherapy among metastatic TNBC patients and showed promising results, especially in the first-line setting and among patients with high PD-1 or PD-L1 tumour expression [244-246]. Further trials assessed the use of ICIs given as a combination with traditional chemotherapy in order boost the host immune response. KEYNOTE-355 evaluated the use of the PD-1 inhibitor pembrolizumab in combination with chemotherapy in patients with advanced TNBC [247]. Results showed an improved PFS and OS among patients whose tumours expressed PD-L1 with a combined positive score (CPS: total number of PD-L1+ tumour cells, lymphocytes and macrophages divided by the number of all viable tumour cells x 100) of greater than or equal to 10 [247]. Based on this the combination was approved in 2020 as first-line treatment for metastatic TNBC patients with CPS ≥ 10 [248]. In early-stage TNBC, based on the results of the large phase III KEYNOTE-522 trial, the addition of pembrolizumab to standard chemotherapy was approved for neoadjuvant treatment of previously untreated patients with stage II/III disease, regardless of PD-L1 status [249].

In HER2+ patients, numerous early phase trials have assessed the use of ICIs given in combination with trastuzumab or trastuzumab-based antibody-drug conjugates in patients with metastatic disease [250-252]. They reported responses mainly in patients with PD-L1+

tumours. IMpassion050, a phase III trial, evaluated the use of chemotherapy in combination with HER2-targeted antibodies with or without the PD-L1 inhibitor atezolizumab in 454 HER2+ patients with localised disease [253]. Overall, the addition of atezolizumab appeared to have no effect on pCR rates even among the PD-L1+ population. An evaluation of event-free survival (EFS) is ongoing [253]. ASTEFANIA [254] and APTneo [255] are two ongoing phase III trials assessing the use of ICIs the HER2+ breast cancer.

In the ER+/HER2- disease setting, which encompasses the vast majority of ILCs, the phase Ib trial KEYNOTE-028, assessed the use of pembrolizumab given as a single agent to 25 heavily pre-treated patients with ER+/HER2- metastatic breast cancer [256]. One of the inclusion criteria was that the level of PD-L1 expression within the tumour immune microenvironment as assessed by the CPS had to be greater than or equal to 1. An overall response rate (ORR) of 12.0% was reported and the median response duration was 12 months. Twenty percent of participants experienced immune-related adverse events which were mainly grade 1 - 2. Ongoing phase III trials assessing the efficacy of ICIs in the ER+ breast cancer population include CHECKMATE 7FL, KEYNOTE-B49 and KEYNOTE-756 [257-259]. Overall reduced immunogenicity and low levels of infiltrating T lymphocytes within the tumour microenvironment present a challenge to the potential success of ICIs in breast cancer, especially in ER+ disease which is characterised by low levels of infiltrating immune cells and as it stands, ICIs have not received approval for the treatment of ER+ breast cancers. Importantly these trials have not reported response rates separately for ILC and IC-NST. The GELATO trial (NCT03147040) is the first clinical immunotherapy trial conducted exclusively in metastatic ILC patients [260] and will be discussed further in section 1.9.

Throughout the course of these studies a range of biomarkers have been under investigation to effectively predict ICI response. When PD-L1 is considered, phase III trials assessing the uses of ICIs in early-stage TNBC, showed an increase in pCR rates irrespective of PD-L1 status, yet PD-L1 positivity was associated with higher pCR rates [249, 261]. However in the metastatic setting, responses to atezolizumab and pembrolizumab have been shown to depend upon the level of PD-L1 expression [247, 262]. Whilst high levels of tumour infiltrating lymphocytes (TILs) have been shown to be predictive of response to neoadjuvant chemotherapy in TNBC and HER2+ breast cancer, their predictive role regarding response to ICIs is less clear and no definitive conclusions have been drawn on their value in this setting [263].

Additional biomarkers of interest to predict ICI response across a range of solid tumours include the level of tumour mutational burden (TMB). This is defined as the number of nonsynonymous mutations per megabase of DNA and it is used as a proxy measure of

neoantigen burden [264]. Tumours with a high TMB have been associated with improved clinical responses to ICIs. Pembrolizumab has been approved by the FDA as a treatment for all solid tumours with a TMB ≥ 10 mutations/megabase assessed using a specific assay (FoundationOne CDx assay). This was based upon results of the KEYNOTE-158 study [265]. This was a single-arm, phase II multi-cohort study evaluating the use of pembrolizumab for previously treated unresectable or metastatic non-colorectal solid tumours which were deficient in DNA mismatch repair (dMMR) therefore having high microsatellite instability (MSI-H) and harbouring thousands of somatic mutations encoding potential neoantigens [265]. Breast tumours however were not included in the study.

Whilst high TMB has shown promise as a predictive marker in cancers such as non-small cell lung cancer [266], and malignant melanoma [267], its applicability in breast cancer is somewhat limited by the fact that breast cancers are generally characterised by a low TMB, especially ER+ disease, and overall there is limited data on its predictive and prognostic role in this setting. There is also extensive ongoing work to identify specific immune cell gene expression signatures associated with response to ICIs including a 27-gene RT-qPCR immuno-oncology gene expression assay [268] designed to identify high-risk patients with early breast cancer who may derive benefit from ICIs.

1.9 Clinical trials in ILC

Historically there has been a paucity of clinical trials focusing on ILC. However there are currently several ongoing clinical trials specifically in ILC. These are in both the early-stage and metastatic settings (Table 4).

A landmark study identified synthetic lethality between E-cadherin deficiency and the inhibition of the tyrosine kinase ROS1 [269], which led to the ROLO non-randomized phase II clinical trial (NCT03620643), assessing the ROS1 inhibitor crizotinib in combination with fulvestrant in patients with metastatic E-cadherin defective ILC [270]. Currently, crizotinib has been approved for the treatment of *ROS1* mutated non-small cell lung cancer. The primary study endpoints are response rate as assessed by RECIST criteria as well as safety/tolerability [270].

A further trial, the ROSALINE trial (NCT04551495), is assessing ROS1 inhibition together with endocrine therapy among early-stage ILC patients [271]. ROSALINE is a non-randomized, single-arm neoadjuvant trial assessing the combination of the AI letrozole and entrectinib, a

small molecule which inhibits both ROS-1 and ALK. The trial involves 4 months of the treatment combination in patients with early-stage disease, followed by surgery and the primary endpoint is residual cancer burden [271].

An additional trial including early-stage ILC patients is the 'Endocrine Response in Women with Invasive Lobular Carcinoma' trial (NCT02206984) [272]. Previous studies have demonstrated that adjuvant endocrine therapy results in significant improvements in PFS and OS among ER+ breast cancer patients with early-stage disease. Among postmenopausal women and a subset of high-risk premenopausal patients who are treated using a combination of endocrine therapy with ovarian suppression, AIs have been shown to result in superior DFS compared to tamoxifen [33, 180]. In ILC the differential benefit of AIs over tamoxifen has been shown to be greater in ILC compared to IC-NST suggesting that endocrine therapy approaches have differential efficacy in ILC and IC-NST [33]. The 'Endocrine Response in Women with Invasive Lobular Carcinoma' trial (NCT02206984) is a window trial assessing 3 alternative endocrine therapy approaches in early-stage postmenopausal women prior to surgery. Patients are randomized to receive either the AI anastrozole, tamoxifen or fulvestrant for 21 days prior to surgery. The primary endpoint is change in the level of expression of Ki67.

In ER+, HER2- metastatic breast cancer, CDK4/6 inhibitors in combination with endocrine therapy have been shown to result in significant improvements in DFS and OS. The efficacy of CDK4/6 inhibitors in early-stage ER+ disease is not fully understood, although early reports from the phase III monarchE trial show that 2 years of treatment with adjuvant abemaciclib in high-risk lymph-node positive patients, results in significant improvements in DFS [202]. PELOPS is an open-label phase II trial (NCT02764541) assessing the efficacy of neoadjuvant palbociclib combined with endocrine therapy for the treatment ER+ early-stage breast cancer [273]. It involves an initial 'window phase' which assesses whether letrozole or tamoxifen is more effective in the treatment of ILC, before the 'treatment phase' whereby participants are randomised to receive endocrine treatment +/- palbociclib, and the primary endpoint is pCR.

In the advanced disease setting, the GELATO trial (NCT03147040) is a non-randomized single arm, phase II trial among ILC patients with metastatic disease [260]. It is the first clinical immunotherapy trial conducted exclusively in metastatic ILC patients. In the study participants who have endocrine therapy resistant disease (in ER+ cases) and who have had a maximum of 2 lines of previous palliative chemotherapy, receive a combination of carboplatin, an alkylating agent (12 cycles) and atezolizumab, a PD-L1 inhibitor (starting from the 3rd cycle of carboplatin) until disease progression or intolerability. The primary endpoint is PFS at the 6-month time point, as assessed by RECIST [260]. Since the majority of ILCs are ER+ and

generally characterised by a low level of immune infiltrate, such features would predict poor response rates. First results of the trial show that of 23 evaluable patients, 4 were progression-free at 6 months therefore meeting the first-stage primary endpoint. One patient had an ongoing response and 2 patients had commenced therapy but not yet reached the endpoint. Four of the 21 patients showed partial responses. In addition, 2 patients had stable disease, producing a clinical benefit rate of 29%. Interestingly when hormone-receptor status was considered, 4 of the patients receiving clinical benefit had triple-negative disease, whilst only 5 of the original 23 ILC patients had triple-negative disease. Moreover, the level of stromal TILs did not appear to be associated with clinical benefit. Overall, these first results suggest an efficacy of PD-L1 inhibition in metastatic ILC, predominantly in those patients with triple-negative disease.

As previously described, patients with ILC can harbour *HER2* alterations including mutations and amplifications which are rare in classic ILC and more prevalent in the pleomorphic subtype. SUMMIT (NCT019539926) is a phase II basket study, for patients who have advanced solid cancers harbouring somatic *HER2* mutations (without *HER2* amplification). In the ER+ breast cancer arm, participants with previous CDK4/6 inhibitor therapy are randomly allocated to receive either i) neratinib (a tyrosine kinase inhibitor), trastuzumab (a monoclonal anti-HER2 antibody) and fulvestrant, or ii) trastuzumab and fulvestrant or iii) fulvestrant alone. Among ER+ patients without previous CDK4/6 inhibitor therapy, patients receive neratinib, trastuzumab and fulvestrant [274]. ORR measured by RECIST is the primary endpoint [274]. Whilst the study includes all ER+ breast cancers, an early report demonstrates that almost half of breast cancer patients included in the study show lobular histology, consistent with the higher frequency of *HER2* mutations observed in metastatic ILC compared to IC-NST. An early report showed that of 13 heavily pre-treated ER+ breast cancer patients receiving the combination of neratinib, trastuzumab and fulvestrant, 5 patients had a partial response resulting in an ORR of 39%. The combination was well-tolerated and shows a promising ORR and based on this the cohort has been therefore expanded to include 50 patients. The results are awaited [275].

Trial Name	Phase	Trial Description
ROSALINE trial (NCT04551495)	II	This trial assesses the efficacy of the ROS1 inhibitor, entrectinib, as a treatment combination with letrozole, as neoadjuvant therapy in earlier stage(1-3) ILC.
'Endocrine response in women with invasive lobular carcinoma' trial (NCT02206984)	II	This trial seeks to optimise and refine adjuvant endocrine therapy specifically in ILC in postmenopausal women with ER+ve disease. Patients receive one of three anti-oestrogen neoadjuvant treatments (anastrozole, fulvestrant or tamoxifen) and relative changes in Ki67 IHC staining are used as a marker of response.
GELATO trial (NCT03147040)	II	This trial investigates responses to immunotherapy in ILC, in the form of a PD-L1 inhibitor; atezolizumab in combination with chemotherapy (carboplatin) in metastatic ILC.
PELOPS trial (NCT02764541)	II	This trial assesses the effect of neoadjuvant palbociclib with endocrine therapy in ER+ve early stage breast cancer. It includes an initial 'window phase' aimed at determining whether letrozole or tamoxifen is more effective in ILC, prior to the 'treatment phase'.
SUMMIT (NCT019539926)	II	This trial is for patients with advanced solid cancers harbouring somatic HER2 mutations, and assesses responses to anti-HER2 therapy. ER+ breast cancer patients (of which ILC patients account for almost a half) are randomly allocated to receive either i) neratinib, trastuzumab and fulvestrant, ii) trastuzumab and fulvestrant or iii) fulvestrant alone.

Table 1.3: Summary of ongoing clinical trials in ILC [260, 271-274]

1.10 Molecular Characteristics of ILC

Three large studies have comprehensively characterised ILC at the molecular level and provided new insights into the multi-omic landscape of ILC and its differences from IC-NST [32, 276, 277].

1.10.1 Genomic Features of ILC

1.10.1.1 E-cadherin

The hallmark feature of ILC is loss of E-cadherin expression, observed in 85 - 95% of cases [5]. E-cadherin is a calcium-dependent transmembrane glycoprotein that mediates cell to cell adhesion in epithelial tissues. Its loss is responsible for the discohesive morphology of ILC. Loss of expression of E-cadherin can result from mutation of the *CDH1* gene and heterozygous loss of chromosome 16q (where *CDH1* is located), resulting in complete loss of the protein. Loss of function mutations in *CDH1* are found in 50 - 65% of ILCs [32, 276]. They are considered to be an early event, often being identified in accompanying LCIS [1]. In addition, the dysregulated expression of catenin-binding proteins (α, β, γ and p120-catenin), which are responsible for anchoring E-cadherin to the membrane and actin cytoskeleton, is another mechanism for E-cadherin loss in ILC [89, 278]. Previous reports have described promoter hypermethylation and resultant down-regulation of *CDH1* expression in 21 – 77% of ILC cases [85, 279]. However, the epigenetic silencing of *CDH1* in ILC was not supported by a more recent molecular characterisation of 127 ILCs from the TCGA dataset which found that reduced expression of E-cadherin was not associated with DNA promotor hypermethylation at *CDH1* [32]. It was also not supported by a recent study which assessed the molecular heterogeneity of E-cadherin expression in ILC, performing whole genome sequencing and methylation profiling in a cohort of cases, which found no *CDH1* promotor methylation [280]. Deleterious germline mutations affecting *CDH1* have been shown to be causative of hereditary diffuse gastric cancer (HDGC) and ILC where they confer a risk of ILC development of up to 42% [281].

1.10.1.2 Additional Mutations and Copy-number alterations

After *CDH1*, gain of function mutations in *PIK3CA* are the second most frequently observed mutation found in a reported 34.8% - 48% of ILC cases [32, 277]. In fact, mutations in phosphatidylinositol 3-kinase (*PIK3CA)*, phosphatase and tensin homologue (*PTEN*), or protein kinase B (*AKT1*) have been found in over half of ILCs and at a higher frequency to molecular subtype-matched IC-NST, resulting in an enriched phosphatidylinositol 3-kinase pathway in ILC [32, 282]. The majority of ILCs are of the luminal A molecular subtype, and therefore a direct comparison of luminal A ILC and luminal A IC-NST was performed using samples from the TCGA dataset to identify ILC-specific features (Figure 1.4) [32]. ILC was enriched for inactivating *PTEN* alterations (14% ILC vs. 3% IC-NST), runt-related transcription factor 1 (*RUNX1*) mutations (10% ILC vs. 3% IC-NST) and forkhead box protein A1 (*FOXA1*) mutations (7% ILC vs. 2% IC-NST) [32]. *FOXA1* mutations were also reported at a similar rate of 9% in an additional large ILC cohort [276]. *PTEN* loss was associated with increased AKT phosphorylation, and this was highly activated in ILC. ILC was also distinguished from IC-NST by its lower frequency of *GATA3* mutations (5% ILC vs. 20% IC-NST, Figure 1.4) [32]. *GATA3* and *FOXA1* are both important transcriptional regulators of ER activity [283, 284]. These findings therefore suggest a preferential requirement for distinct modulators of ER activity in ILC and IC-NST and the mutually exclusive roles of *FOXA1* and *GATA3* in the development of the two distinct disease entities.

Of clinical significance, actionable mutations have been reported in ILC with *HER2* and *HER3* mutations identified in a cohort of 170 ILCs (including several different ILC histological subtypes) at rates of 5.1% and 3.6% respectively [276]. These mutations were enriched in the 'mixed non-classic' histological subtype although of note a 'pleomorphic' category was not included and thus pleomorphic ILCs would have been included in the mixed non-classic group. 'Mixed non-classic' is not currently used as an ILC histological subtype. In addition, *AKT1* mutations were identified in 4.1% of cases presenting an additional potential therapeutic target in a small proportion of ILCs [276]. Interestingly patients harbouring *HER2* and *AKT1* mutations showed an increased risk of early disease recurrence [276] and for *HER2* mutations, worse overall prognosis [285]. This further highlights the need to target these aberrations to help improve clinical outcomes for these patients. As previously described, higher rates of HER2 alterations are reported in pleomorphic ILCs. A lobular specific prognostic gene expression signature has identified that as well as *HER2* and *AKT1* mutations, *HER3*, *TP53* and *ROS1* mutations in ILC primary tumours are also associated with poor prognosis [286].

With regards to copy number alterations (CNAs), as well as loss of chromosome 16q, deletions of 8p23-p21, 11q14.1-q25, 13q, gains of chromosome 1q, 8q, 11p, 16p and high-level amplifications at 1q32, 8p12 (*FGFR1* locus) and 11q13 (*CCND1* locus) are also genomic features widely reported in ILC [1, 32, 276]. Interestingly 1q gains have been associated with a better clinical outcome, whilst 17q12 and 11p gains have been associated with worse outcome [276].

Table 4 summarises the reported rates of somatic mutations in LCIS, ILC and metastatic ILC from various studies. Of note, the rates of mutations in *TP53*, *HER2* and *ESR1*, appear to increase with increasing disease severity in the *in situ*, invasive and metastatic settings, thus highlighting the role of these mutations in driving an aggressive tumour biology. There is however an ongoing need to identify additional genomic alterations present in the primary tumours of ILC patients which may be predictive of aggressive tumour behaviour and early relapse. This would enable the identification at outset of those patients requiring additional treatments to prevent, or at least delay the onset of disease metastases.

		1	2	3	4	5	6	7	8	9		10		11
		CLCIS (*n* = 43)	PLCIS (17) FLCIS (2) (*n* = 19)	PLCIS (10) FLCIS (6) (*n* = 16)	ILC (*n* = 127)	ILC (*n* = 144)	ILC (*n* = 413)	PILC (*n* = 27)	PILC (*n* = 17)	Primary ILC (*n* = 32)	Metastatic ILC (*n* = 32)	Primary ILC (*n* = 127)	Metastatic ILC (*n* = 132)	Metastatic ILC (*n* = 180)*
Mutations	*CDH1*	81.0%	94.7%	94.0%	63.0%	42.7%	65.4%	89.0%	59.8%	53%	62.5%	82%	76%	77%
	PIK3CA	32.0%	31.6%	56.2%	48.0%	33.8%	43.3%	33.0%	52.9%	44%	44%	57%	52%	52.9%
	ERBB2	7.0%	68.4%	37.5%	4.0%	4.3%	5.1%	26.0%	17.6%	12.5%	15.6%	2%	12%	8.3%
	ERBB3	NR	21.1%	18.8%	NR	2.9%	3.6%	NR	23.5%	3%	0%	NR	NR	< 5%
	RUNX1	2.3%	21.1%	NR	10.0%	NR	3.4%	NR	11.8%	3%	6%	9%	5%	~ 7%
	CBFB	19.0%	15.8%	12.5%	2.0%	NR	NR	NR	17.6%	NR	NR	NR	NR	NR
	FOXA1	4.7%	10.5%	31.3%	7.0%	NR	9.0%	NR	5.9%	15.6%	15.6%	8%	11%	NR
	GATA3	4.7%	10.5%	12.5%	5.0%	5.1%	7.3%	7.0%	5.9%	15.6%	15.6%	3%	7%	2.2%
	ARID1A	2.3%	10.5%	6.3%	17.0%	7.0%	6.3%	15.0%	5.9%	12.5%	12.5%	8%	11%	~ 12%
	TP53	0%	5.3%	18.8%	8.0%	3.6%	7.3%	19.0%	11.8%	18.7%	9%	9%	20%	23.9%
	TBX3	9.3%	NR	NR	9.0%	8.0%	13.3%	7.0%	23.5%	21.8%	18.7%	10%	16%	12.8%
	KMT2C	2.3%	NR	NR	7.0%	10.0%	8.0%	19.0%	35.3%	NR	NR	NR	NR	< 5%
	MAP3K1	4.7%	NR	NR	6.0%	5.1%	5.0%	19.0%	35.3%	18.7%	15.6%	10%	10%	< 5%
	ESR1	NR	NR	NR	NR	1.4%	NR	NR	5.9%	12.5%	15.6%	2%	15%	17%
	AKT1	4.7%	NR	NR	2.4%	5.1%	4.1%	7.0%	5.9%	6.2%	9.4%	NR	NR	~ 5%
	PTEN	0%	NR	6.2%	7.0%	1.4%	3.9%	NR	0%	3.1%	0%	9%	9%	~ 10%
	NF1	NR	NR	0%	NR	4.3%	1.0%	7.0%	24%	3.1%	6.2%	2%	8%	12.2
Amp	*CCND1*	NR	26.3%	18.8%	17.0%	15.0%	38.0%	11.0%	12.0%	33.3%	38.0%	17%	17%	22%
	ERBB2	4.7%	31.5%	12.5%	7.0%	4.0%	0.0%	4.0%	6.0%	33.3%	15.6%	2%	5%	NR
	FGFR1	0%	0%	NR	NR	NR	25.3%	7.0%	23.5%	6.2%	6.2%	7%	11%	7.8%

Table 1.4: Somatic mutations in LCIS, ILC and metastatic lobular carcinoma- adapted from [36]. 1: Lee et al [287], 2: Harrison et al [288], 3: Shamir et al [289], 4: Ciriello et al [32], 5: Michaut et al [277], 6: Desmedt et al [276], 7: Rosa-Rosa et al [101], 8: Zhu et al [103], 9: Richard et al [290], 10: Pareja et al [291], 11: Sokol et al [292]

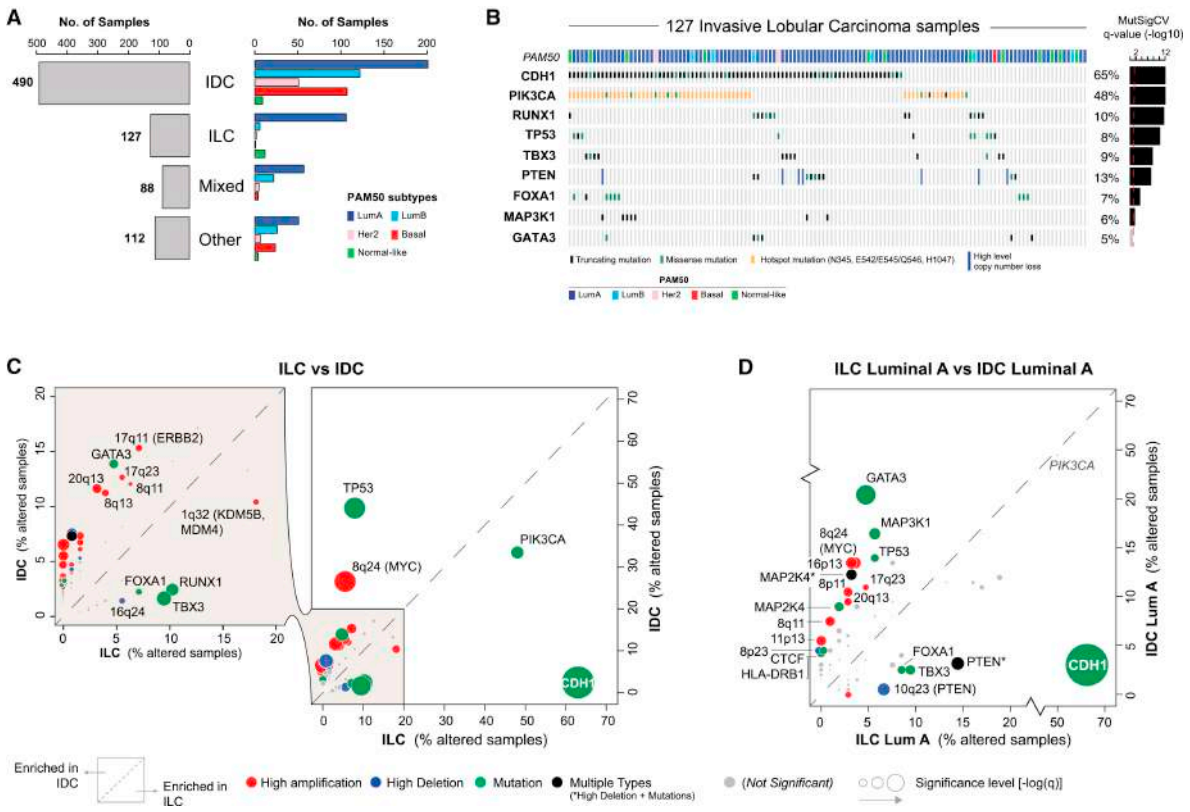


Figure 1.4: Molecular determinants of ILC from the TCGA dataset [32]
A) Histopathological breast cancer subtypes: invasive breast cancer of no special type (IC-NST), invasive lobular (ILC), mixed ductal-lobular (Mixed), and other-type (Other) carcinoma
B) Recurrently mutated genes (MutSigCV2) in ILC
C) Comparison of the alteration frequency for 153 recurrent genomic alterations in ILC versus IC-NST
D) Comparison of the alteration frequency for 153 recurrent genomic alterations in ILC Luminal A versus IC-NST Luminal A

1.10.2 Genomic Features of Pleomorphic ILC

Pleomorphic ILC represents a molecularly heterogenous subgroup of ILC. Studies show higher rates of mutations in *TP53*, *KMT2C*, *ISR2, MAP3K1*, *NCOR1*, *NF1*, *TBX3*, *ARID1A* and *ARID1B* in pleomorphic ILC compared to classic ILC [101, 103, 293]. Indeed, *TP53* and *ISR2* mutations, loss of ARID1A expression and amplification of *PIK3CA* and *CCND1* have been associated with progression from *in situ* to invasive disease and/or lymph node metastases [101, 103]. In addition, *HER2* alterations are also more frequent in pleomorphic ILC. Analysis of the genomic features of 31 pleomorphic ILC from TCGA showed that 32% had activating *HER2* mutations and 19% had *HER2* amplifications [100]. Rates of amplification/overexpression have reported in up to 33% of cases in other small pleomorphic ILC cohorts [100-103]. In contrast *HER2* alterations are rare in classic ILC with no cases having activating *HER2* mutations in the TCGA cohort and just one case showing *HER2* amplification (Figure 1.5) [100]. This is clinically relevant as *HER2* alterations can be targeted therapeutically suggesting subsets of pleomorphic ILC patients may derive benefit from targeted adjuvant anti-HER2 treatment. Moreover, *HER2* mutation analysis may be important in identifying *HER2* mutated pleomorphic ILC patients who do not show *HER2* amplification who could also benefit from such treatment. For example one study showed that 6 of 39 (15.4%) grade 3 HER2- tumours in fact harboured at least one activating *HER2* mutation [294]. Further clinical trials are required to better understand how and when to therapeutically target the HER2 pathway in relevant ILC patient subsets, particularly pleomorphic ILC.

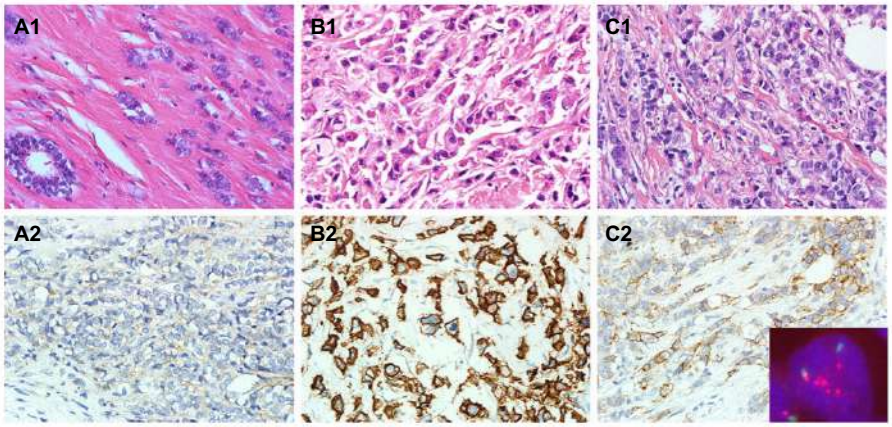


Figure 1.5: *HER2* alterations in classic vs pleomorphic ILC: Histology, IHC and FISH in classic and pleomorphic lobular carcinomas A1) Histology and A2) corresponding IHC staining for HER2 in classic ILC case showing negative HER2 staining on IHC B1) Histology and B2) corresponding IHC staining for HER2 in a pleomorphic ILC case showing strongly positive HER2 staining C1) Histology and C2) corresponding IHC staining for HER2 in a pleomorphic ILC case showing equivocal HER2 staining but *HER2* FISH demonstrated amplification. Taken from [32]

1.10.3 Genomic Features of Other ILC Subtypes

A range of mutations and CNAs have also been associated with other ILC subtypes. An enrichment of *HER2*, *TP53*, and *ARID1A* mutations was observed in solid ILC together with 11p and 6q25.1 (*ESR1*) gains and 1p36.22 (*ARID1A*) deletions [276]. The alveolar subtype was associated with 11q13.3 (*CCND1*) and 11q14 (*PAK1*) gains and the 'mixed non-classic' type was associated with *HER2* and *TP53* mutations [276].

Based on the identification that specific molecular alterations are associated with the various subtypes of ILC, a multi-step model of the evolution of classic ILC and its morphological variants has been proposed which suggests that these commonly arise on a background of normal epithelium through the acquisition of various genomic alterations (Figure 1.6) [1].

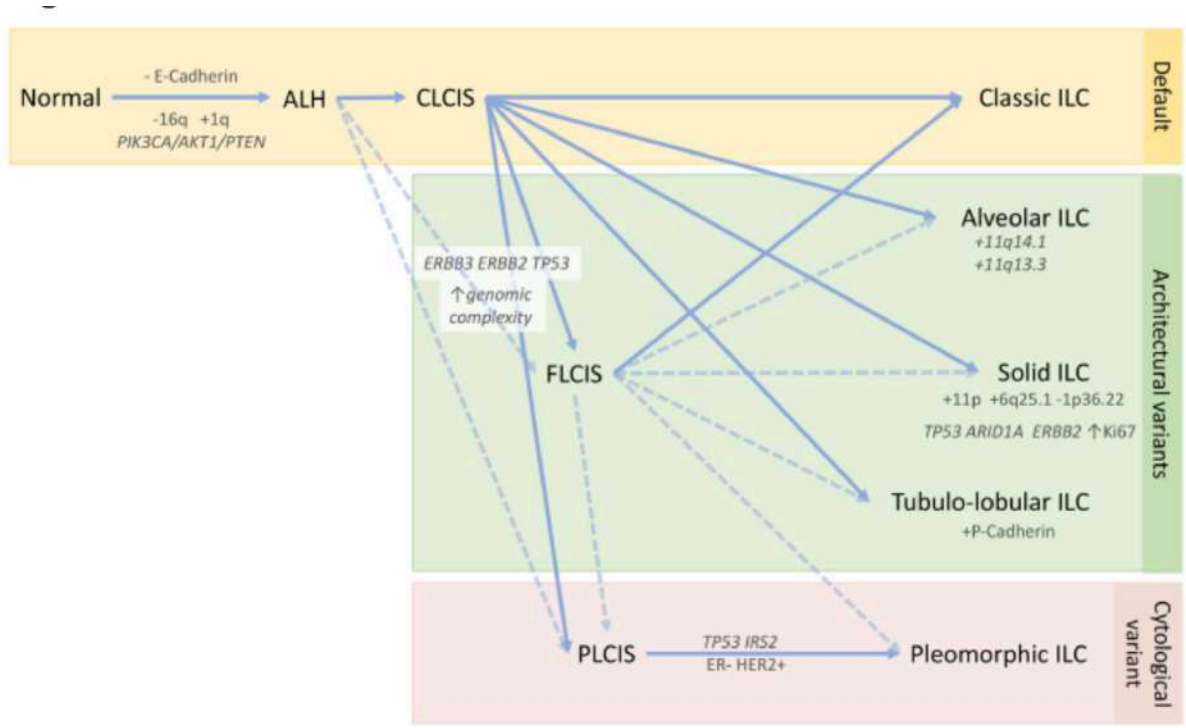


Figure 1.6: Multistep model of the evolution of classic ILC and its morphological variants (ALH = Atypical Lobular hyperplasia, CLCIS – Classic lobular carcinoma in situ, PLCIS = Pleomorphic in situ, FLCIS = Florid in situ). Taken from [1]

1.10.4 Mutational drivers of endocrine resistance and metastases in ILC

Since the occurrence of *de novo* metastatic ILC is rare and a large proportion of patients even with early-stage tumours receive subsequent postoperative treatment in the form of hormonal therapy and/or chemotherapy, distinguishing between metastasis-specific and therapy-induced mutations is challenging since they go hand in hand.

1.10.4.1 *ESR1*

ESR1 mutations have been associated with resistance to endocrine therapy and disease relapse in breast cancer and a recent study compared the frequency of *ESR1* mutations between IC-NST and ILC patients using metastatic samples from MSKCC-IMPACT (n = 595 IC-NST and n = 116 ILC) and circulating tumour DNA from the SoFEA and PALOMA-3 trials (n = 416 IC-NST and n = 76 ILC). This found no differences in the prevalence and distribution of *ESR1* mutations between IC-NST and ILC [295]. Interestingly a separate study assessed CNAs in a cohort of 70 primary ILCs (n = 70) and identified gains and amplifications in 14% and 10% of cases respectively and these alterations were significantly associated with subsequent disease recurrence [296]. When metastatic ILC lesions have been assessed, a significant enrichment of *ESR1* copy number gains has been identified in bone metastatic deposits [297].

1.10.4.2 Additional Genomic Alterations

Hybrid-capture-based genomic profiling of 180 ILC and 191 ER+ IC-NST metastatic biopsies identified an enrichment of *ESR1* alterations in both ILC and IC-NST metastases in relation to the primary breast specimens, yet *NF1* alterations were only enriched in the ILC metastases [292]. Interestingly there was also a significant increase in TMB in metastatic ILC biopsies compared to both primary ILCs and metastatic IC-NSTs [292]. A further retrospective study assembled a multicentric series of matched primary and metastatic samples from 94 ER+ ILC patients and performed targeted sequencing and whole-genome sequencing to identify mutations and CNAs. Matched primary tumour and metastasis analysis identified mutations (*AKT1, ARID1A, ESR1, ERBB2*, or *NF1*) as well as CNAs (*NF1* deletions, *PTEN* deletions, *CYP19A1* amplifications) which were unique to the metastases in 22% (7/32) and 19% (4/21) of cases, respectively. This reflects the molecular heterogeneity and clonal evolution in ILC, with distinct mutational repertoires in the primary tumour and metastatic lesions.

An assessment of the genomic landscape of 692 advanced ER+ breast cancers, which included post-treatment biopsies in the neoadjuvant and metastatic settings was carried out using the MSK-IMPACT targeted gene panel. This identified an enrichment of not only *ESR1* but also *HER2* and *NF1* mutations in post hormone treatment vs pre-treatment ILC samples [173]. Hotspot mutations in *HER2* as well as loss of function mutations of *NF1* were over twice as common in post-treatment samples. Of note, *ESR1* mutations were mutually exclusive with *ERBB2* and *NF1* mutations [173]. Based upon the genomic alterations identified in refractory lesions, the endocrine therapy resistant tumours were grouped into 4 categories: those harbouring *ESR1* mutations (accounting for 18%), alterations affecting the MAPK signalling pathway (affecting 13%) and mutations involving *MYC* (accounting for 9%). In the remaining 60% of cases the underlying mechanisms of resistance were unknown (Figure 1.7) [173]. This suggests that disease recurrence and endocrine therapy resistance is not fully explained by genomic alterations and that other features, such as transcriptomic heterogeneity may play a key role. Tumour subclones with unique transcriptomic profiles which confer a survival advantage may be critical in driving disease progression under the selective pressure of endocrine therapy.

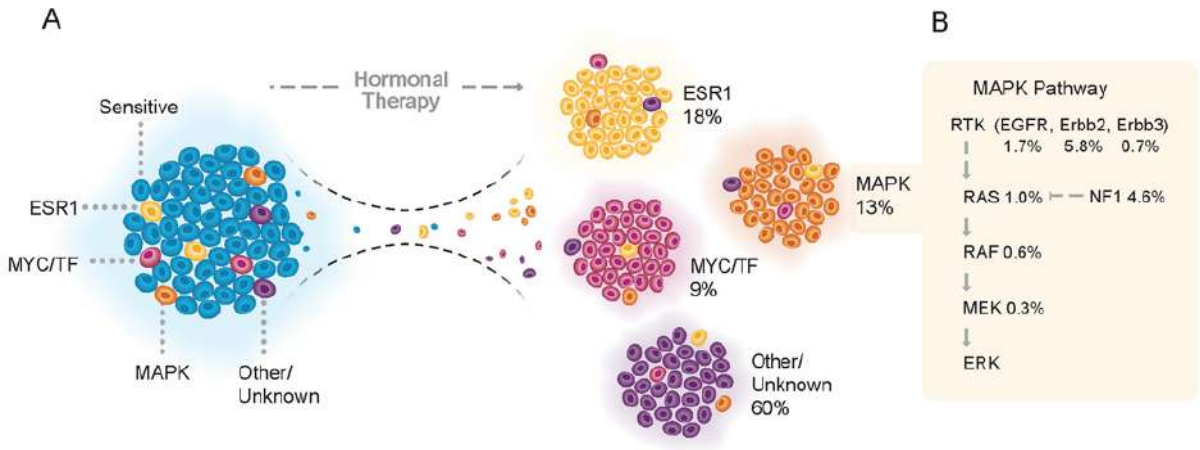


Figure 1.7: Mechanisms of resistance to endocrine therapy in hormone receptor positive breast cancer: A) Genomic alterations affecting ER+ tumours which are resistant to endocrine therapy B) Frequency of alterations affecting the MAPK pathway in tumours following endocrine therapy. Taken from [173]

1.10.5 Transcriptomic Features of ILC

Clustering analyses of mRNA expression levels has been used to identify gene expression subtypes of ILC in two large independent ILC cohorts: the Rational Therapy for Breast Cancer (RATHER) consortia and TCGA (n = 106 and n = 144 respectively) [32, 277]. Two and three gene expression subtypes were identified in these cohorts (RATHER: immune-related and hormone-related subtypes TCGA: immune-related, reactive, and proliferative subtypes). Of note both included an 'immune-related' subtype. This subtype was associated with overexpression of various chemokine and interleukin transcripts, macrophage-associated signalling, and TCR gene expression signatures in the TGCA cohort [32]. In the RATHER cohort the immune-related subtype was also associated with mRNA up-regulation of *PD-L1, PD-1* and *CTLA-4* and these tumours were associated with higher TILs assessed histologically [277]. However, these two separate 'immune' subtyping approaches failed to identify the same ILC cases when they were applied to the same dataset [282]. The RATHER 'hormone-related' subtype was associated with higher levels of oestrogen and progesterone receptors, oestrogen receptor target genes, up-regulation of cell-cycle genes as well as epithelial to mesenchymal transition (EMT) [277]. The TCGA 60-gene classifier identified that a 'reactive-like' subtype was characterised by a strong microenvironment and/or cancer fibroblast signalling, and the 'proliferative' subtype was associated with increased expression of cell-cycle proteins [32].

When these subtypes were correlated with clinical outcome, there was no significant difference in survival outcomes between the two RATHER gene expression subtypes [277]. In the TCGA analysis, the 'reactive-like' subtype was associated with an improved prognosis compared to the 'proliferative' subtype [32]. These studies have highlighted the transcriptomic heterogeneity of ILC.

1.10.6 Molecular Prognostication

In the clinical setting prognostication of breast cancers is completed using standard clinicopathological information most notably with the Nottingham Prognostic Index Plus (NPI+). This prognostic index uses tumour size, tumour grade, lymph node status, in addition to hormone (ER, PR, HER2) status [298]. However, the value of this index specifically within the context of ILC is not clear, since the majority of ILCs are grades 1 or 2, and they are also predominantly ER/PR+ and HER2-. As a result, the majority of ILCs fall within the 'good to moderate' NPI category and there is little to differentiate between ILCs which will have poor outcomes and those with better outcomes, and therefore it would appear to require further refinement for ILC [298].

Several molecular signature-based tests are available commercially yet none of these account for tumour morphology in their algorithms. Their prognostic value in the context of ILC is only starting to emerge with some variable yet promising results. For example, the Genomic Grade Index (GGI/MapQuantDx™) panel has been demonstrated to be superior to grade alone in ILC and reclassed certain pleomorphic ILCs to a lower risk group, whilst upgrading some classic ILC cases to higher groups as a result of their genomic grade [299]. The clinical value of MammaPrint®, used to assess risk of recurrence in early-stage breast cancers, has been shown in node-negative ILC [300]. Prosigna® is another commercially available diagnostic test using the PAM50 intrinsic subtypes to produce a risk of recurrence score which has been shown to provide additional prognostic value in patients with ILC [301]. The EndoPredict test, EPClin [302], has also been shown to be highly prognostic in node positive and negative ILC patients [303]. On the other hand, the clinic value of OncotypeDx®, a 21-gene test used to guide decisions on the potential benefits to be derived from chemotherapy, remains less clear in ILC [304-306].

Recently the first gene signature has been created for specifically prognosticating ILC patients [286]. LobSig is a 194-gene signature which uses *in silico* integrative analysis combining genome copy number and transcriptomic data. It has been shown to outperform the non-ILC specific signatures in univariate and multivariate analysis [286] In particular, it demonstrated significant value in grade 2 ILC cases which are clinically difficult to prognosticate, for example it predicted clinical outcome for patients identified as moderate-risk (using the NPI+ Index) in the METABRIC cohort with 94.6% accuracy. It therefore demonstrates its potential to offer a clinically relevant prognostic tool in ILC.

1.11 The Tumour Microenvironment

A tumour does not simply consist of a collection of malignant cells, but instead represents a heterogeneous network of diverse resident and infiltrating host cells. The tumour cells can stimulate a range of different cellular, molecular, and physical changes in their local tissues [307]. The tumour microenvironment (TME) describes a highly complex and constantly evolving entity. The overall composition of a TME can vary between differing tumours but the hallmark components include immune cells, stromal cells, endothelial cells, as well as extracellular matrix [307]. The TME is not considered to represent a silent bystander in tumorigenesis, but instead can actively promote cancer progression [307].

In the early stage of tumour growth, a highly dynamic and reciprocal interplay develops between tumour cells and various elements of the TME which ultimately supports the survival of tumour cells, local invasion, and resultant metastatic spread [307]. To counteract and overcome hypoxic and acidic conditions, the TME coordinates a strategy which facilitates angiogenesis thus helping to restore an adequate supply of oxygen and nutrients and ensuring the removal of metabolic wastes [307]. Tumours can further become infiltrated with a diverse range of different immune subpopulations which can either promote or hinder tumour progression [307, 308].

Tumour cells also recruit important supporting cells from the surrounding tissue stroma to assist in promoting key steps of solid tumour formation. Stromal composition shows diversity between different tumour types, but key stromal cells include fibroblasts, cancer-associated fibroblasts (CAFs), vascular endothelial cells, adipocytes, and stellate cells [307] (Figure 1.8). Following recruitment to the TME, stromal cells can secrete a range of cytokines and growth factors which influence tumour cell proliferation, angiogenesis, invasion, and metastasis [307].

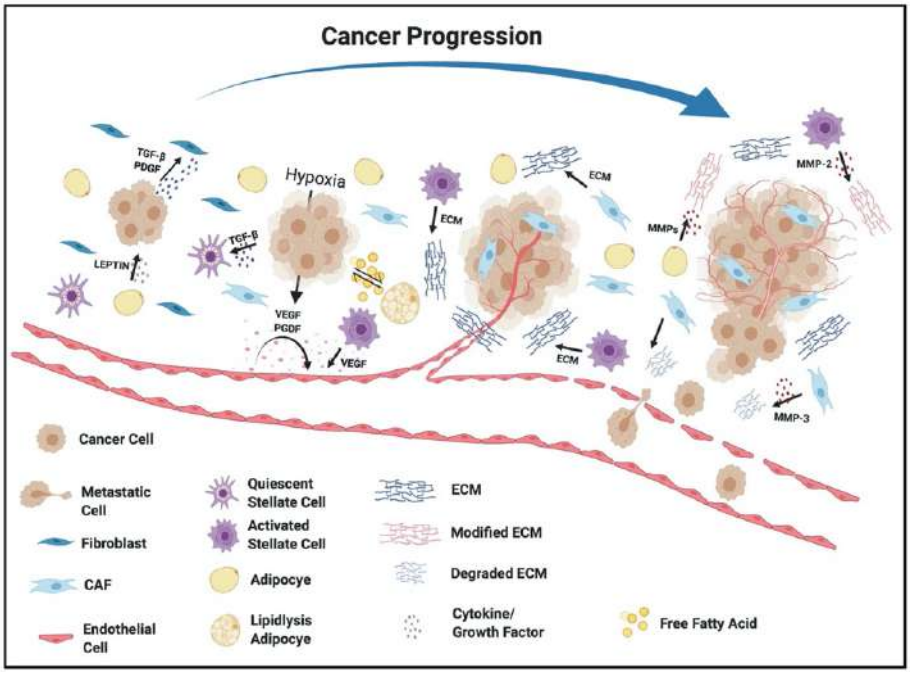


Figure 1.8: The role of stromal cells in promoting tumour progression: Schematic representation summarising the dynamic relationships and interplay between various stromal cells and cancer cells which promote and facilitate tumorigenesis. Image taken from [307]

1.11.1 Cancer-associated fibroblasts (CAFs)

Cancer associated fibroblasts (CAFs) represent a large component of the tumour stroma and have essential roles in facilitating crosstalk between tumour cells and other elements of the TME [307]. They represent a diverse stromal population, commonly arising from resident fibroblasts. CAFs may also arise from adipocytes, pericytes, stellate cells, endothelial cells as well as bone-marrow derived mesenchymal stem cells [307]. Following injury, resident tissue fibroblasts are induced to become myofibroblasts, which play an active role in wound healing for example they develop contractile properties, secretory phenotypes and promote the formation of extracellular matrix [307]. In 1986 tumours were described as 'wounds that do not heal' [309]. Within the TME, tumour cells and some stromal cells actively secrete transforming growth factors e.g. TGF-β, PGDF and FGF2 which result in the conversion of fibroblasts into CAFs. The accumulation of CAFs in the TME has been associated with poor outcomes in a range of cancer types [307].

CAFs can shape the TME by facilitating tumour proliferation and metastasis, promoting angiogenesis, extra-cellular matrix remodelling and by promoting an immunosuppressive

microenvironment. For example, CAFs secrete TGF-β, which is required for epithelial-mesenchymal transition enabling epithelial tumour cells to develop migratory and invasive properties and it also promotes angiogenesis [307]. Moreover the secretion of MMP-3 by CAFs results in the degradation of E-cadherin thus promoting tumour cell invasion [307].

In breast cancer CAFs have been shown to arise from several cell types and promote breast cancer progression through the secretion of various cytokines, generation of exosomes, release of nutrients, reshaping of the extracellular matrix, and through the suppression of the anti-tumour functions of various immune cells [310]. Moreover, CAFs have become therapeutic targets in breast cancer and treatments targeting CAFs are in clinical trials [310]. Interestingly ILC has been shown to have a distinct tumour microenvironment compared to IC-NST and is characterised by a more pronounced growth of CAFs as well as endothelial cells involved in angiogenesis [311]. Furthermore, an in-depth characterisation of human ILC stroma at the transcriptomic level has identified significant differences in the stromal gene expression signatures in ILC vs IC-NST with a range of genes which are involved in extra cellular matrix regulation being more highly expressed in the stromal cells vs tumour epithelial cells in ILC, but not in IC-NST or normal breast tissue [312].

1.12 Heterogeneity of the immune microenvironment in cancer

The tumour immune microenvironment is increasingly recognised to play a key role in modulating tumorigenesis and enabling effective tumour clearance and targeting the immune system using Immunotherapy is revolutionising the management of multiple solid malignancies [313, 314]. The composition of the immune microenvironment and the presence of immune subpopulations involved in the innate and adaptive immune responses varies depending on tumour location and histological subtype. Studies associate most immune subsets with either pro or anti-tumoral functions (Figure 1.9). Mouse models have shown that myeloid lineage leukocytes, such as dendritic cells, myeloid-derived suppressor cells and tumour-associated macrophages (TAMs), play a key role in shaping the nature of the immune microenvironment. Depending on the factors that these cell populations secrete, they can lead to the formation of either an immunostimulatory pro-inflammatory anti-tumoral microenvironment or rather a 'wound healing' immunosuppressive pro-tumoral milieu. T lymphocytes which infiltrate the tumour microenvironment into these opposing settings are therefore either activated or suppressed. Another key immune cell subpopulation, TAMs, can either be polarized towards anti-tumoral M1-like or pro-tumoral M2-like functional phenotypes and this is regulated through their interactions with T cells. The cellular interactions and cross-talk between these various immune subpopulations are key in shaping the nature of the tumour immune microenvironment.

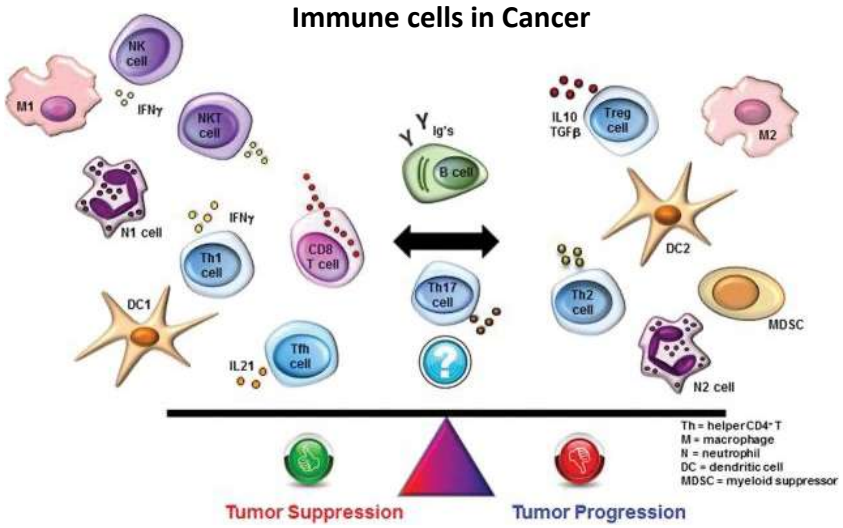


Figure 1.9: Immune cell subpopulations within the tumour immune microenvironment: the immune microenvironment consists of a diverse range of immune subpopulations. The relative proportions of these cells type can tip the balance between tumour suppression and tumour progression - Adapted from [308]

1.12.1 The role of Tumour Infiltrating Lymphocytes (TILs)

Traditionally at the histological level, immunogenicity is assessed by the quantification of lymphocytes. Lymphocytes represent a diverse population of immune cells with a range of functions. There are two main types of lymphocyte: T lymphocytes and B lymphocytes. They are involved in the adaptive/antigen-specific immune response. Morphologically they are indistinguishable and are small cells (8 -10 microns) which possess a large nucleus containing dense hetero-chromatin and a cytoplasmic border with few mitochondria and ribosomes [315]. It is therefore impossible to further characterise them into lymphocyte subpopulations based on H&E alone. Once activated through their interactions with antigenic stimuli, they may enlarge through an increase in their cytoplasm and organelle number [316]. Lymphocytes are responsible for presenting receptors (T-cell receptors and B-cell receptors) for antigenic recognition with a diverse range of different specificities expressed on their surfaces. The genes encoding these receptors undergo a series of DNA recombinations, providing them with huge phenotypic diversity [316]. Humoral immunity otherwise known as antibody-mediated immunity, depends on B cells, which can transform into plasma cells. Plasma cells are responsible for the production of antibodies. B cells also Involved in antibody-presentation and the production of cytokines [317].

Cellular immunity requires T lymphocytes. T lymphocytes play a key role in immune surveillance and the development of the adaptive immune response against infection and cancer [318]. T lymphocytes are classed as either CD4+ or CD8+ cells. $CD8^+$ T cells respond to peptide sequences present on the surface of antigen presenting cells and contribute to an inflammatory immune microenvironment. They are responsible for mediating cell contact-dependent cytotoxicity of tumour cells through the release of various cytotoxins such as perforin and granzymes which trigger the caspase cascade leading to apoptosis as well as through the secretion of cytokines such as interferon γ (IFNγ) and tumour necrosis factor α (TNFα) [319, 320]. Memory CD8+ T cells survive long-term and undergo rapid proliferation and acquisition of effector function upon re-exposure to a specific antigen [321].

In contrast $CD4^+$ T cells represent a highly versatile, polyfunctional cell group that exhibit a diverse repertoire of effector functions and show phenotypic plasticity and heterogeneity subject to various context-specific and microenvironmental factors [322, 323]. They are central coordinators of the adaptive immune response and can differentiate into a diverse range of functional subtypes to help effector immune populations. A key function of $CD4^+$ T cells is to help $CD8^+$ cytotoxic T cells via direct and indirect mechanisms. Once activated $CD4^+$ T cells secrete interleukin-2 (IL-2), which in turn directly activates $CD8^+$ cytotoxic T cells which express CD25, the high-affinity IL-2 receptor α subunit. This drives CD8+ cytotoxic T cell

differentiation, proliferation, and effector function [323]. In addition $CD4^+$ T cells offer indirect help to $CD8^+$ cytotoxic T cells by maintaining and supporting pro-inflammatory dendritic cells which themselves provide activating signals to CD8+ cytotoxic T cells [322].

An important CD4+ T cell population in the context of the tumour immune microenvironment is regulatory T-cells. These cells are characterised by the expression of the master transcription factor forkhead box protein P3 known as FOXP3 and they contribute to an immunosuppressive microenvironment [324]. Ordinarily, they play critical roles in maintaining self-tolerance, thus preventing auto-immune disease, however in the context of cancer, they suppress an appropriate anti-tumour response therefore promoting tumour progression [324]. Multiple meta-analyses have found an association between high levels of FOXP3 TILs and higher tumour grade, lymph node involvement and poor survival outcomes [325-327] in ER+ breast cancer although in multivariate analysis regulatory T-cell levels were not an independent prognostic factor [328, 329].

1.12.2 The role of Macrophages in cancer

Macrophages are myeloid cells which form a key element of the innate immune response. They originate from monocyte precursor cells in the blood and differentiate under the influence of various growth factors and cytokines within the tissues that they infiltrate [330, 331]. They are a heterogenous cells population with a diverse range of functions which include exogenous antigen presentation, phagocytosis, and immunomodulation. M1-like and M2-like macrophages show different metabolic characteristics and distinct gene expression and immunohistochemical profiles. M1-like macrophages promote and contribute to inflammation in response to pathogens and tumour cells, contributing to tumour suppression. They secrete inflammatory cytokines such as IL-12, IFNγ, TNFα and nitric oxide synthase [331]. They are characterised by the expression of CD68 and CD64 among other IHC markers [332]. In contrast M2-like macrophages create an immunosuppressive microenvironment promoting tissue repair and tumour progression through the secretion of range of anti-inflammatory cytokines e.g. IL-4, IL-10 and IL-13 [333]. They are characterised by the expression of high levels of CD206, and CD163 [332]. Macrophage polarisation describes the process by which macrophages, under the influence of different microenvironmental signals and stimuli become the distinct M1-like and M2-like phenotypes. M1-like macrophages can switch to an M2-like phenotype and vice versa under the influence of various microenvironmental factors, such as cytokine secretion, hypoxia, infection, and inflammation [332]. TAMs are generally considered

to represent M2-like type phenotype macrophages promoting tumour progression through a range of mechanisms (Figure 1.10).

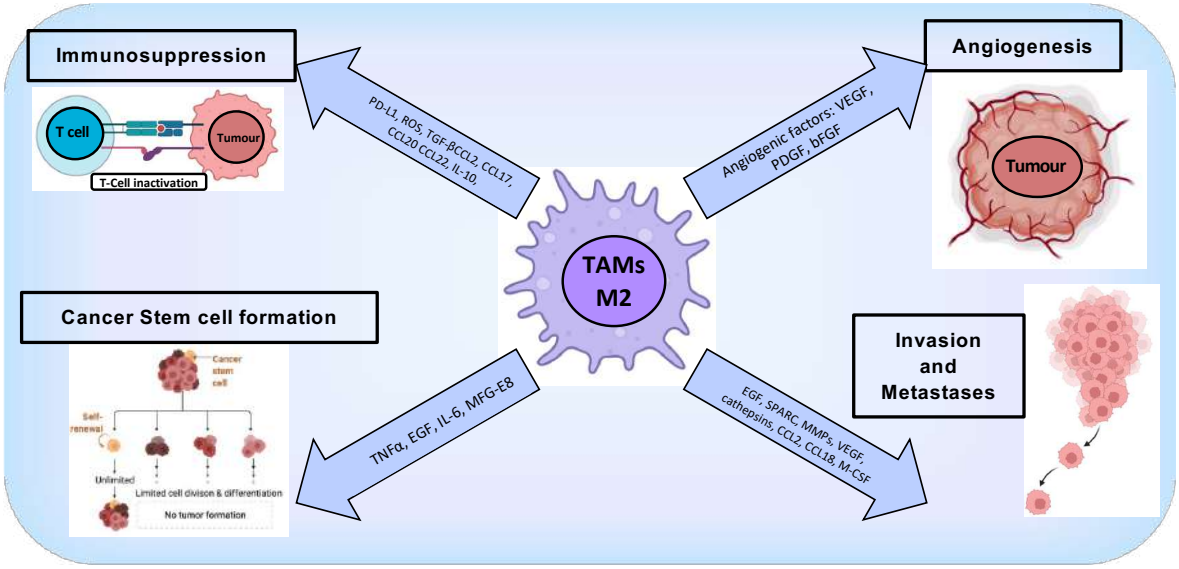


Figure 1.10: Tumour associated M2-lke macrophages contribute to tumour progression through a range of mechanisms: these include promotion of angiogenesis, invasion and metastases, cancer stem cell formation and through the promotion of an immunosuppressive tumour microenvironment.

1.12.3 The Immune Landscape in ILC

Breast cancer has previously been classed as a relatively weakly immunogenic tumour, yet there is growing evidence that subsets of breast cancers are associated with significant numbers of TILs, and that their quantification holds prognostic and predictive value [313, 314]. In TNBC and HER2+ disease increased TILs are associated with improved DFS and with higher response rates to neoadjuvant therapy [313, 314, 334-338].

The prognostic significance of TILs in ER+ breast cancer is less clear [338]. ER+ tumours are generally characterised by lower TILs than TNBC and ER-/HER2+ disease, and ILCs show lower TILs than IC-NST [337]. In addition, low levels of stromal TILs are seen in ILC metastases, with the 'mixed non-classic' histologies showing higher levels [282]. Furthermore, no significant differences in TILs levels have been observed between different ILC metastatic sites [290].

Two recent studies have studied TILs in large cohorts of ILC and identified that a subset of cases show high levels of stromal TILs [282, 339]. Moreover, high TILs scores (defined as stromal TILs score > 10% [282] and > 5% [339] in the two studies respectively) were associated with younger patient age, lymph node involvement and more proliferative tumours in the first study [282], as well as a range of poor prognostic factors and worse OS and DFS

in the second (Figure 1.11) [339]. When ILC histological subtypes were considered, one study showed that compared to classic ILC, significantly lower TILs were present in alveolar ILC ($p = 0.02$) and higher levels in 'mixed non-classic' ILC ($p < 0.001$) [282]. A distinct pleomorphic group was not defined in these studies. These studies suggest that the role of TILs contrasts that observed in TNBC and HER-2 positive disease.

A more recent study examined phenotypic, transcriptional, and functional diversity in cohort of primary untreated ER+ ILCs (n = 87) and IC-NSTs (n = 94) using flow cytometry, multiplexed IHC and single cell RNA sequencing technologies [340]. This revealed that macrophages, rather than T cells, were the predominant immune subpopulation infiltrating the tumour bed in ILC and the most transcriptionally diverse immune subpopulation between ILC and IC-NST [340]. Multiplexed IHC revealed a higher proportion of macrophages in ILC, with this immune subpopulation accounting for 82% of the immune infiltrate within the tumour and 50% within the stromal compartment, compared to IC-NST, where the macrophages contributed to less than 50% of the immune infiltrate in both tumour and stromal compartments [340]. Cell quantification data showed that the increased proportion of macrophages in ILC resulted from reduced T cell counts as opposed to expansion of the macrophage immune subset [340].

Overall, there is an increasing body of work providing new insights into ILC within the immune-oncology space. There is however an ongoing need to delineate the significance of the immune microenvironment within different ILC subtypes, especially pleomorphic ILC. Moreover, there is a need to further characterise TILs into the various immune subpopulations, as it may be the quantification of specific immune subpopulations which determines clinical outcome, or indeed their spatial location, rather than the simple quantification of TILs identified histologically. These findings will provide a strong foundation for the development of future clinical trials assessing the efficacy of immunotherapy in clinically aggressive, immunogenic ILCs.

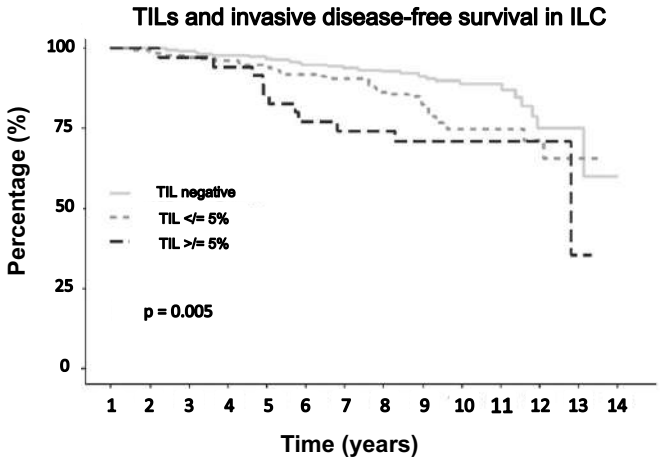


Figure 1.11: TILs are associated with poor survival in ILC: Kaplan–Meier (KM) plot showing IDFS of ILC patients stratified according to tumour TILs quantification: negative (n = 239), ≤ 5% (n = 185), > 5% (n = 35). Taken from [339]

1.13 Project aims

There is a significant unmet clinical need in aggressive ILCs. To improve the treatment options and clinical outcomes for these patients, there is a need to characterise the transcriptomic and immune heterogeneity in aggressive ILCs, to enable the development of novel molecular and immune therapies with accompanying biomarker stratification for treatment.

To address this unmet need, the purpose of this PhD is to characterise the transcriptional and immune heterogeneity in clinically aggressive ILCs including pleomorphic ILC. The key aims of the project are to:

1) Identify the molecular drivers of clinically aggressive ILC
2) Evaluate the immune landscape and prognostic associations in ILC
3) Map the heterogeneity of immune infiltrate in pleomorphic ILC
4) Map the sub-clonal heterogeneity of metastatic ILC

Chapter 2: Methods

2.1 Clinical samples

A retrospective series of primary untreated ILC tumours (n = 163) was obtained with appropriate ethical approval using two independent FFPE retrospective ILC cohorts: an ILC cohort from the KHP tissue bank with detailed clinical outcome and treatment data (n = 149), (REC Number: 12/EE/0493), and cases collected through the GLACIER study (n = 14) with outcome data provided by Public Health England (MREC 06/Q1702/64). All patients provided written consent for the use of material for research purposes. Appropriate representative formalin-fixed paraffin-wax embedded (FFPE) tissue blocks were chosen based upon their histology reports. Haematoxylin and eosin (H&E) sections of each case were reviewed by two independent histopathologists (Dr Ioannis Roxanis, consultant breast histopathologist, the Institute of Cancer Research, and myself) to confirm the ILC histological subtype and tumour content. ILCs were classed as classic, pleomorphic, solid, alveolar, mixed with a pleomorphic component, and mixed without a pleomorphic component. For the purposes of the analysis pleomorphic cases encompassed pleomorphic and mixed cases with a pleomorphic component, whilst non-pleomorphic cases encompassed classic, solid, alveolar and mixed cases without a pleomorphic component. Patients were stratified based on clinical outcome data into those with ‘early relapse’, (defined as relapse within 3 years of initial diagnosis), ‘late-relapse’ (relapse after 6 years of initial diagnosis), ‘intermediate’ relapse (relapse between 3 and 6 years of initial diagnosis) and ‘no-relapse’. An anonymised summary table of the clinical features of the cohort is provided in Supplementary Table 1.

2.2 Validation of histological classification using digital pathology

I performed an objective assessment of nuclear size using QuPath (an open-source platform for the analysis of pathology images [341]) in a subset of pleomorphic and non-pleomorphic ILCs. A total of 200 - 250 representative tumour cells were selected in each case and nuclei were segmented using StarDist, a cell detection algorithm that utilises star-convex shape priors for microscopy images [342]. Segmentation was achieved using the publicly available model “he_heavy_augment”. An average (mean) nuclear size and the standard deviation of nuclear size was calculated for each case. Tumour cell nuclear size and standard deviation of nuclear size were compared between histologically defined pleomorphic and non-pleomorphic cases using a Mann-Whitney U test.

2.3 DNA extraction

I performed DNA extraction of 95 ILCs (52 pleomorphic and 43 non-pleomorphic cases) at the Research and Innovation Hub in Cancer, King's College London. James Rosekilly (KHP tissue bank) prepared 10 x 10µm FFPE tissue sections per tumour case and I used needle macrodissection to enrich for tumour content. Genomic DNA from the FFPE tissue sections was extracted using QIAamp FFPE Tissue Kit (Qiagen, Manchester, UK) according to the manufacturer's instructions. I assessed DNA quantity using the Qubit Fluorometer (Fisher Scientific, Loughborough, UK).

2.4 Targeted DNA sequencing

Next Generation Sequencing (NGS) took place at The Centre for Molecular Pathology, The Royal Marsden NHS Foundation Trust, Sutton. Ninety-five samples were sequenced using a targeted, custom-designed capture panel (RMH200Solid panel) consisting of 233 cancer related genes (Supplementary Table 2). Library preparation was performed by Paula Proszek, Sabri Jamal and Ridwan Shaikh. NGS libraries were prepared from 200 - 400ng of DNA using the KAPA HyperPlus Kit (Kapa Biosystems, Wilmington, MA, USA) and IDT UDI 8bp adapters (Integrated DNA Technologies, Coralville, USA), following the manufacturer's protocol. Sequencing was performed on a NovaSeq6000 (Illumina, San Diego, CA, USA) with 100bp paired end reads and v1 chemistry.

Data was analysed using an in-house pipeline (MDIMSv4) by aligning reads to the reference genome build GRCh37/Hg19 followed by the marking of PCR duplicates and calculation of various quality control (QC) metrics using Picard v2.8.1. Copy number was calculated using an in-house developed pipeline utilising Picard tools. Segmentation was performed using the R package DNAcopy v1.64.0. A log2 ratio > 1 is considered an amplification while < 0.5 is a suspected deletion.

Manta v1.2.2 [343] and Pindel v0.2.5b8 [344] were used for the detection of structural variants. GATK was used for variant calling using MuTect2 [345]. Variant Call Format (VCF) files from unpaired samples were annotated using Personal Cancer Genome Reporter v.0.6 (PCGR) and checked manually on IGV. A minimum variant allele depth x 10 and Variant Allele Frequency (VAF) threshold of 5% were applied. Low frequency mutations with VAF < 5% were subsequently assessed by myself in the sequencing data. Splice variants were included if within +/- 2bp from the exon boundary and silent variants were excluded, unless known to

impact exon expression. GnomAD was used to remove SNPs of germline origin excluding SNPs with a total population frequency $>10^{-4}$. Mutational load was calculated as the total number of nonsynonymous mutated bases in the tumour genome divided by the Mb of the genome covered [346].

2.5 RNA extraction

I performed RNA extraction of 47 pleomorphic ILCs (KHP pleomorphic cohort) at the Research and Innovation Hub in Cancer, King's College London. Tissue needle macro dissection was used to enrich for tumour content and 10 x 10µm FFPE tissue sections (prepared by James Rosekilly, KHP tissue bank) were used per case. RNA from FFPE tissue sections was extracted using QIAamp FFPE Tissue Kit (Qiagen, Manchester, UK) according to the manufacturer's instructions. I assessed RNA quantity using the Qubit Fluorometer (Fisher Scientific, Loughborough, UK).

2.6 RNA sequencing of KHP pleomorphic cohort

Next Generation Sequencing (NGS) took place at The Genomics Facility, the Institute of Cancer Research, Sutton. A total of 250 -1000ng of RNA, from the 47 samples was treated with TurboDNase (Invitrogen, #AM2239) to remove genomic DNA contamination. Ribosomal RNA was then removed from the samples using the NEBNext rRNA Depletion Kit (NEB, #E6310X) following the manufacturer's instructions. From the resulting RNA, strand-specific libraries were created using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB, #E7760) on the Agilent Bravo (option B). Final libraries were quantified using TapeStation (Agilent) and qPCR (Roche, #KK4835), then clustered at a molarity of 300pM. Sequencing was performed on an Illumina NovaSeq 6000 using PE x 100 cycles v1.5 chemistry, to achieve coverage of 100 million reads per sample.

2.6.1 KHP dataset RNA sequencing data processing

RNA sequencing generated 25.6 to 143.8 million paired-end reads per sample (n = 47). Library quality was assessed using FastQC, FastQ Screen [347] and MultiQC (v1.9) [348] .Reads were trimmed using Trim Galore (v0.6.6). Paired-end reads (100bp) were aligned to the human reference genome GRCh38, using STAR 2.7.6a (PMID: 23104886) with quantMode

GeneCounts and --twopassMode basic alignment settings. GENCODE (v22) was used for feature annotations.

2.6.2 KHP dataset RNA sequencing data analysis

2.6.2.1 Differential gene expression analysis

Differential mRNA abundance analysis was performed using R package edgeR (v3.34.0) [349] in R statistical programming environment (v4.1.0) by Shaun Tan (BCR bioinformatics team). Genes with low expression were filtered out using edgeR's function 'filterByExpr()' with parameters: min.count = 5 and min.prop = 0.5. Raw counts were normalised using edgeR's TMM (trimmed mean of M-values) method and differential mRNA abundance was performed using the quasi-likelihood (QL) F-test using the model ~0 + group. ENSEMBL gene identifiers were annotated with gene symbols using the R package org.Hs.eg.db (v3.10.0) [350]. TPM normalised counts were transformed to log2 counts per million (CPM) for visualisations and survival modelling.

Differential gene expression analysis was performed in 45 samples. Two of the 47 samples which were sequenced were excluded due to inadequate follow-up required to define relapse status. Differential gene expression was compared in relapse (n = 24) vs no relapse (n = 21) and early (n = 11) vs late relapse (n = 6), Using the transformed $\log_2$ CPM, top hits of the differentially expressed genes were visualised in both volcano plots and heatmaps. Genes were considered as significantly differentially expressed if they fulfilled both a p-value threshold of 0.001 and a $\log_2$ fold change threshold of at least ±1.

2.6.2.2 Univariable survival analysis

Potential prognostic genes associated with survival in the pleomorphic KHP discovery cohort were first identified using univariable survival analysis. Genes associated with overall survival (OS) and metastasis free survival (MFS) were identified independently. Two of the 47 patients in the cohort presented with *de novo* metastatic disease and were excluded from survival analysis. Clinical and gene expression data in the remaining 45 pleomorphic samples was used. To account for potential clinical factors, age, tumour stage and lymph node status were adjusted for when conducting the analysis. To limit the extent of the survival data, a cut-off of

10 years was applied. Prognostic genes were identified if they fulfilled an un-adjusted p-value criteria of < 0.05.

2.6.2.3 Random Forest model training

Random Forest survival modelling was conducted by Shaun Tan (BCR bioinformatics team) using R version 4.1.0 with the R package randomforestSRC [351] and the rfsrc function to create a classifier specific to pleomorphic ILCs. This approach was conducted to validate the genes identified in univariable survival analysis from the KHP discovery cohort, on external published datasets (SCAN-B [352], TCGA [32] and METABRIC [353]), which included both OS and MFS associated genes. The training set used the gene expression data of the selected genes from univariable survival analysis and each of the sample's survival data. To ensure that the trained model could be applied to the published datasets, only the gene expression data common among the KHP discovery cohort and the 3 published datasets was used. Gene expression data from OS and MFS associated genes were trained with OS and MFS outcomes respectively. To obtain the most accurate model, optimization was performed prior to validation in the published datasets. Parameters that were optimized for were the number of trees (ntree), node size (nodesize) and the number of variables to randomly sample at each split (mtry). Other parameters used but were not optimized for were sampling without replacement ("swor" for samptype) and a 75/25 train/test split (sampsize). Seed 42 was used to enable consistency when producing results.

The optimal number of trees was first determined by varying the number of trees and fixing the other set of parameters. Mtry and node size were fixed to the square root of the number of genes and 10 respectively. Out-of-bag (OOB) error rates produced by the model for the different number of trees used were plotted on a line plot to identify the point at which the OOB error rates have stabilised. It is at this point that the optimal number of trees was chosen. With optimal number of trees identified, a grid search was performed on node size and mtry to identify the optimal parameters. Node size ranging from 3 to 10 and increasing by 1 each time was used, while mtry ranging from the square root of the number of genes to 1/3 of the genes and increasing by 1 each time was used.

Once the optimal parameters were identified, the random forest model was trained using these parameters on data from KHP discovery cohort.

2.6.2.4 Validation in independent purified ILC datasets

Because of the highly pure nature of the KHP discovery cohort due to tissue macrodissection of the tumour cells prior to RNA sequencing, there was a possibility that the random forest modelling prediction would perform poorly on unpurified validation datasets, as they are more heterogenous containing signal from non-tumour cells in addition to tumour cells. To ensure the robustness of the model, *in-silico* purification of the gene expression data in the 3 published validation datasets was performed using the ISOpureR package [354].

The purified gene expression data of the validation datasets were then fed into the random forest model to predict risk scores. associated with each sample in the published datasets. These risk scores are mortality values that represent estimated risk for each individual and were calibrated to the scale of the number of events of the KHP discovery cohort. Risk scores obtained from the model were then divided into tertiles, ranging from 0 - 33%, 34 - 66% and > 66%. The 3 groups of risk scores were then used in a cox-proportional hazards model to evaluate the ability of predicted risk scores to accurately determine survival outcomes up to a 10-year cut-off period. P-values of the different risk groups shown are relative to the 0 - 33% group. A trend test p-value was also calculated to determine if the survival of the different groups followed a specific trend. Additionally, Kaplan-Meier (KM) plots were generated alongside the cox statistics to visualize the survival outcomes in the 3 groups. Validation of OS and MFS genes were only performed on OS data in the published datasets, as only OS data was present, except for SCAN-B. In SCAN-B, MFS genes were also validated on MFS outcomes.

2.7 RNA sequencing of non-pleomorphic ILCs

Fifteen non-pleomorphic (all classic) ILCs were characterised at the RNA level using the HTG Molecular Diagnostics Transcriptome panel. Of these cases, 5 patients relapsed early (< 3 years of primary diagnosis) and 10 patients relapsed late (> 6 years after primary diagnosis). The HTG panel provides coverage of most human mRNA transcripts including isoforms and interrogates 19,616 targets using FFPE tissue sections. For each case an H&E and 5µm thick unstained FFPE tissue section were prepared by James Rosekilly (KHP tissue bank). I annotated the H&Es identifying areas of tumour content of at least 50% and a minimum tumour area of 11 mm^2 was selected per case, as per the HTG protocol. Annotated H&Es and unstained tissue sections were shipped to HTG Molecular Diagnostics (Tucson, Arizona) and

the annotated H&Es were used to select the tumour areas for sequencing on the corresponding unstained tissue section.

The samples together with an HCT-15 control technical replicate were randomized prior to placement on the HTG EdgeSeq sample plate to reduce any potential intra-plate biases during processing. The samples were randomized by HTG. Samples were processed in accordance with the manufacturer's recommendations, HTG EdgeSeq processing (instrument method). Samples were transferred to a standard 96-well micro-titre plate, referred to as the sample plate.

Target capture was performed by the HTG EdgeSeq chemistry. Briefly, the nuclease protection probes (NPPs) were added to the lysed samples in the sample plate in excess amount and hybridized to the target messenger RNA (mRNA). An S1 nuclease was subsequently added to digest non-hybridized RNA and excess NPPs, thus producing a stoichiometric amount of target mRNA NPP duplexes. After the S1 digestion was complete, the processed sample was transferred to a new 96-well micro-titre plate with a v-bottom, referred to as the stop plate, and S1 digestion was terminated by the addition of a termination solution followed by heat denaturation of the S1 enzyme.

2.7.1 Library preparation and sequencing

Each processed sample from the stop plate was used as a template to set up polymerase chain reaction (PCR) reactions with specially designed primers, referred to as tags. These tags share common sequences that are complementary to 5'-end and 3'-end "wing" sequences of the probes and common adaptors required for cluster generation on an Illumina sequencing platform. In addition, each tag contains a unique barcode that was used for sample identification and multiplexing. HTG developed 12 forward and eight reverse primer tags and therefore all samples were simultaneously analysed on an Illumina sequencing platform. The library was prepared using a PCR with OneTaq (New England Biolabs) and EdgeSeq PCR tag primers (HTG molecular). Following the PCR, a clean-up procedure was performed with AMPure clean-up beads (Beckman). The library was then quantified in with Accuclear fluorescent dye (Biotium) and a Molecular Devices SpectraMax plate reader. All samples and controls were quantified in triplicate.

Following quantification, the HTG Library Calculator software was used to ensure that there was a sufficient concentration of sample for library pooling and to determine the appropriate

dilutions for building the library pool. All samples processed within the study had sufficient PCR product to be pooled for sequencing. The HTG Library Calculator also determined the volume and specific type of denaturation reagents to be used for the library. The library was denatured by adding 0.2N NaOH to the library tube. The tube was then vortexed, spun down, and incubated for five minutes at room temperature. Cold HT1 buffer (Illumina) was added to the library tube, followed by the addition of 200mM Tris pH 7.4 to neutralize the NaOH. The sample was vortexed and spun down.

PhiX control adaptor-ligated library was spiked in at a concentration of 12.5pM to the pooled library. The library was vortexed and spun down for a final time. The denatured library was loaded (1,300µL) into the appropriate well of the NextSeq sequencing cartridge. The concentration of the pooled library loaded on the NextSeq flow cell was 1.7pM.

The sequencing was performed on the Illumina NextSeq 550 sequencer in accordance with manufacturer's recommendations but also including two HTG custom sequencing primers. The sequencing data on mRNA expression of target genes was imported into HTG EdgeSeq Parser software for alignment and quantification of the reads. The HTG EdgeSeq Reveal Application was utilized to quality check and normalize the data.

2.7.2 Post-sequencing quality control (QC)

Post-sequencing QC metrics (QC0, QC1, QC2, and QC3) were used to detect four different types of sample failure. QC0 detects degraded RNA or poor quality / quantity samples, by assessing the percentage of overall reads being allocated to the positive process control probe for each sample. QC1 detects samples with insufficient read depth; this is evaluated by setting the cut-off at the minimum number of reads that can be allocated to each sample for the data set to be repeatable. QC2 detects samples with minimal expression variability, which is determined by the median log2 counts per million (CPM) for negative control probes. QC3 detects samples with elevated gDNA background, which is caused by incomplete digestion of gDNA by the DNase enzyme.

2.7.3 Data processing and analysis

Data was returned from the sequencer in the form of demultiplexed FASTQ files, with four files per original well of the assay. The HTG EdgeSeq Parser was used to align the FASTQ files to

the probe list to collate the data. Data are provided as data tables of raw, QC raw, log_2 CPM and median normalized data.

The data was processed by Syed Haider and Hui Xiao (BCR bioinformatics team). mRNA profiling was performed by using the HTG Transcriptome Panel which contains 19,616 probes including 4 positive process control probes, 100 negative process control probes, 22 genomic DNA probes, 92 external RNA control consortium (ERCC) probes, and 19,398 probes specific to human mRNA transcripts.

The raw counts of human mRNA specific probes were kept for analysis of gene expression. Probes with minimum counts lower than 10 in more than 30% of samples in the cohort were filtered out. The probe counts were then normalised by using the trimmed mean of M-values (TMM) method implemented in R package edgeR to correct library size biases between samples. The normalised probe counts were transformed into log2 CPM.

Differentially expressed genes (DEGs) between the early relapse (< 3 years) and late relapse (> 6 years) groups were identified based on the normalised counts by using quasi-likelihood negative binomial generalized log-linear model (implemented in R package edgeR) with the FDR threshold of 5%. An unadjusted p-value threshold of $p < 0.001$ and log_2FC +/-1 were used to identify significantly differentially expressed genes.

Cancer Hallmarks was used for pathway analysis which was performed using gene set enrichment analysis (GSEA) via the 'fgsea' library in R. Genes were ranked according to their signed p-values in which the signs were defined by the changing directions (up-regulation as '+' and down-regulation as '-'). 10000 gene permutations were used to calculate statistical significance for pathways. P-values were adjusted by false discovery rate (FDR) for multiple test correction. An adjusted p-value of 0.05 was required for statistical significance of a cancer hallmark pathway.

2.8 Differential gene expression in the TCGA dataset

Differential gene expression was performed in pleomorphic vs non-pleomorphic ILCs from TCGA [32]. This first involved the identification of pleomorphic cases using QuPath.

2.8.1 Identification of pleomorphic cases in TCGA

Image analysis was performed by Chris Starling (Breast Cancer Now Histopathology facility) using QuPath, an open-source platform for the analysis of pathology images [341]. Colour vectors were corrected for each image using QuPath's automated pre-processing tool. Tumours were annotated using the "Simple Tissue Detection" plugin and nuclei were detected with StarDist [342] using the publicly available model "he_heavy_augment". Nuclear features were calculated using the "Intensity Features" plugin with a pixel size of 0.5μm, tile size 25μm and a Haralick distance of 1 with 32 bins. To remove false detections, a minimum optical density of haematoxylin (HxOD) was set at 0.15 and maximum and minimum nuclear areas were set at 250μm^2 and 15μm^2 respectively and these were removed from the analysis.

Nuclei with a circularity of < 0.85 were classified as "Stroma". Of the remaining cells, those with a nuclear area > 35μm^2 were classified as "Tumour", those with nuclear area < 20μm^2 were classified as "Immune cells" and the remaining cells were classified as "Other". The mean and standard deviation (SD) HxOD of all cells classified as "Tumour" in each image was calculated. This was compared with the value of each cell in the image classified as "Other". For lightly stained images (those with a HxOD mean + SD < 1) the HxOD threshold was set at mean + 1.5 SD. For strongly stained images (HxOD mean + SD > 1) the HxOD threshold was set to mean + SD. Those cells classed as "Other" with HxOD below the threshold were classed as "Tumour" and those above as "Immune cells."

A threshold for pleomorphic tumour cell nuclear area was set at 100μm^2 based on four times the average size of a lymphocyte [78]. Those "Tumour" cells above this threshold were classified as "Tumour: Positive" and those below as "Tumour: Negative". To remove false positives the mean and SD of all pleomorphic cells in each image was calculated. Any cell classified as pleomorphic ("Tumour: Positive") with an HxOD more than 2 SD below the mean was analysed for Haralick entropy (F8) in the nuclear haematoxylin staining. Those cells with entropy > 6.5 were considered true detections. Those with entropy < 6.5 were considered too "smooth" and represented false positives. The data was then used to calculate the proportion

of pleomorphic tumour cells (tumour cells with nuclear areas > 100μm^2) and the density of pleomorphic cells per mm^2 of tumour tissue calculated.

2.8.2 Differential gene expression based on histology in TCGA

Differential gene expression was performed in pleomorphic (n = 36) vs non-pleomorphic (n = 143) ILC by Shaun Tan (BCR bioinformatics team). Hormone status information was available for some of the patients and using this a further analysis was performed in ER+/HER2- pleomorphic (n = 17) vs ER+/HER2- non-pleomorphic (n = 71) cases. A linear model fit and empirical Bayes moderation was used to determine the differentially expressed genes. Top differentially expressed gene hits were then identified if they fulfilled the criteria of an adjusted p-value threshold < 0.05 and $\log_2$FC +/-1.5.

2.9 Histological Assessment of TILs

TILs were histologically characterised using the 'Salgado' scoring method in which TILs are reported for the stromal compartment only with the 'Salgado' score representing the percentage of the stromal area occupied by TILs (Figure 2.1) [308]. TILs were quantified by two independent histopathologists (Dr Ioannis Roxanis, consultant breast histopathologist, the Institute of Cancer Research, and myself) and an average of the scores was used for analysis.

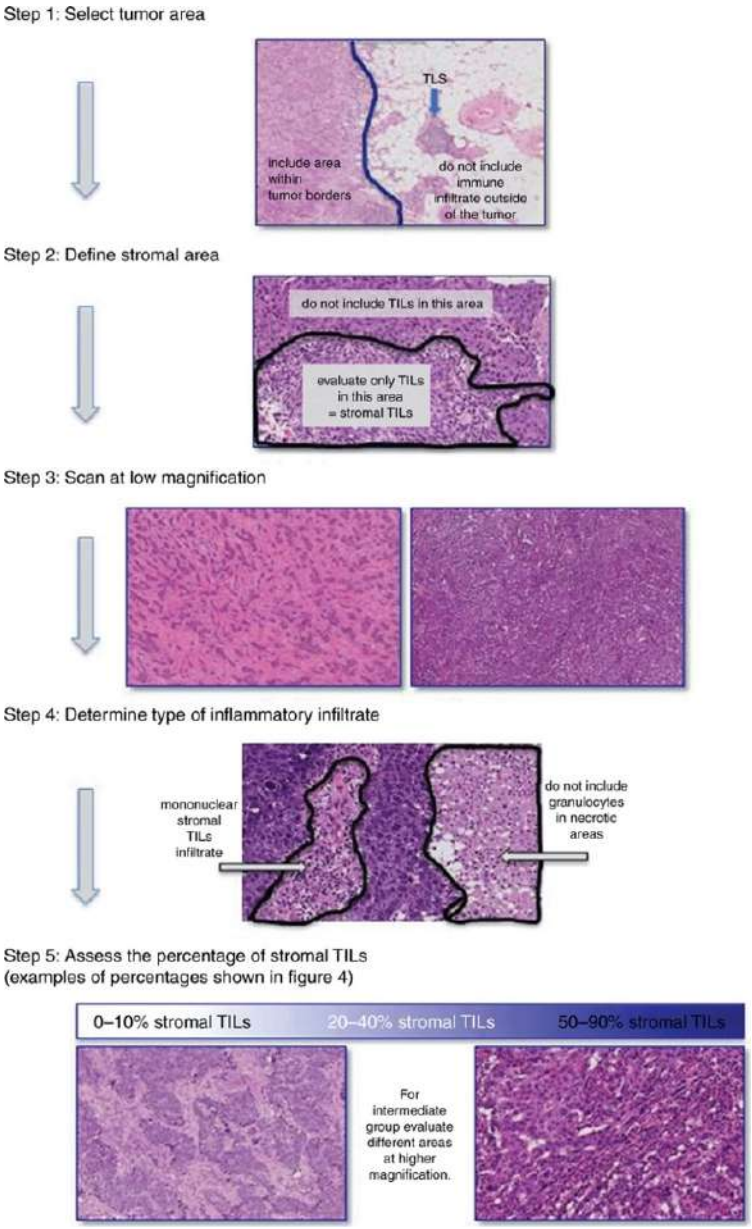


Figure 2.1: Salgado scoring methodology: Diagrammatic representation of the standardised methodology used to quantify stromal TILs in breast cancer [308].

2.10 Association of stromal TILs with clinical outcome

Univariable survival analysis using a cox-proportional hazards model was performed on the KHP cohort with stromal TILs used as an independent variable. This analysis was conducted to identify the association between stromal TILs scores MFS and OS. Separate analysis was performed on pleomorphic and non-pleomorphic patients. Seven patients with *de novo* metastases were removed prior to survival analysis. The follow-up time was truncated at 15 years, therefore patients with metastatic events occurring after 15 years were censored.

For analysis of TILs scores, each individual stromal TILs score was categorised as either ≤ 5% or > 5%. Univariable survival analysis was conducted on these TILs score categories to identify association with MFS and OS. Clinical covariates such as tumour stage (T1, T2, and T3), lymph node status (negative: 0, positive: > 0) and hormone status of the tumour were adjusted for (triple negative+ HER2 positive: 1, others: 0). To visualise survival outcomes KM plots were generated.

2.11 NanoString GeoMx® Digital Spatial Profiling (DSP)

NanoString GeoMx® DSP was performed on 20 pleomorphic ILCs with at least 5% stromal TILs. FFPE blocks were sectioned at 6 µm thickness, and sections mounted on adhesive, positively charged slides (SuperFrost Plus™ adhesion slides, Epredia™). Multiple tumour sections were placed on each NanoString DSP slide. Seven NanoString DSP slides were used in total, with a total of 20 ILCs. After sectioning, air-dry sections were shipped to NanoString at ambient temperature.

The slides were deparaffined and subjected to antigen retrieval procedures. Tissue sections were co-incubated with fluorescent-labelled antibodies as morphology markers: Pan-Cytokeratin (PanCK, epithelial), CD45 (pan-immune), CD3 (T-cell) and syto83 (DNA). The slides were also incubated with photocleavable oligonucleotide-labelled primary antibodies targeting 71 immuno-oncology markers from the NanoString GeoMx® Human Immuno-Oncology (IO) Panel, 3 housekeeping targets and 3 negative controls (Supplementary Table 3). Once the staining was complete, slides were loaded onto the GeoMx DSP instrument, and scanned producing a digital fluorescent image of the tissue. Individual regions of interest (ROIs) were then selected. Twelve ROIs were selected per slide. These were selected based on the presence of abundant CD45+ cells (immune cells) in proximity to PanCk+ cells (tumour cells).

ROIs were segmented into the tumour and immune cell compartments, producing two distinct areas of interest (AOIs) and oligos from the immune cell (CD45+ve) compartment were selected to be released upon exposure to UV light. Photocleaved oligos were collected via microcapillary aspiration and dispensed into a 96-well plate, then hybridized to 4-color, 6-spot optical barcodes and digitally counted in the nCounter system (NanoString). Digital counts from barcodes corresponding to protein probes were normalized to internal spike-in controls (ERCCs), and then normalized to signal-to-noise ratio (SNR) calculated based on the mean of the 3 negative controls. A SNR < 1 was considered non-specific background. Therefore, only those proteins markers were included in the analysis whereby SNR > 1 in at least three ROIs was observed. Given that this was a small discovery cohort relaxed statistical thresholds of $p < 0.1$ and $\log_2 FC$ +/- 0.585 were used to identify significantly differentially expressed immune-oncology proteins between patient comparison groups.

2.12 Immunohistochemistry (IHC)

CD68 IHC was performed in 54 pleomorphic and 78 non-pleomorphic ILCs. The staining was completed by James Rosekilly (KHP tissue bank) using 3 - 4µm thick whole FFPE tissue sections. Slides were incubated with anti-CD68, rabbit monoclonal antibody 1:1000, EPR13501 (Abcam, UK), using the Dako-Autostainer Link 48 with the EnVision FLEX kit according to the manufacturer's instructions (Agilent Technologies, UK). Human breast, tonsil, appendix, prostate, and kidney tissue were used as positive controls. Stromal cells were used as an internal positive control. CD68 staining was scored by me using QuPath [341]. A pixel classifier was trained to recognise DAB staining using Random Trees methodology at high resolution.

Dual IHC was performed in 35 pleomorphic ILCs. Eighteen patients relapsed and 17 patients did not relapse. FFPE sections which were 4µm thick were stained on a Ventana Discovery plus (Roche). Epitopes were unmasked at 95^{o}C and high pH for 64 minutes. After endogenous peroxide enzymes were blocked, sections were incubated in CD163 primary antibody (Abcam) diluted 1:500 with casein protein block for 32 minutes at room temperature. The primary was detected using UltraMap anti-Rabbit HRP polymer (Roche) for 8 minutes at room temperature and visualised with DAB (ChromoMap - Roche). Antibodies were stripped at 100^{o}C, low pH, for 8 minutes and any remaining peroxidase enzymes were blocked again. The second primary, CD68 (Merck Millipore), was applied at 1:500 with protein block for 32 minutes at room temperature. This was detected using UltraMap anti-Mouse HRP polymer (Roche) for 8 minutes and visualised with Discovery Green (Roche) for 32 minutes with the green substrate followed by 16 minutes with the activator. Slides were counterstained with Haematoxylin for 4 minutes and blued in blueing reagent. The liquid coverslip was removed with detergent and sections dehydrated, cleared, and mounted.

Slides were digitised using a Nanozoomer XR (Hamamatsu) using a x 20 dry objective and the ndpi images imported into QuPath (Bankhead, 2017). Colour vectors were set manually to Haematoxylin (0.785 0.525 0.328), DAB (0.453 0.493 0.742), and Green (0.733 0.337 0.591) with background 8-bit RGB values 231, 236 and 227.

Images were manually annotated with green staining classified as 'M1', DAB staining as 'M2', and cells negative for both markers as "Other". Whitespace was classified as "*Ignore". These annotations were used to train a pixel classifier using Random Trees at a resolution of 1.81 µm/pixel and utilising all multiscale features. The area occupied by each stain was used to

calculate the proportion of M1 and M2 macrophages in each tumour. A Mann-Whitney U test was used to study differences in M2/M1 ratios in relapsing vs non relapsing patients.

Univariable survival analysis using a cox-proportional hazards model was performed on the CD68 IHC cohort with CD68 scores used as an independent variable. This analysis was conducted to identify the association between CD68 scores and MFS and OS. Separate analysis was performed on pleomorphic and non-pleomorphic patients. Seven patients with *de novo* metastases were removed prior to survival analysis. The follow-up time was truncated at 15 years, therefore patients with metastatic events occurring after 15 years were censored. Clinical covariates such as tumour stage (T1, T2, and T3), lymph node status (negative: 0, positive: > 0) and hormone status of the cancer were adjusted for (triple negative + HER2 positive: 1, others: 0). CD68 scores were split into tertiles. The tertiles of the scores were then used in univariable survival analysis to identify any association with MFS and OS. KM plots depicting the outcomes seen in the different CD68 score levels were plotted.

Univariable survival analysis using a cox-proportional hazards model was performed to assess M2/M1 ratio in the KHP cohort with M2/M1 ratio used as an independent variable to assess any association with MFS. The tertiles of the M2/M1 ratios were used and KM curves plotted.

2.13 NanoString GeoMx® Whole Transcriptome Atlas (WTA)

NanoString GeoMx® WTA was performed on 10 pleomorphic ILCs with at least 4% stromal TILs. Blocks were sectioned at 6 µm thickness, and sections mounted on adhesive, positively charged slides. Three tumour sections were placed on each NanoString slide. Three NanoString slides were used in total. After sectioning, air-dry sections were sent to the Genome Centre, Queen Mary, University of London at ambient temperature.

The slides were deparaffined and subjected to antigen retrieval procedures. They were stained with RNA scope probes and GeoMx DSP oligo-conjugated RNA detection probes. The RNA *in situ* hybridization (ISH) probes were conjugated to unique DNA indexing-oligonucleotides (DSP barcodes) via a UV photocleavable linker. The slides were also co-incubated with fluorescent-labelled visualization antibodies: Pan-Cytokeratin (PanCK, epithelial cell marker), CD45 (pan-immune), alpha smooth muscle actin (*α*-SMA, cancer-associated fibroblast) and syto83 (DNA).

Slides were subsequently loaded onto the GeoMx DSP instrument and scanned producing a digital fluorescent image of the tissue. Individual ROIs were then selected. ROIs were selected based on the presence of areas of tumour cells, immune cells, and cancer-associated

fibroblasts (CAFs). Six ROIs (3 'immune hot' and 3 'immune cold') were selected per tumour resulting in 18 ROIs per slide. ROIs were segmented into 3 AOIs for each 'immune-hot' ROI which were CD45+ (immune), a-SMA+ (CAF) and PanCK (tumour) compartments and 2 AOIs per 'immune-cold' ROI which were a-SMA+ (cancer-associated fibroblast) and PanCK (tumour) compartments. Following selection of the ROIs and segmentation into AOIs the DSP barcodes were UV cleaved and collected. The photocleaved oligos were collected via microcapillary aspiration and dispensed into a 96-well plate. During the library preparation, the DSP barcodes were tagged with their ROI location and subsequently sequenced on an Illumina sequencer. The DNA oligonucleotide sequences contained ROI indices which mapped them back to their tissue location, an RNA target identification sequence which matched them to their ISH probes, and a unique molecular identifier (UMI) to deduplicate reads. Sequenced oligonucleotides were processed and imported back into the GeoMx DSP analysis software for integration with the slide images and ROI selections for spatially-resolved RNA expression.

Bioinformatics pre-processing was performed using the NanoString GeoMx DSP analysis software. ROIs with < 5% detected genes were removed. Genes that were detected in < 5% of ROIs were also removed. ROI-level whole transcriptome profiles were upper-quartile (Q3) normalised and subsequently used for statistical analysis. For statistical analysis, gene expression data were modelled using linear mixed effect models with patient as a random effect (R package lmerTest v3.1-3) and variable of interest (e.g. immune hot or cold) as a fixed effect. Deconvolved cell type estimates in METABRIC and TCGA breast cancers were calculated using consensusTME [355]. This was performed by Syed Haider (BCR Bioinformatics team).

2.14 Macrophage co-culture experiments

THP-1 cells, a human leukaemia monocytic cell line, were cultured in suspension in RPMI+10% FBS. ILC cell line MDA-MB-134 previously tagged with red fluorescent protein and luciferase (MM134-RFP-Luc2) were cultured as previously described [356]. THP-1 cells were differentiated by the addition of phorbol 12-myristate 13-acetate (PMA) (Sigma, UK) 10ng/ml for 24hrs. Cells were then polarised by the addition of IFN-y (20ng/ml) and LPS (100ng/ml) into M1-like and IL-4 (20ng/ml) and IL-13 (20ng/ml) to M2-like macrophage phenotypes (cytokines from Peprotech, UK) [357]. Unpolarized macrophages (M0) were also used as a control.

ILC cell line MDA-MB-134 previously tagged with red fluorescent protein and luciferase (MM134-RFP-Luc2) were cultured as previously described [356]. MM134-RFP-Luc2 and polarised THP-1 cells were collected by using cell scrapers to detach cells from the 10cm dishes and 1 × 10^6 cells resuspended in 1 ml of serum free media. THP-1 cells were tagged with a green lipophilic dye (cat no: KH67GL-1KT) and combined with ILC cells in a ratio of 2000:5000 in 30µl/well of a 96-well Ultra Low Attachment plate (Corning, UK). The plates were centrifuged at 300g for 5 minutes and incubated at 37°C and with 5% atmospheric CO_2 for 5 days. Spheroids were imaged at 12-hour intervals on the Incucyte® S3 System. N = 3 biological replicates were performed.

2.15: Single nuclei sequencing of MM134-RFP-Luc2 primary mouse mammary glands and metastases

2.15.1 Generation of primary and metastatic MM134-RFP-Luc2 ILC lesions

All the experiments were performed in accordance with Swiss guidelines for animal safety by George Sflomos (ISREC-Swiss Institute for Experimental Cancer Research). NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ and NSG-EGFP (JAX stock #021937) mice were purchased from Jackson Laboratories and further expanded in EPFL with a 12-h-light-12-h-dark cycle, controlled temperature and food and water ad libitum. Eight to 12-week-old female mice were anesthetized by intraperitoneal injection with xylazine 10 mg/kg and ketamine 90 mg/kg (Graeub AG) and intraductally injected into the 3rd and the 4th pair of glands with 5-8 ul of PBS containing 400,000 cells. Luciferase-based imaging was performed with Xenogen IVIS Imaging System 200 (Caliper Life Sciences) in accordance with the manufacturer's protocols and used to monitor individual mammary glands. At sacrifice (11 months 20 days after MIND injection), engrafted mammary glands were harvested, and carcinomas were viably frozen. In brief, freshly surgically resected xenograft tissues (1-2 mm3 pieces) were transferred into 1ml prewarmed (37 degrees) freezing mix of 95% Fetal Bovine Serum and 5% DMSO in a cell culture cryovial and shipped to ICR on dry ice.

2.15.2 10x sample preparation and sequencing

Frozen tissue was weighed and cut into small pieces using a chilled razor blade on dry ice and transferred into a 7ml douncer (Millipore Sigma, cat. no. D9063) on ice. 1 mL of cold lysis buffer comprising 0.32M Sucrose, 5mM $CaCl_2$, 3mM Mg $(Ac)_2$, 20mM Tris-HCl pH 7.5, 0.10%

Triton X-100, 0.1mM EDTA pH 8.0, DEPC Water, 40U/mL RNase Inhibitor (Millipore Sigma, cat. No. 3335402001) was added to the douncer for 5-minute lysis. The tissue was homogenized tissue with 2 - 5 strokes with the loose pestle (pestle A) and 2 - 5 strokes with the tight pestle (pestle B). After the 5-minute lysis, nuclei were pelleted by spinning at 4^0C for 10 minutes at 800 x g. The supernatant was removed, and the nuclei pellet was resuspended with 2 ml of wash buffer comprising 1 x PBS, 1% BSA, and 0.1U/uL RNase Inhibitor. Centrifugation and resuspension with wash buffer was repeated for a total of 2 washes. After the second wash, the nuclei suspension was filtered using a 40um Flowmi cell strainer (Millipore Sigma, cat. no. BAH136800040).

The nuclei suspension was stained with 4′,6-diamidino-2-phenylindole (DAPI) for (Fluorescence-activated cell sorting) FACS sorting. DAPI positive nuclei was collected and debris and nuclei aggregates within the DAPI positive gating was excluded. After FACS, sorted nuclei was spun down at 4C for 10 minutes at 800 x g. Nuclei counting was done using a haemocytometer (Thermo Fisher, cat. no. 22-600-100). Following counting, the appropriate volume for each sample was calculated for a target capture of 10,000 cells and loaded onto 10x single cell G chip. After droplet generation, samples were transferred to a pre-chilled PCR strip tube (MJS BioLynx, cat. no. US14024700) and incubated overnight in a Veriti 96-well thermal cycler (Thermo Fisher). The next day, sample cDNA was recovered using Recovery Agent provided by 10x and subsequently cleaned up using a Silane DynaBead mix as outlined by the user guide 3' Reagent Kits v3.1. Purified cDNA was amplified for 11 cycles before being cleaned up using SPRIselect beads (Beckman Coulter, cat. no. B23318). Samples were run on a Bioanalyzer (Agilent Technologies) to determine cDNA concentration. cDNA libraries were prepared as outlined by the Single Cell 3' Reagent Kits v3.1 user guide with modifications to the PCR cycles based on the calculated cDNA concentration.

The molarity of each library was calculated based on library size as measured bioanalyzer (Agilent Technologies) and qPCR amplification data (Roche, cat. no. 07960140001). Samples were sequenced on Illumina's NovaSeq 6000 with the following run parameters: Read 1 – 28 cycles, read 2 – 90 cycles, index 1 – 10 cycles and index 2 – 10 cycles.

2.15.3 Data analysis

The data was analysed by Syed Haider and Fatemeh Ahmadi Moughari (BCR Bioinformatics team). Reads were mapped to the combined human hg38 and mouse mm10 reference genome using CellRanger V6 [358] to quantify the reads for each of the cells. The cells were labeled as human & mouse cells based on their proportion of reads mapped to either of

species. The downstream analysis was performed in 3 main steps: preprocessing, individual sample analysis and cross-sample analysis. In the preprocessing step mouse cells were removed, to focus on the human ILC tumour cells. Cells with less than 1500 or more than 32000 reads (Unique Molecular Identifier - UMI counts) were removed to remove outliers. Cells with high number of reads mapping to the mitochondrial genome were removed. Doublet cells were detected using both scDblFinder and DoubletFinder methods, and cells that were called as the doublet by both methods were removed. Genes that were not expressed in at least 90% of cells (zero-inflated genes) were removed.

Each sample was then analyzed using Seurat R package V4 [359] and were normalized using SCTransform V1 [360]. The top 50 principal components were used for Uniform Manifold Approximation and Projection (UMAP) visualization and clustering. Clustering was performed using the Louvian algorithm [361] with default parameters. To find the shared transcriptional patterns between samples, the CIDER meta-clustering algorithm was used [362]. This approach was chosen as previous studies have widely used batch effect correction-based integration to find clusters that are shared between samples. However, applying batch effect correction methods to integrate only epithelial cells removes true biological signals and over-corrects the data. It has been shown that batch correction methods rarely lead to analysis improvement [363]. To overcome this CIDER, a meta-clustering method that does not require batch effect correction was used. Instead, it uses batch covariate modelling to compute similarities between initial clusters [362]. It computes the pairwise similarity of the initial clusters and runs the hierarchical clustering on the similarity matrix to define the meta-clusters; i.e. each meta-cluster is composed of a group of initial clusters that are very similar to each other and lass similar to the rest.

Next to gain further insights into the gene expression profiles of each individual CIDER meta-cluster, the markers of the individual meta-clusters were defined by the aggregation of markers of the constituent subclusters. Firstly, the markers of the subclusters were computed by running differential gene analysis on each individual sample, using the Wilcoxon rank test with a fold change threshold of 1.25 and an adjusted p-value < 0.05. The markers from all subclusters of a meta-cluster were then aggregated using an adjusted combined p-value using Benjamini-Hochberg [364]. Genes with adjusted combined p-values < 0.01 were considered and ranked as the top markers of the meta-clusters.

Chapter 3: Identification of molecular drivers of clinically aggressive ILCs

3.1 Background

ILC as a special breast cancer subtype is recognised for its propensity to cause late disease recurrence whereby patients relapse several years after primary diagnosis. This is attributed to the presence of dormant disseminated cancer cells which are distributed away from the primary tumour and exist as minimal residual disease following primary surgery. They have the ability to rest dormant for several years and during this time they can develop genetic and epigenetic features which subsequently result in aggressive tumour growth and overt metastases at distant sites [65]. However despite this, there is a subset of ILC patients who are treated for primary disease but relapse early (< 3 years of initial diagnosis) despite standard of care treatment as well as patients who present with metastatic disease (*de novo* metastases). These patients have clinically aggressive disease and limited treatment options, representing a clinically unmet need. The molecular drivers of the aggressive disease biology are poorly understood in these patients. Moreover within ILC itself, the rare histological subtype pleomorphic ILC is associated with clinically aggressive disease with numerous studies associating pleomorphic histology with a range of poor prognostic features and early disease recurrence [69-72, 77, 125].

Despite advances in the understanding of the molecular features of ILC as a special breast cancer subtype, as well as rarer subtypes such as pleomorphic ILC, histological analysis remains the most frequently and widely used method for the diagnosis of these tumours. A major challenge which exists within the field of diagnostic pathology including the diagnosis of ILC, is inter-observer variability. This refers to the degree of variation in the diagnostic interpretations of two or more independent pathologists when the same case is analysed [365]. Studies suggest that inter-observer variability appears to be greater in cancer diagnostics compared to non-neoplastic lesions [366, 367]. Moreover, the degree of variability can depend upon the tumour type and frequency. Interestingly breast and gynaecological tumours and less common tumour types show greater levels of inter-observer variability compared to skin, gastrointestinal tumours and more frequently occurring tumour subtypes [367-369]. Given that pleomorphic ILC is a breast tumour, and an uncommon tumour accounting for less than 1% of all invasive breast cancers, which general pathologists are unlikely to be presented with on a regular basis, the degree of inter-observer variability in the diagnosis of this rare tumour type

would be relatively high, although no large studies have specifically assessed this. One small study assessed inter-observer variability in the diagnosis of 5 pleomorphic ILCs among 17 pathologists with varying levels of experience and showed that the level of agreement for pleomorphic ILC failed to exceed substantial concordance [370]. Given the clinically more aggressive nature of pleomorphic ILC compared to the more common classic ILC, the correct diagnosis of these lesions remains important, as it identifies an ILC patient subset who may require closer follow-up or more intensive or prolonged treatment.

Over recent years there has been progress and increased interest in the use of digital pathology and Artificial Intelligence (AI) techniques which provide new opportunities for addressing the challenge of inter-observer variability in diagnostic pathology. AI approaches use diagnostic algorithms to contribute to histological diagnosis in a range of malignancies [371, 372]. Such approaches typically rely on supervised machine learning, whereby software is first trained to recognise specific tissue components. Many manually annotated tumour features are needed for machine learning based classification which can therefore be time-consuming. Whilst digital image analysis techniques are starting to be used in diagnostic pathology, they have been adopted more widely in radiology [373]. The low levels of digitization of the diagnostic workflow in many pathology laboratories, together with a reluctance from some pathologists to replace the traditional microscope is the reason for this [374]. However with the development of digital slide scanners, which enable fast whole slide scanning, digital pathology is becoming adopted in a growing number of pathology departments [372, 375]. Digital image analysis when incorporated into the standard histopathology workflow, together with the use of AI techniques, helps to overcome inter-observer variability and standardise the diagnostic process, challenging the use of the human eye alone as the gold standard. QuPath is an example of a bioimage analysis software tool for digital pathology image analysis, providing a platform with a range of algorithms and image analysis applications for the study of complex and heterogenous tissue images [341]. One such application which QuPath offers, is a measurement of nuclear size across a tumour which can aid in distinguishing pleomorphic from non-pleomorphic ILCs.

ILC has been previously characterised at the molecular level as detailed in chapter 1. However studies which have thoroughly assessed distinct ILC subtypes, notably pleomorphic ILC are limited with very small pleomorphic ILC patient numbers. They have previously concluded that in addition to an enrichment of *HER2* alterations, pleomorphic ILCs show higher rates of mutations in *TP53*, *KMT2C*, *ISR2, MAP3K1*, *NCOR1*, *NF1*, *TBX3*, *ARID1A* and *ARID1B* compared to classic ILC [101, 103, 293]. When metastatic ILC on the whole is considered, including all ILC histological subtypes, hybrid-capture-based genomic profiling of 180 ILC

matched primary and metastatic biopsies identified an enrichment of *ESR1* alterations in ILC metastases in relation to the primary breast specimens, as well as an enrichment of *NF1* alterations in the metastases [292]. A further study using a multicentric series of matched primary and metastatic samples from 94 ER+ ILC patients identified mutations (*AKT1, ARID1A, ESR1, ERBB2*, or *NF1*) as well as CNAs (*NF1* deletions, *PTEN* deletions, *CYP19A1* amplifications) which were unique to the metastases in 22% (7/32) and 19% (4/21) of ILCs respectively, reflecting molecular heterogeneity and clonal evolution in ILC [290].

At the gene expression level, distinct ILC gene expression subtypes have been identified in the Rational Therapy for Breast Cancer (RATHER) consortia and TCGA cohort (n = 106 and n = 144 respectively) [32, 277]. Two and three gene-expression subtypes were identified in these cohorts (RATHER: immune-related and hormone-related subtypes TCGA: immune-related, reactive, proliferative subtypes). Interestingly both included an 'immune-related' subtype as further detailed in chapter 1. However, these two separate 'immune' subtyping approaches failed to identify the same ILC cases when they were applied to the same dataset [282]. Moreover there was no significant difference in survival outcomes between the two RATHER gene-expression subtypes [277].

So far published studies have not explicitly profiled pleomorphic tumours. There is therefore a need to understand drivers of clinically aggressive disease among ILCs at both the genomic and transcriptomic level to potentially identify high-risk ILC patients.

3.2 Chapter Aims

Overall the purpose of this chapter is to:

1) Characterise somatic driver mutations in aggressive ILCs.

2) Characterise the transcriptomic features associated with aggressive ILC.

3.3 Results

3.31 Pleomorphic ILCs show an association with early disease recurrence

In the first instance, a cohort of 64 cases classed as ILC from the KHP tissue bank were reviewed histologically by two independent pathologists (Dr Ioannis Roxanis and myself). Most of the histology reports for these cases classed them as 'lobular carcinomas' without further details on ILC subtype histology. This reflects the age of the samples, since many of the cases were from the early 90s and refinements to the pathology reporting classifications have occurred over time such that nowadays more detailed classifications for ILC exist (see chapter 1). The cases selected were relapsing ILC patients and the initial aim was to assess any association between the onset of relapse (early vs late) and the level of immune infiltrate.

Upon histological review, 10/64 (15.6%) showed pleomorphic histology which is consistent with reported incidence rates of pleomorphic ILC [77, 118]. The remainder of cases were non-pleomorphic, and mainly classed as classic ILC (Supplementary Table 1). One pleomorphic patient presented with *de novo* metastatic disease as did 6 non-pleomorphic cases. After excluding the *de novo* metastatic patients, when the time from primary diagnosis to disease recurrence was compared in pleomorphic (n = 9) vs non-pleomorphic patients (n = 48), the pleomorphic patients showed significantly shorter periods between primary disease and metastasis (p = 0.0390, Mann-Whitney U test, Figure 3.1). When a 3 years cut off was used to define early relapse, a higher proportion (4/9 - 44.4%) of pleomorphic patients relapsed early compared to non-pleomorphic (11/48 - 22.9%) patients although this difference was not statistically significant (p = 0.223, Fisher's exact test).

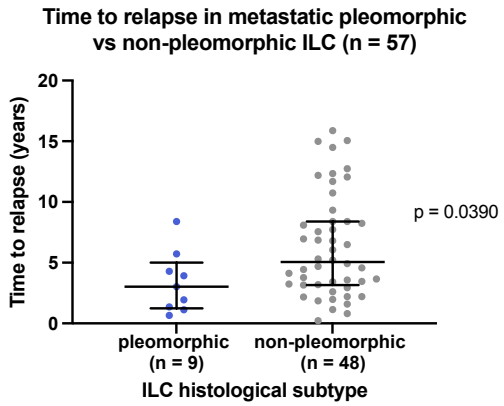


Figure 3.1: Pleomorphic ILCs relapse earlier than non-pleomorphic ILCs: Scatterplot showing time between primary diagnosis and disease recurrence in pleomorphic (n = 9) vs non-pleomorphic (n=48) ILC (p = 0.0390, Mann-Whitney U test).

3.32 Digital histopathology assessment in an extended ILC cohort

Having identified in a small metastatic ILC cohort that pleomorphic histology appears to be associated with shorter time intervals between primary diagnosis and recurrence and given that pleomorphic ILC is a rare and understudied tumour type, a larger cohort of pleomorphic ILCs from patients with and without disease recurrence was subsequently requested and reviewed from the KHP tissue bank as well as pleomorphic ILC cases collected through the GLACIER study. In addition, further non-pleomorphic ILC cases from patients without disease recurrence were also acquired from the KHP tissue bank (Supplementary Table 1). In total an ILC study cohort of 163 patients with a range of different clinical outcomes was reviewed histologically as summarised in Table 3.1. Throughout the project, these cases were subject to a range of different histological and molecular based assessments (Figure 3.2).

A range of different ILC subtypes were identified, including classic (52.1%), pleomorphic (33.7%), mixed - without a pleomorphic component (5.52%), mixed with a pleomorphic component (4.91%), solid (1.84%) and alveolar (1.84%). Consistent with the WHO diagnostic criteria, the classic cases were characterised by small, uniform discohesive cells, with round or notched oval nuclei and occasional intra-cytoplasmic lumina, arranged as single files or individually dispersed cells throughout the stroma with minimal disturbance of the normal tissue architecture, or desmoplastic reaction (Figure 3.3 A, B, D). Associated LCIS was occasionally seen (Figure 3.3 C). In contrast pleomorphic cases were identified based upon a greater degree of cellular atypia and pleomorphism than the classic and other subtypes. They showed nuclear enlargement and greater variability in the size and shape of the nuclei, with the presence of nuclei > 4 times the size of lymphocytes, hyperchromasia and occasional prominent nucleoli (Figure 3.3 E - F). Pleomorphic LCIS, more frequent mitoses and lymphovascular invasion were also features identified in cases classed as pleomorphic ILC (Figure 3.3 G - H). The rare solid and alveolar cases were identified by the presence of solid sheets and globular aggregates of tumour cells respectively (Figure 3.3 I - J). Mixed lesions showed an admixture of the classic type with one or more additional subtypes (Figure 3.3 K - L).

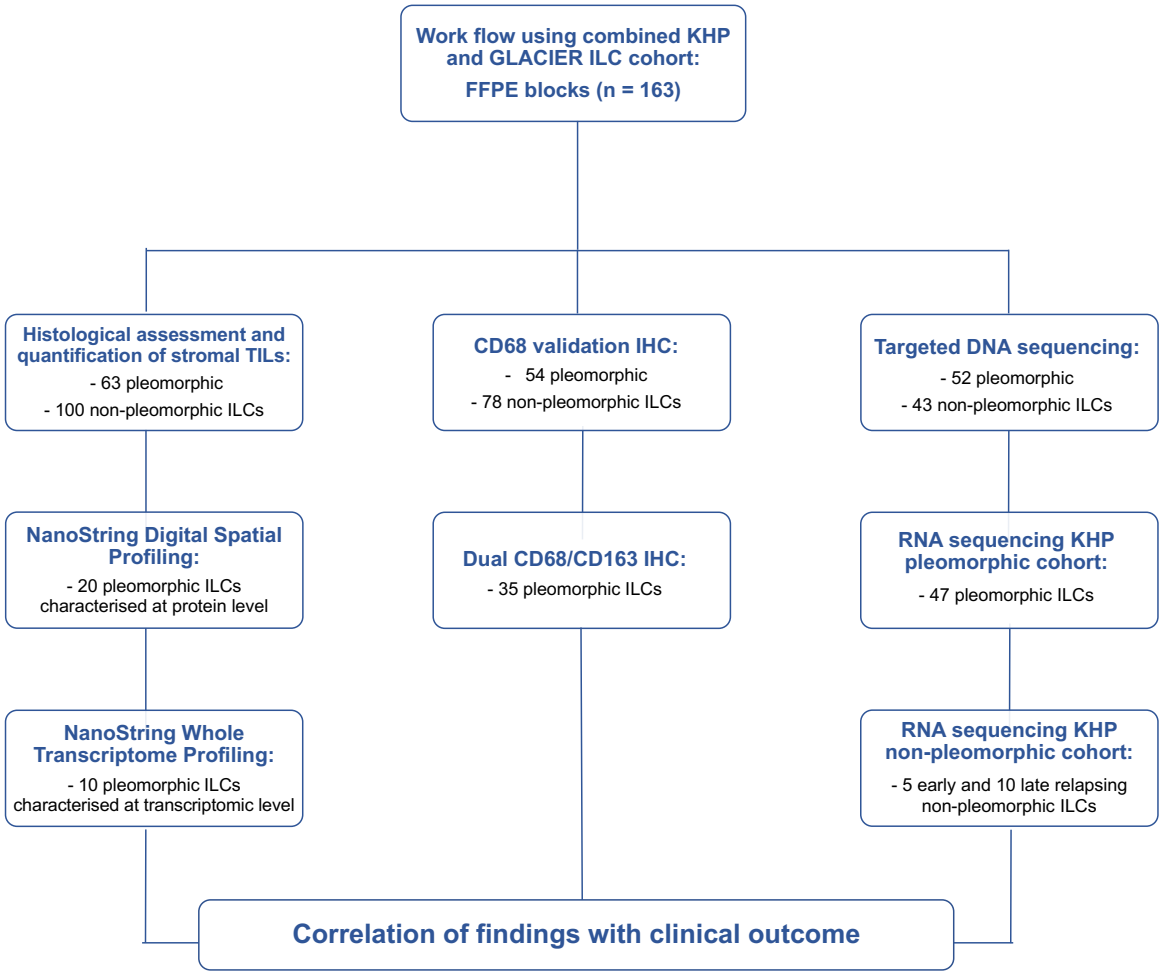


Figure 3.2: Overview of study workflow: Consort diagram summarising study workflow using combined KHP and GLACIER ILC cases (n = 163).

Parameter	Number (n)
ILC tumour samples	163
Histological subtype	
Classic ILC	85
Pleomorphic ILC	55
Mixed ILC without pleomorphic component	9
Mixed ILC with pleomorphic component	8
Solid ILC	3
Alveolar ILC	3
Average patient age (years)	59.8
Disease relapse status at 15-year follow-up	
No Relapse	76
Relapse	83
< 3 years of primary diagnosis	34
3 - 6 years after primary diagnosis	22
> 6 years after primary diagnosis	26
Onset unknown	1
Inadequate follow-up	4
Tumour size	
≤ 20 mm	39
> 20 mm, ≤ 50 mm	92
> 50 mm	31
Unknown	1
Number of involved lymph nodes	
0	74
1 – 3	41
4 – 9	27
≥ 10	19
Sites of metastases	
Bone	51
Liver	18
Pleura	18
Skin	15
Peritoneum	11
Lung	10
Brain	3
Meningeal	2
Mediastinal	2
Brachial Plexopathy	2
Other / Unknown	20
Tumour grade	
I	3
II	97
III	27
Unknown	36
Oestrogen (ER) Status	
ER+	151
ER-	8
Unknown	4
Progesterone (PR) Status	
PR+	102
PR-	48
Unknown	13
HER2 Status	
HER2+	11
HER2-	127
Unknown	25

Table 3.1: Clinicopathological features of the extended ILC cohort

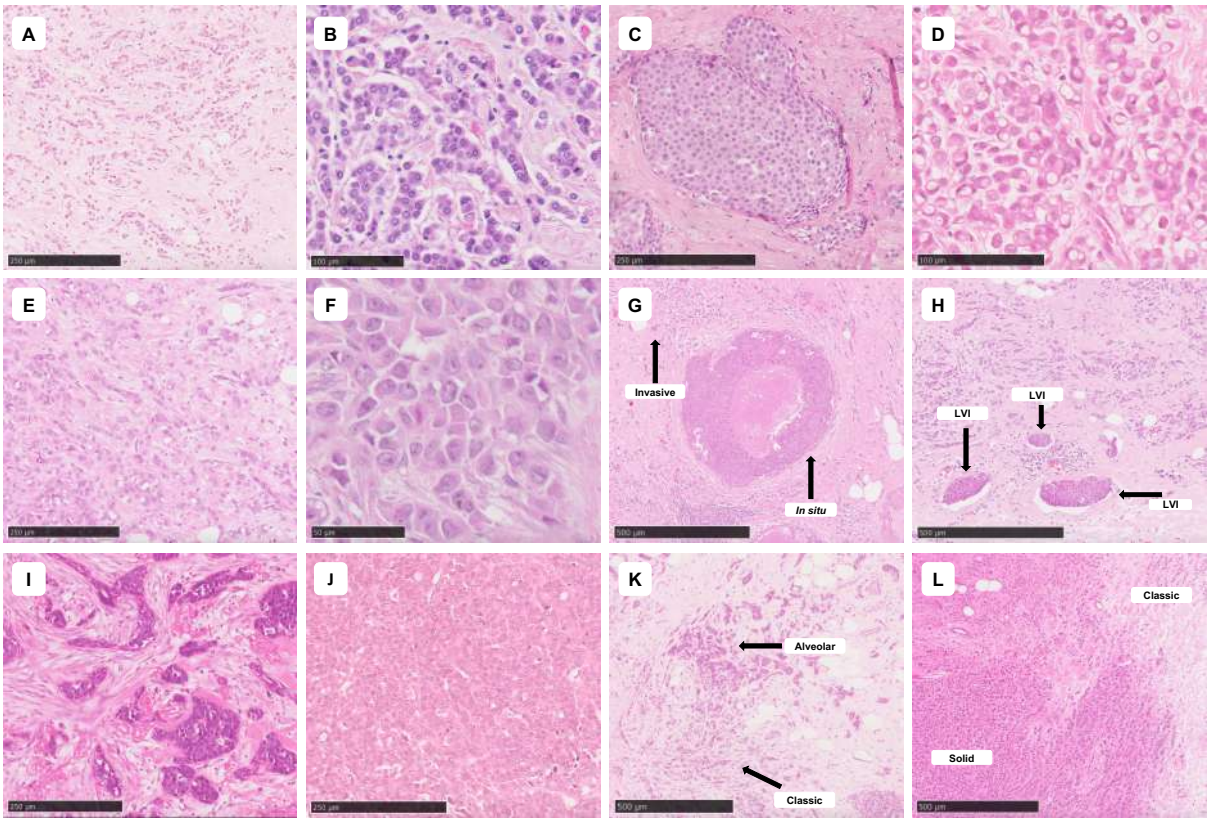


Figure 3.3: Histological subtypes and features of the ILC cohort: A) Classic ILC case at low power showing characteristic single file pattern B) Classic ILC case at high power showing uniform cells with small round nuclei C) Classic LCIS identified in a classic ILC case D) Classic ILC case where tumour cells show abundant intracytoplasmic lumina E) Pleomorphic ILC case at low power showing single file and single cell infiltrative pattern with greater cellular atypia than classic ILC F) Pleomorphic ILC case at higher power showing nuclear pleomorphism and prominent nucleoli G) Pleomorphic LCIS with central necrosis and surrounding invasive pleomorphic ILC H) Lymphovascular invasion (LVI) identified in a pleomorphic ILC case I) Alveolar ILC showing aggregates of tumour cells J) Solid ILC K) Mixed ILC without a pleomorphic component showing classic and alveolar areas L) Mixed ILC with a pleomorphic component showing solid and pleomorphic areas.

To objectively assess and confirm the histological diagnoses made of pleomorphic and non-pleomorphic ILC, a subset of 25 pleomorphic and 25 non-pleomorphic cases were further assessed using QuPath [341] (Table 3.2). A total of 200 - 250 representative tumour cells were selected in each case and nuclei were segmented using StarDist, a cell detection algorithm [342]. Cases which were identified histologically as pleomorphic showed a significantly higher nuclear size compared to the subset of cases identified as non-pleomorphic ($p < 0.0001$, Mann-Whitney U test, Figure 3.4 A). Moreover, cases identified as pleomorphic showed greater variability in nuclear size (pleomorphism) with a significantly higher standard deviation in nuclear size across the tumour cells assessed for each of the pleomorphic cases compared to the non-pleomorphic cases ($p < 0.0001$ Mann-Whitney U test, Figure 3.4 B).

Case	Histology (pleomorphic = 1/ non-pleomorphic = 0)	Average nuclear size (μm)	Standard deviation of nuclear size
3117	1	37.37	17.84
9068	1	44.10	25.39
9252	1	47.75	18.60
9894	1	45.07	22.07
TG00506	1	40.31	14.66
TG01139	1	50.87	23.62
TG01189	1	38.72	11.44
TG02222	1	46.79	16.71
7756	1	42.94	17.31
3117	1	36.60	17.10
9350	1	59.66	26.32
17011789	1	41.22	15.44
TG01941	1	44.93	19.93
TG01958	1	50.04	25.02
TG01964	1	64.24	23.45
TG02006	1	40.34	13.86
TG02072	1	44.77	19.42
TG02166	1	43.08	17.07
TG02210	1	43.51	19.44
17015028	1	45.79	23.48
17020462	1	39.93	16.40
17023387	1	79.74	33.64
17037789	1	57.85	24.48
17047324	1	50.63	18.16
17051586	1	53.83	23.85
684	0	25.81	8.99
758	0	21.24	7.33
759	0	30.31	12.16
843	0	27.47	10.69
1190	0	30.68	11.36
1529	0	23.52	8.66
1605	0	34.03	11.17
137	0	35.34	12.71
245	0	27.46	11.77
144	0	27.64	11.04
1867	0	32.01	10.52
2272	0	27.21	9.50
5817	0	31.54	12.07
6389	0	37.61	13.38
6813	0	27.02	8.47
7490	0	38.62	14.80
8785	0	25.86	9.71
9005	0	30.82	11.33
9129	0	26.43	9.85
9353	0	29.40	11.15
9423	0	27.29	12.74
9573	0	25.09	10.94
10293	0	35.32	13.21
17037705	0	26.83	11.01
4148/17020161	0	31.33	10.91

Table 3.2: Nuclear size and standard deviation of nuclear size results for 25 pleomorphic and 25 non-pleomorphic ILCs assessed using QuPath.

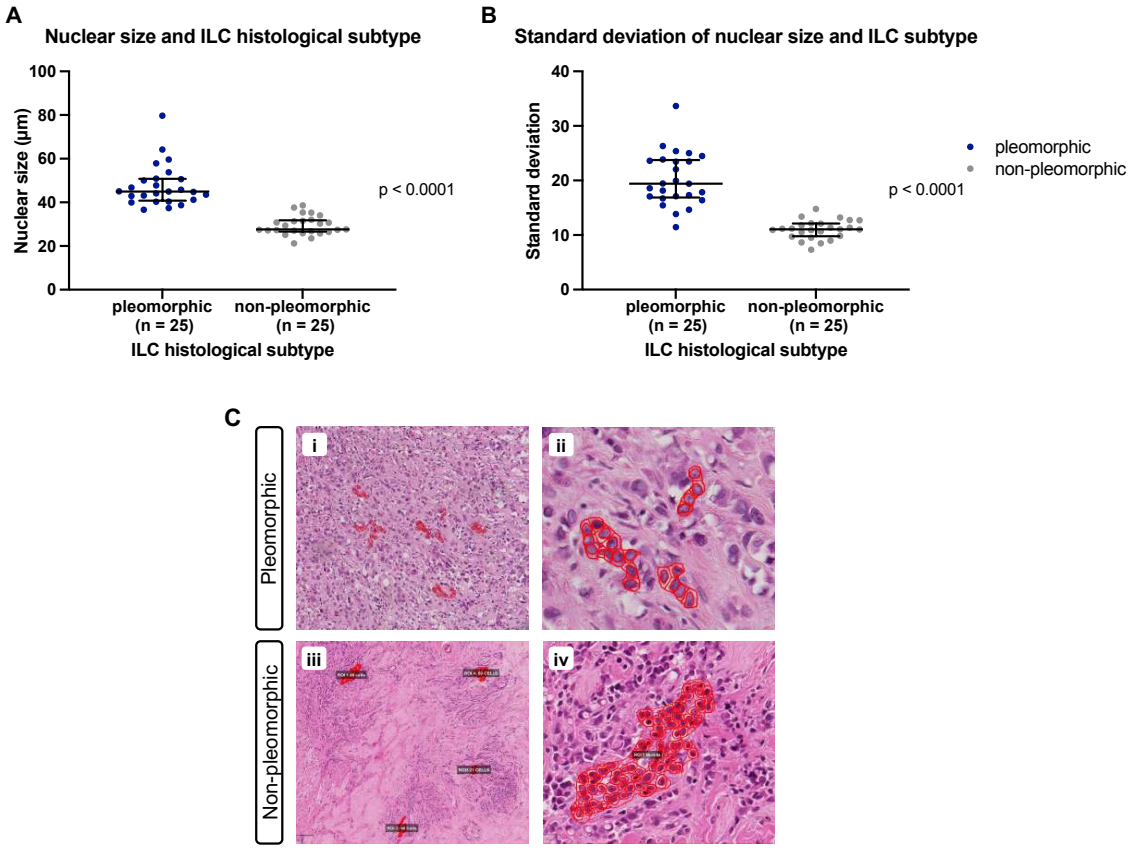


Figure 3.4: Pleomorphic ILCs show increased nuclear size and variability: A) Scatterplot comparing the average tumour cell nuclear size in pleomorphic (n = 25) vs non-pleomorphic (n = 25) ILC showing significantly greater nuclear size in pleomorphic ILC ($p < 0.0001$, Mann-Whitney U test) B) Scatterplot comparing the degree of variation (standard deviation) in nuclear size in pleomorphic (n = 25) vs non-pleomorphic (n = 25) ILC showing significantly greater nuclear size variation in pleomorphic ILC ($p < 0.0001$, Mann-Whitney U test) C) QuPath selection of tumour cells and assessment of nuclear size in representative pleomorphic and non-pleomorphic cases: top panel shows pleomorphic case at i) low (x 10) and ii) high (x 40) power with cells segmented to identify nuclear area, bottom panel shows non-pleomorphic case at iii) low (x 4) and iv) high (x 30) power with cells segmented to identify nuclear area.

3.33 Pleomorphic ILC is associated with earlier disease recurrence

Following the initial analysis in the original 64 relapsing ILCs which identified an association between pleomorphic ILC histology and earlier onset of disease recurrence, an assessment was next made of this association in the extended ILC cohort having acquired further case numbers for assessment.

Nine patients (2 pleomorphic and 7 non-pleomorphic) presented with *de novo* metastatic disease and were not included in the analysis. Among patients presenting with primary disease, using a 15-year follow-up time, 24 pleomorphic patients and 50 non-pleomorphic patients developed disease recurrence. The median time to recurrence in the pleomorphic group was 2.63 years (IQR: 1.36 - 6.09 years) compared to 4.63 years (IQR: 2.86 - 8.28 years) in the non-pleomorphic group. Overall the pleomorphic patients relapsed significantly earlier than the non-pleomorphic patients (p = 0.0456, Mann Whitney U test, Figure 3.5A). Given that pleomorphic ILCs have higher rates of ER- and HER2+ disease, further analysis was performed excluding ER- and HER2+ cases, and cases for whom this information was unavailable. The median time to recurrence in the remaining 15 ER+/HER2- pleomorphic ILCs was 2.22 years (IQR: 1.36 – 8.40 years) compared to 4.63 years (IQR: 2.86 - 8.28 years) in the 50 ER+/HER2- non-pleomorphic cases. Overall the difference in time to relapse between ER+/HER2- pleomorphic and non-pleomorphic ILCs did not reach statistical significance (p = 0.0970, Mann Whitney U test Figure 3.5B), suggesting that earlier relapse observed in pleomorphic ILC is driven by hormone status rather than nuclear pleomorphism.

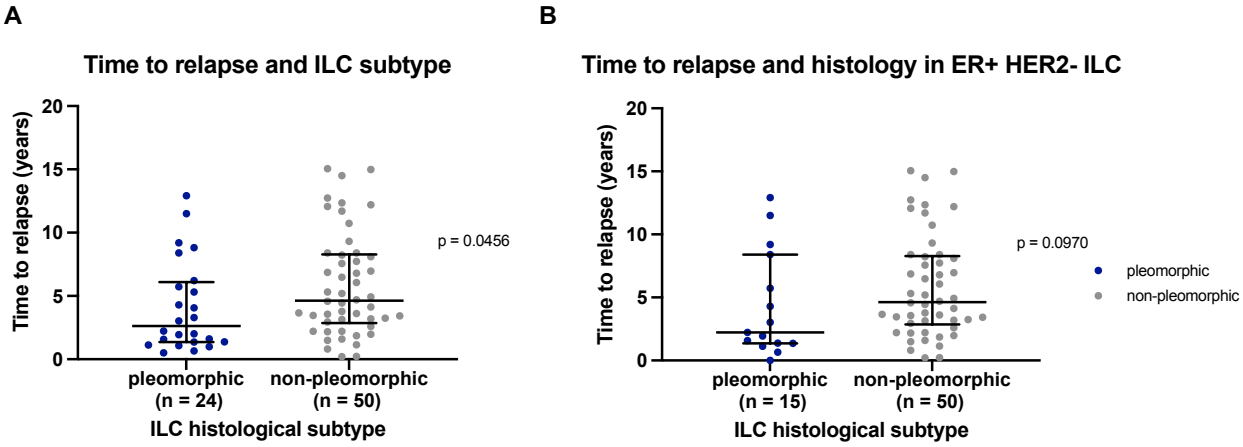


Figure 3.5: Pleomorphic ILC is associated with earlier disease recurrence in relapsing ILC: A) Scatterplot comparing time to disease recurrence in relapsing pleomorphic (n = 24) vs non-pleomorphic (n = 50) ILCs (p = 0.0456, Mann-Whitney U test) B) Scatterplot comparing time to disease recurrence in ER+/HER2- relapsing pleomorphic (n = 15) vs non-pleomorphic (n = 50) ILCs (p = 0.0970, Mann-Whitney test).

3.34 Pleomorphic ILC is distinct at the genomic level

Ninety-five samples (52 pleomorphic and 43 non-pleomorphic) were sequenced using a targeted, custom-designed capture panel (RMH200 v2 panel) consisting of 233 cancer related genes (Supplementary Table 2). This identified differences between the two ILC histological subtypes at the genomic level with respect to the frequency of mutations and copy number alterations (CNAs) (Figure 3.6, 3.7).

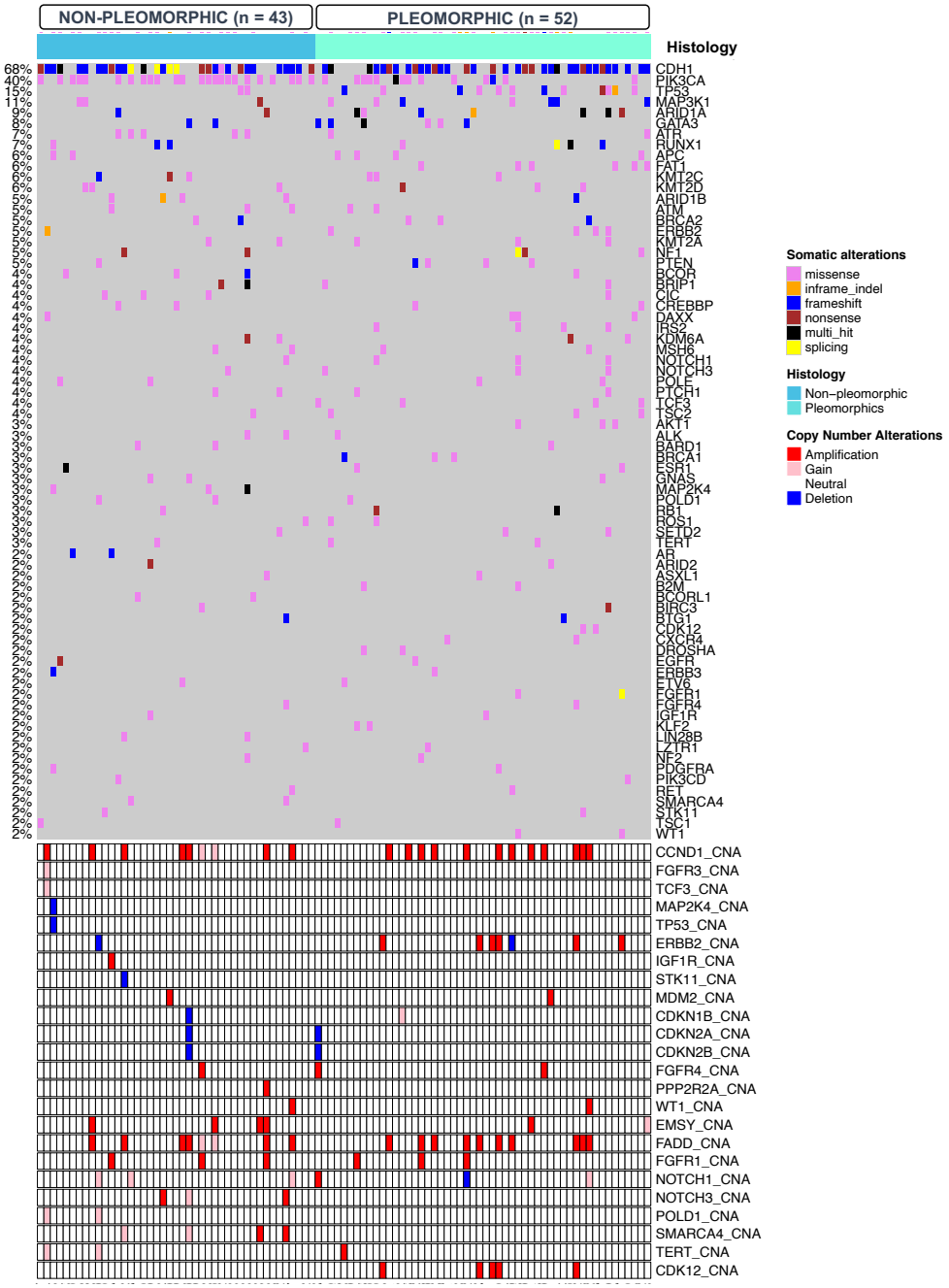


Figure 3.6: Pleomorphic ILC is distinct from non-pleomorphic ILC at the genomic level: Heatmap summarising the mutations and copy number alterations observed in 95 ILC including 43 non-pleomorphic cases and 52 pleomorphic cases.

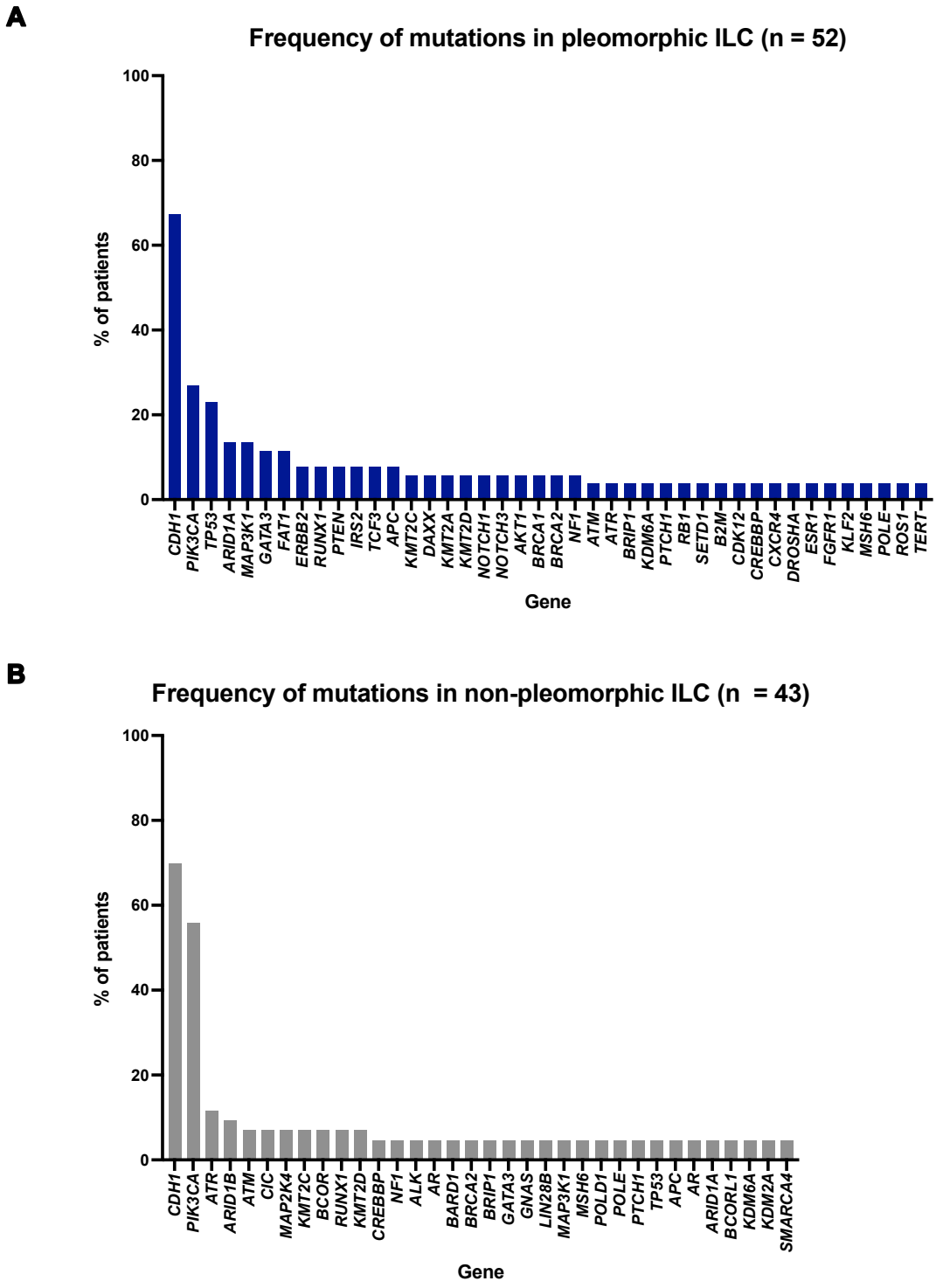


Figure 3.7: Pleomorphic ILC differs from non-pleomorphic ILC in the frequency of mutations observed: Longtail barchart showing the frequency of mutations in A) pleomorphic (n = 52) and B) non-pleomorphic ILC (n = 43) (genes mutated in fewer than two patients in each subtype are excluded).

Pleomorphic ILC had significantly higher rates of *TP53* and *FAT1* mutations and significantly lower rates of *PIK3CA* mutations compared to non-pleomorphic ILC (Table 3.3). Of the 12 pleomorphic patients with *TP53* mutations 9 were either ER- or HER2+ and when the frequency of *TP53* mutations only in ER+/HER2- pleomorphic (n = 32) vs non-pleomorphic (n = 43) ILCs was completed there was no significant difference according to histology (p = 0.645, Fisher's exact test).

Gene	Mutation status	Pleomorphic (n)	Non-pleomorphic (n)	Total (n)	p-value
TP53	mutant	12 (23.1%)	2 (4.7%)	14	0.0179
	wild-type	40 (76.9%)	41 (95.3%)	81	
FAT1	mutant	6 (11.5%)	0 (0%)	6	0.0304
	wild-type	46 (88.5%)	43 (100%)	89	
PIK3CA	mutant	14 (26.9%)	24 (55.8%)	38	0.0061
	wild-type	38 (73.1%)	19 (44.2%)	57	

Table 3.3: Number and percentage of *TP53, FAT1* and *PIK3CA* mutations in pleomorphic vs non-pleomorphic ILC (% represents the percentage of cases within each histological subtype, p-value represents pleomorphic vs non-pleomorphic frequency comparison using the Fisher's exact test).

Consistent with previous studies which report higher rates of *HER2* alterations in pleomorphic vs non-pleomorphic ILC [100-103], when CNAs were considered, *HER2* amplifications were present in 7/52 (13.5%) of pleomorphic cases compared to 0/43 (0%) of non-pleomorphic cases (p = 0.0150, Fisher's exact test, Table 3.4), although no significant difference was observed in the *HER2* mutation frequency. When analysis was performed assessing mutations with a variant allele frequency (VAF) < 5%, a further pleomorphic case harboured a low frequency *HER2* mutation with VAF 1.4%. Additionally, *CDK12* alterations (mutations and/or CNA's) were observed only in pleomorphic ILCs (Table 3.4), with 6/7 patients also harbouring *HER2* amplifications, (Table 3.5).

Gene	Mutation status	Pleomorphic (n)	Non-pleomorphic (n)	Total (n)	p-value
HER2	mutant	4 (7.69%)	1 (2.33%)	5	0.373
	wild-type	48 (92.3%)	42 (97.7%)	90	
HER2	amplified	7 (13.5%)	0 (0%)	7	0.015
	non-amplified	45 (86.5%)	43 (100%)	88	
CDK12	altered*	7 (13.5%)	0 (0%)	7	0.015
	wild-type	45 (86.5%)	43 (100%)	88	

Table 3.4: Number and percentage of *HER2* and *CDK12* alterations in pleomorphic vs non-pleomorphic ILC (% represents the percentage of cases within each histological subtype, p-value represents pleomorphic vs non-pleomorphic frequency comparison using the Fisher's exact test, *altered encompasses *CDK12* amplifications (n = 5) and mutations (n = 2), mutations with VAF > 5% only are included).

	CDK12 amplified / mutant (n)	*CDK12* wild-type (n)	Total (n)
***HER2* amplified / mutant (n)**	6 (85.7%)	4 (8.9%)	10
***HER2* wild-type (n)**	1 (14.3%)	41 (91.1%)	42
Total (n)	7 (100%)	45 (100%)	52

Table 3.5: Summary of overlap between *CDK12* and *HER2* alterations in pleomorphic ILC (% represents the percentage of the column totals i.e. the % of total *CDK12* altered and wild-type cases harbouring or not harbouring *HER2* alterations).

3.34.1 Pleomorphic ILCs have a higher mutational load

Aside from specific mutations alone, an assessment was next made to establish if pleomorphic ILC differs from non-pleomorphic ILC in terms of the total mutational load. The number of mutations (per Mb of the sequencing panel) was calculated for each tumour case and compared between pleomorphic (n = 52) and non-pleomorphic (n = 43) cases. This showed a significantly higher mutational load in pleomorphic ILCs ($p < 0.0001$, Mann-Whitney U test, Figure 3.8A). To control for hormone status, and the fact that pleomorphic ILCs show a higher rate of ER- and HER2+ disease, a further analysis was completed excluding ER- and HER2+ cases. Again a significantly higher tumour mutational load was observed in the ER+/HER2- pleomorphic (n = 32) vs non-pleomorphic (n = 43) ILCs ($p < 0.0001$, Mann-Whitney U test, Figure 3.8B).

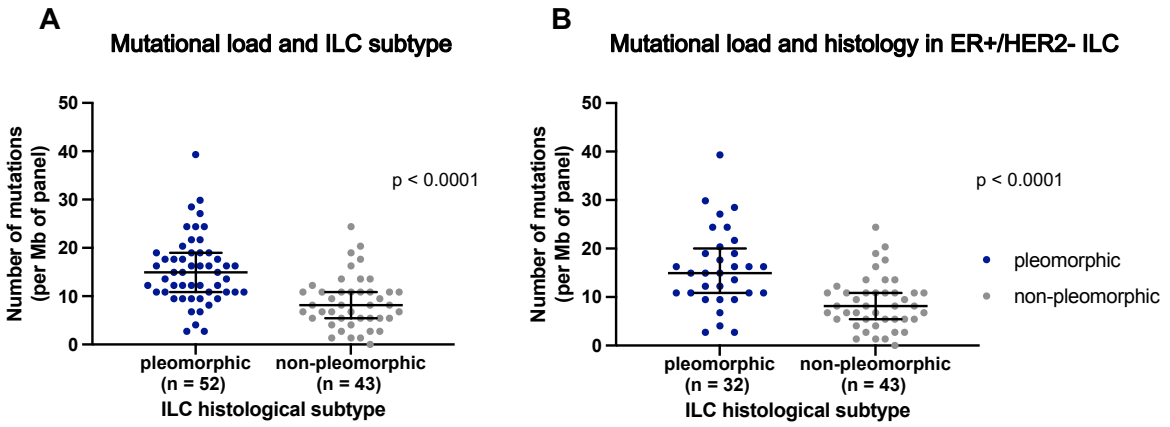


Figure 3.8: Pleomorphic ILCs have a higher mutational load than non-pleomorphic ILC: A) Scatterplot comparing the number of mutations (per Mb of panel) in pleomorphic (n = 52) vs non-pleomorphic ILC (n = 43) B) Scatterplot comparing the number of mutations (per Mb of panel) in ER+/HER2- pleomorphic (n = 32) vs non-pleomorphic ILC (n = 43).

Given that pleomorphic ILC is characterised by a larger nuclear size than non-pleomorphic ILC, and that pleomorphic cases have a significantly higher mutational load than non-pleomorphic ILC, an assessment was next made to establish any direct relationship between the number of mutations and nuclear size in ILC. To do this nuclear size was accurately assessed in the patients with mutational data using QuPath [341]. Some cases had to be excluded due to suboptimal staining on H&E which resulted in inaccurate cell-detection. Overall mutational load and nuclear size were compared in a total of 59 ILCs (26 pleomorphic and 33 non-pleomorphic cases). This showed a statistically significant positive correlation between tumour mutational load and average nuclear size ($\rho = 0.411$, $P = 0.001$, Spearman correlation test).

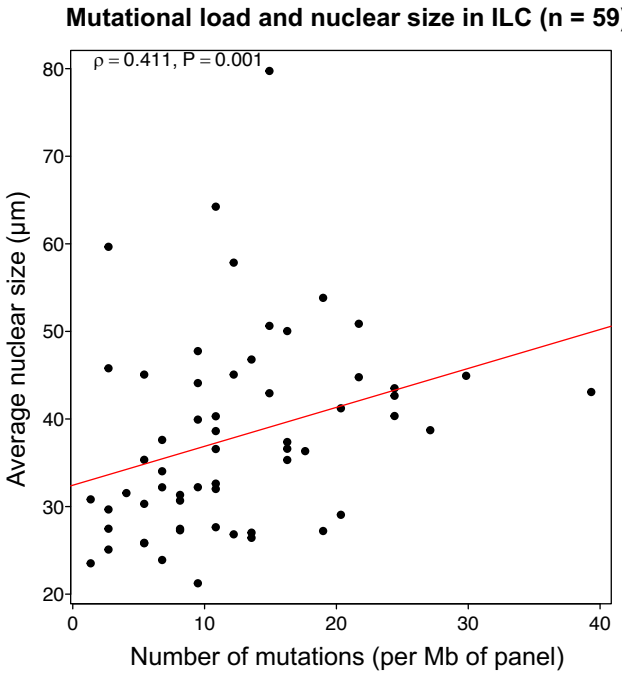


Figure 3.9: Tumour mutational load is positively correlated with nuclear size in ILC: Scatterplot showing a positive correlation between the number of mutations (per Mb of panel) and average nuclear size (assessed using QuPath) in 59 ILCs ($\rho = 0.411$, $P = 0.001$, Spearman correlation test).

3.34.2 *FGFR1* alterations are associated with relapse in pleomorphic ILC

Having identified differences in the genomic landscape based upon ILC histological subtype, an assessment was next made to determine if any specific genomic alterations held prognostic significance in pleomorphic and non-pleomorphic ILC. Within the pleomorphic group, long-term follow-up data was available for 49/52 patients. Of the 49 patients, 24 had disease recurrence and 25 were recurrence-free at censorship. Of the 49 patients with outcome data, *FGFR1* alterations were observed in relapsing patients only. Three relapsing patients had *FGFR1* amplifications and a further 2 patients had *FGFR1* mutations (Figure 3.10, Table 3.6) resulting in an overall rate of *FGFR1* alterations of 20.8% in the relapsing group, compared to an absence of *FGFR1* alterations in non-relapsing patients ($p = 0.0223$, Fisher's exact test). Of the relapsing patients with *FGFR1* alterations, 2 relapsed early (< 3 years), 1 had intermediate relapse (>3 years, ≤ 6 years), 1 had late relapse (> 6 years), and the remaining patient relapsed but the specific date of relapse was unavailable. Aside from *FGFR1*, no other genomic alterations were associated with outcome in the pleomorphic ILCs. Follow-up data was available for 42/43 non-pleomorphic ILCs and there were also no genomic alterations associated with relapse among these patients (Figure 3.11).

	***FGFR1* mutation or amplification (n)**	***FGFR1* wild-type (n)**	**Total (n)**
Relapse (n)	5 (20.8%)	19 (79.2%)	24
No-relapse (n)	0 (0%)	25 (100%)	25
Total (n)	5	44	49

Table 3.6: Number of patients with *FGFR1* alterations and clinical outcome in pleomorphic ILC: (% represents the percentage of the row totals).

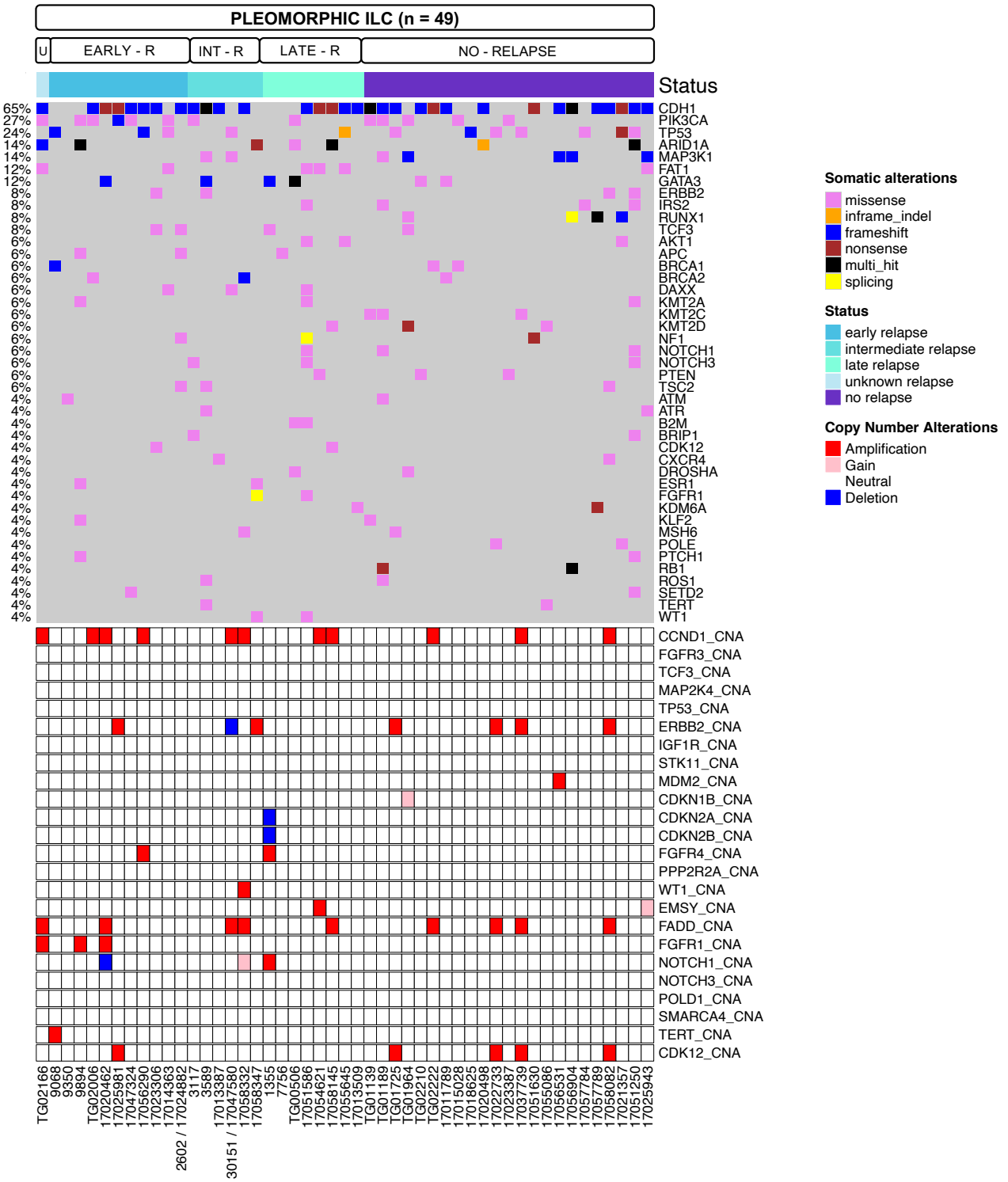


Figure 3.10: *FGFR1* alterations are associated with relapse in pleomorphic ILC: Heatmap assessing association between genomic alterations and clinical outcome in 49 pleomorphic ILCs with clinical outcome data, showing a higher frequency of *FGFR1* alterations in relapsing vs non-relapsing patients (p = 0.0223, Fisher's exact test) (U = unknown onset of relapse, EARLY-R = early relapse, INT-R = intermediate relapse, LATE-R = late relapse).

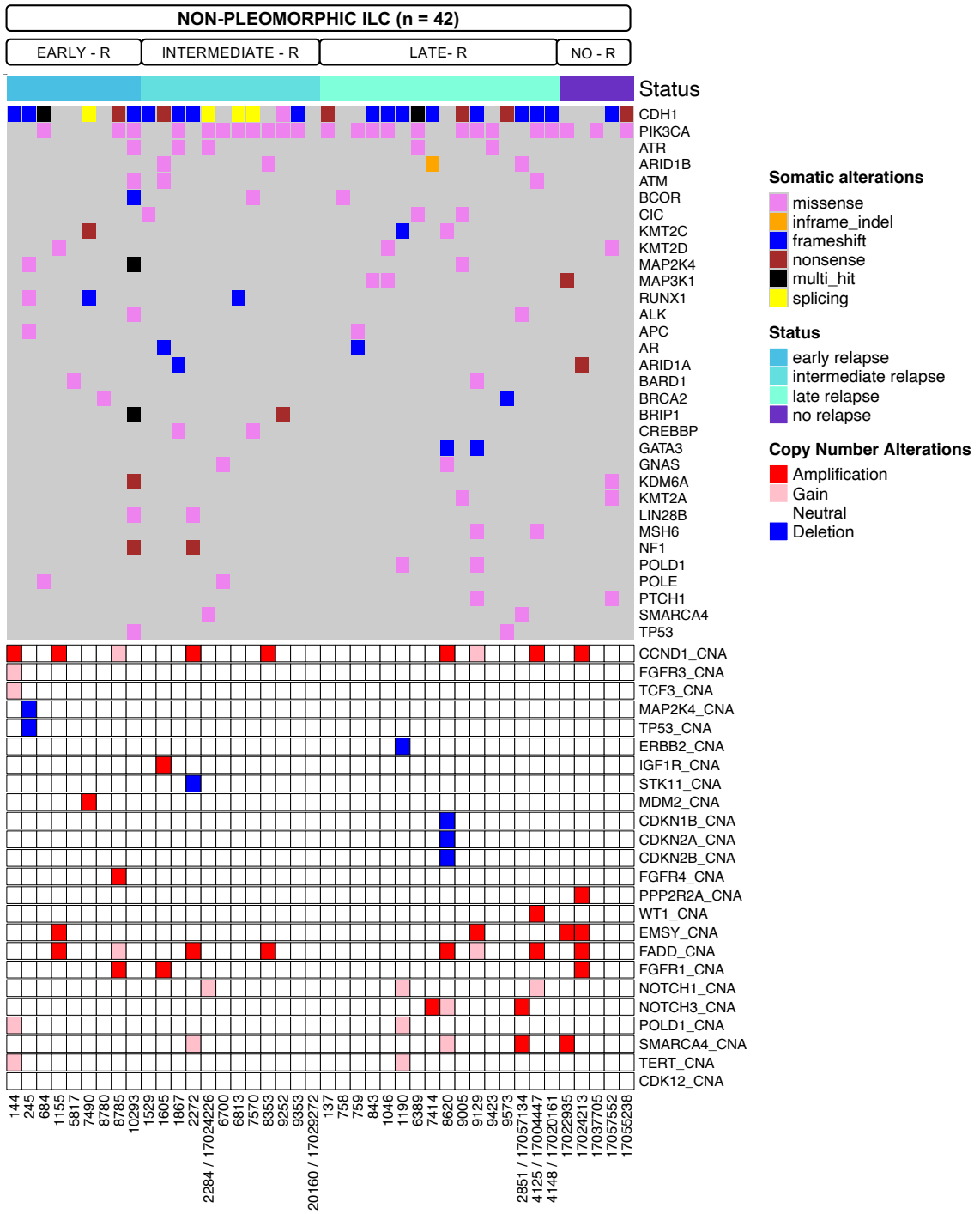


Figure 3.11: No genomic alterations are significantly associated with clinical outcome in non-pleomorphic ILC: Heatmap summarising the mutations and CNAs observed in 42 non-pleomorphic ILCs with clinical outcome data (Early – R = early relapse, INTERMEDIATE – R = intermediate relapse, LATE – R = late relapse, NO – R = no relapse).

3.35 Identification of a prognostic gene expression risk predictor in pleomorphic ILC

Given that aside from *FGFR1* alterations, no other genes appeared to be associated with clinical outcome from the targeted DNA sequencing, an assessment of the transcriptomic landscape of clinically aggressive ILCs was next made to better understand the drivers of aggressive disease biology at the gene expression level. To do this a multi-pronged approach was undertaken (Figure 3.12).

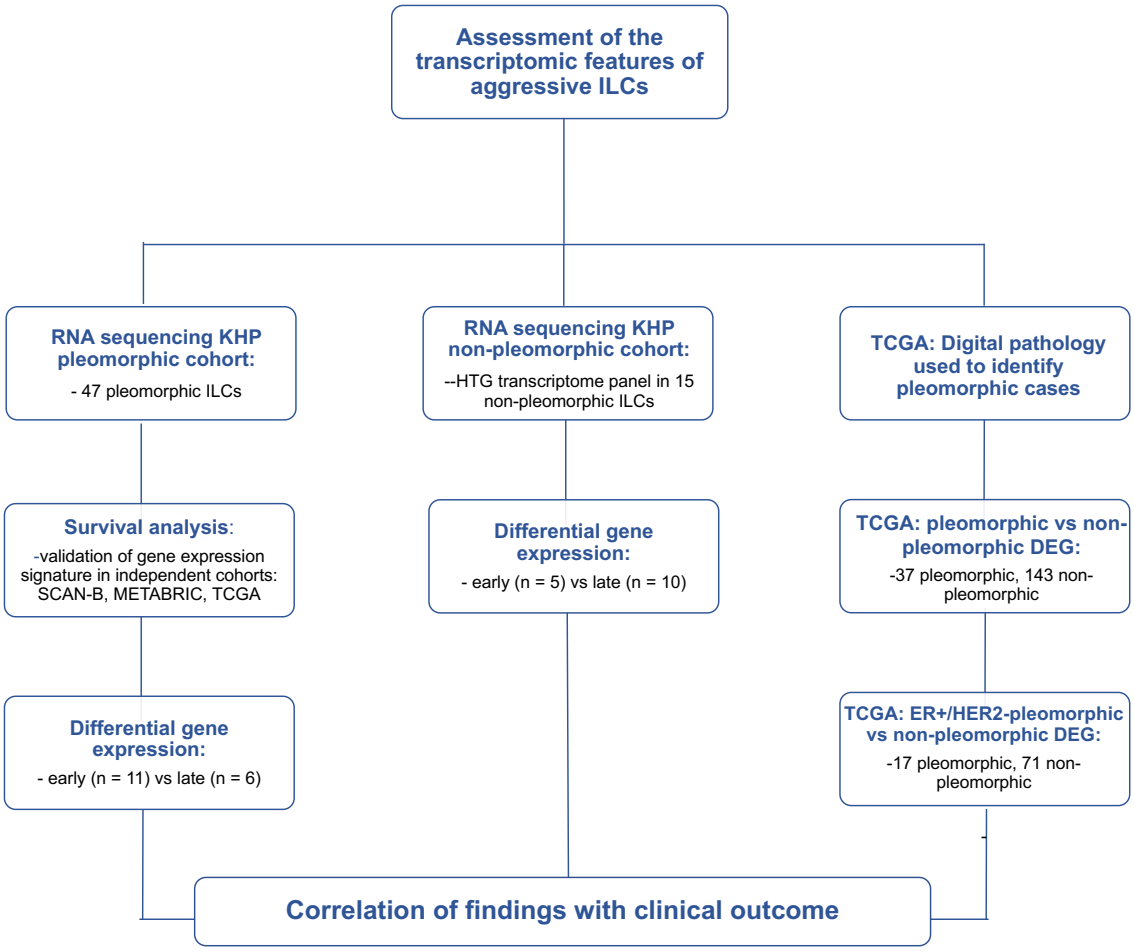


Figure 3.12: Overview of RNA study workflow: Consort diagram summarising study workflow to assess the transcriptomic features of aggressive ILCs.

Given that there are no publicly available RNA datasets derived specifically from pleomorphic ILC, a cohort of 47 pleomorphic ILCs from the KHP ILC cohort was assembled and subject to RNA sequencing (Table 3.7). The tumours were needle macro-dissected from FFPE-derived tissue sections prior to RNA extraction and sequencing to enrich for tumour epithelial-cell content. Univariable survival analysis in the KHP discovery cohort was first performed in 45

samples (2 *de novo* metastatic patients were excluded from survival analysis). This identified 157 and 394 genes which were significantly associated with OS and MFS respectively, using an unadjusted p-value threshold of < 0.05. Of these 134 OS genes and 356 MFS genes were common between the KHP and validation datasets (Supplementary Table 6, 7).

Parameter	Number (n)
Pleomorphic ILC tumour samples	47
Average patient age (years)	59.5
Disease relapse status at 15-year follow-up	
No Relapse	25
Relapse	22
de novo metastatic	2
< 3 years of primary diagnosis	9
3 - 6 years after primary diagnosis	7
> 6 years after primary diagnosis	3
Onset unknown	1
Tumour size	
≤ 20 mm	12
> 20 mm, ≤ 50 mm	27
> 50 mm	8
Number of involved lymph nodes	
0	18
1 – 3	15
4 – 9	9
≥ 10	5
Sites of metastases	
Bone	11
Liver	2
Pleura	2
Abdominal (not liver)	2
Peritoneum	2
Lung	1
Skin	1
Brain	1
Brachial Plexopathy	1
Other / Unknown	4
Tumour grade	
II	23
III	20
Unknown	4
Oestrogen (ER) Status	
ER+	36
ER-	7
Unknown	4
Progesterone (PR) Status	
PR+	29
PR-	11
Unknown	7
HER2 Status	
HER2+	7
HER2-	36
Unknown	4

Table 3.7: Clinical features of KHP pleomorphic ILC RNA cohort

Sixty-seven of the 134 OS genes were also included within the 356 MFS genes. Enrichr [376] was used to study the pathways associated with the 134 OS and 356 MFS genes. Pathways associated with the 134 OS genes identified in the 'WikiPathway 2021 Human' database included TGF-beta receptor signaling and interleukin-1 signaling (Table 3.8). There were no statistically significant pathways identified within the 356 MFS genes using an adjusted p-value < 0.05.

Pathway	Adjusted p-value
TGF-beta Receptor Signaling WP560	0.004273
TGF-beta receptor signaling in skeletal dysplasias WP4816	0.004273
Structural Pathway of Interleukin 1 (IL-1) WP2637	0.02133

Table 3.8: Pathways enriched in 134 OS genes from 'WikiPathway 2021 Human' database

A random forest model was used to validate the genes identified in univariable survival analysis from the KHP discovery cohort. The validation cohorts were TCGA (n = 161), METABRIC (n = 109) and SCAN-B (n = 917). The KHP training set used the gene expression data of the selected genes from univariable survival analysis and each of the sample's survival data. To ensure that the trained model could be applied to the published validation datasets, only the gene expression data common among the KHP discovery cohort and the 3 published datasets was used. Gene expression data from OS and MFS associated genes were trained with OS and MFS outcomes respectively. To obtain the most accurate model, optimization was performed prior to validation in the published datasets (see methods). Once the optimal parameters were identified, the random forest model was trained using these parameters on data from KHP discovery cohort.

Because of the highly pure nature of the KHP discovery cohort due to tissue macrodissection of the tumour cells prior to RNA sequencing, there was a possibility that the random forest modelling prediction would perform poorly on unpurified validation datasets, as they are more heterogenous containing signal from non-tumour cells in addition to tumour cells. To ensure the robustness of the model, *in-silico* purification of the gene expression data in the 3 published validation datasets was performed using the ISOpureR package [354].

The purified gene expression data of the validation datasets were then fed into the random forest model to predict risk scores. associated with each sample in the published datasets. These risk scores estimated risk for each individual and were calibrated to the scale of the number of events of the KHP discovery cohort. Risk scores obtained from the model were then divided into tertiles, ranging from 0 - 33%, 34 - 66% and > 66%. The 3 groups of risk scores were then used in a cox-proportional hazards model to evaluate the ability of predicted

risk scores to accurately determine OS up to a 10-year cut-off period, since only OS survival data was available for all of the validation cohorts. The risk predictor derived from the 356 MFS genes validated as a predictor of OS in TCGA ($p = 0.002$), METABRIC ($p = 0.004$) and SCAN-B ($p = 0.018$) (Figure 3.13 I, iii, v). The risk predictor derived from the 134 OS did not validate as a predictor of OS in TCGA ($p = 0.07$) but did validate in METABRIC ($p = 0.022$) and SCAN-B ($p = 0.015$) (Figure 3.13 ii, iv. vi).

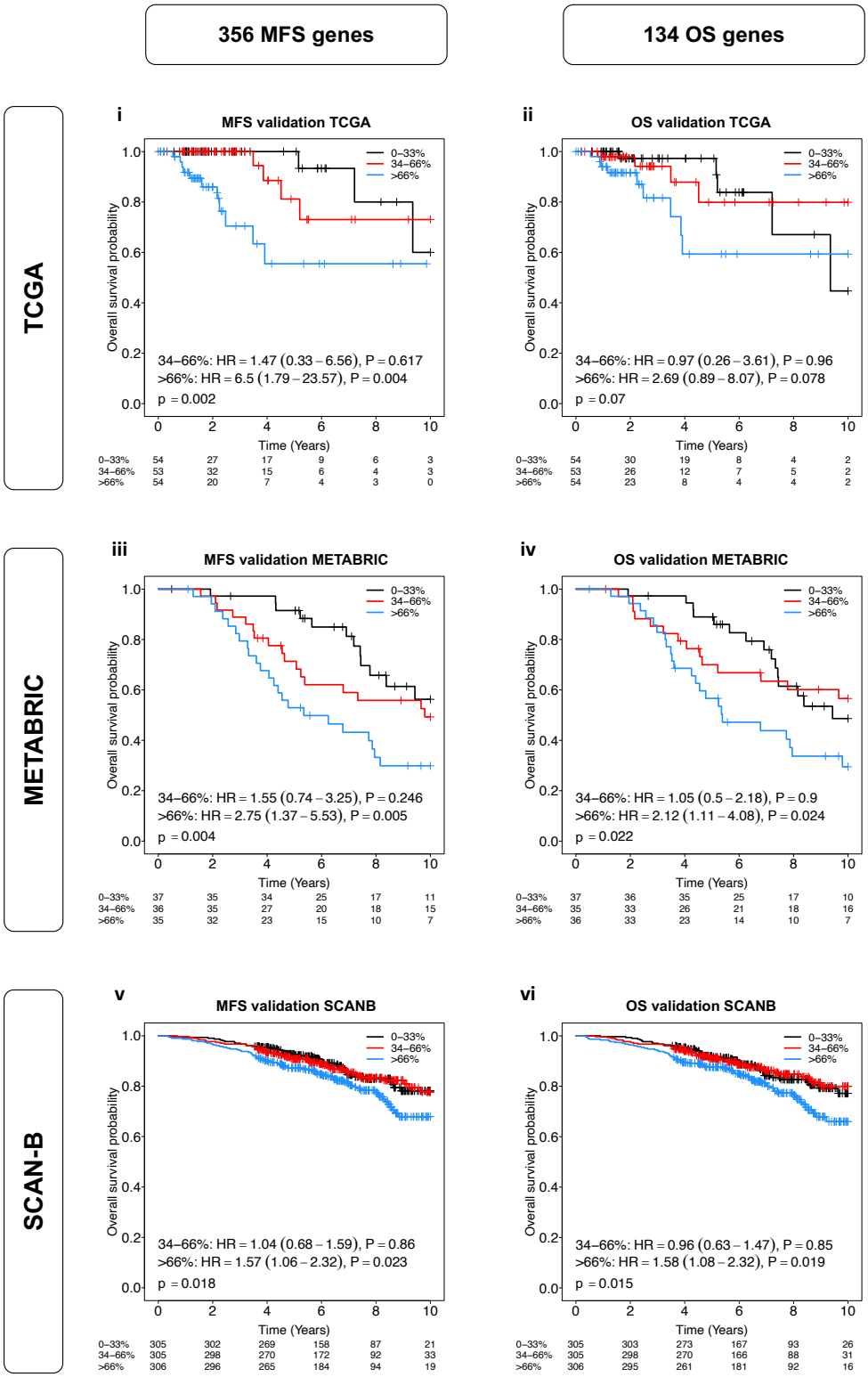


Figure 3.13: Prognostic KHP risk predictor validates in independent ILC cohorts: P-values of the different risk groups shown are relative to the 0 - 33% group. Lower-case p denotes the trend test p - value used to determine if the survival of the different groups followed a specific trend. Kaplan-Meier (KM) plots show OS in the 3 risk groups in i) TCGA for risk scores derived from KHP MFS genes (p = 0.002, trend test) and ii) TCGA for risk scores derived from KHP OS genes (p = 0.07, trend test) iii) METABRIC for risk scores derived from KHP MFS genes (p = 0.004, trend test) and iv) METABRIC for risk scores derived from KHP OS genes (p = 0.022, trend test) v) SCAN-B for risk scores derived from KHP MFS genes (p = 0.018, trend test) and ii) SCAN-B for risk scores derived from KHP OS genes (p = 0.015, trend test).

3.36 Transcriptomic differences are identified between early and late relapsing pleomorphic ILCs

Next focusing on the relapsing patients within the KHP pleomorphic ILC cohort (n = 24) differential gene expression analysis was performed to identify transcriptomic differences between tumours of patients with early vs late relapse. Of the 24 relapsing pleomorphic patients within the cohort 11 relapsed early (< 3 years of primary diagnosis) and 6 relapsed late (> 6 years after primary diagnosis). Patients with intermediate relapse (> 3 years, < 6 years) were excluded. Using an unadjusted p-value < 0.001 and a log_2FC+/-1, 7 genes were identified as differentially expressed between early and late-relapsing patients (Figure 3.14, Table 3.9). High expression of three genes; *CBX2*, *CSPG4* and *TG* were significantly associated with early-relapse patients, and four genes; *IGFBP5*, *ENSG00000232931.4*, *ATP8* and *MUC19* with late-relapse. Of these genes *IGFBP5* was present in the 134 genes associated with OS from the univariable survival analysis which were used to train the random forest model and showed an association with improved survival (coefficient: -0.296, p = 0.0134).

Gene	logFC	p-Value
IGFBP5	-3.90	0.000001
ENSG00000232931.4	-3.26	0.000015
ATP8	-2.48	0.000360
MUC19	-6.72	0.000724
CBX2	4.50	0.000859
CSPG4	3.49	0.000461
TG	6.32	0.000676

Table 3.9: Differentially expressed genes in early vs late relapse in the KHP pleomorphic ILC cohort: blue = higher expression in early relapse, red = higher expression in late relapse.

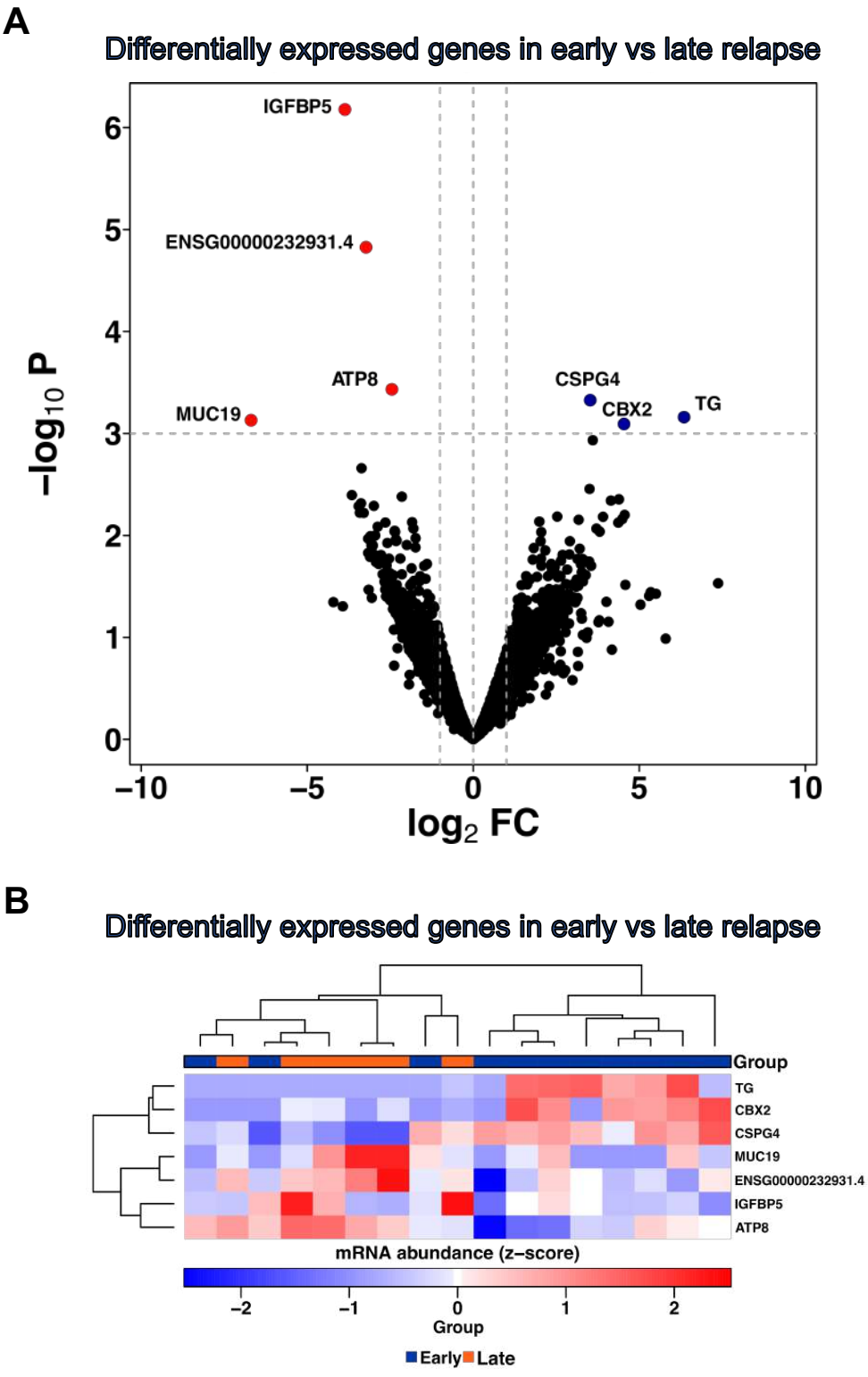


Figure 3.14: Differentially expressed genes between early and late relapsing pleomorphic ILCs: A) Volcano plot showing 7 differentially expressed genes in early (n = 11) vs late (n = 6) relapsing pleomorphic ILCs (red data points show genes more highly expressed in late relapse and blue data points show genes more highly expressed in early relapse). B) Heatmap showing the same 7 differentially expressed genes according to onset of disease relapse.

3.37 Transcriptomic differences are identified between early and late relapsing non-pleomorphic ILCs

Having identified transcriptomic features of aggressive pleomorphic ILCs an assessment was next made of the transcriptomic features of aggressive non-pleomorphic ILCs in a cohort of 15 relapsing non-pleomorphic ILCs from the KHP ILC cohort (Table 3.10) using the HTG Molecular Diagnostics Transcriptome panel which provides coverage of most human mRNA transcripts including isoforms and interrogates 19,398 targets using FFPE tissue sections. This platform was used as it was free having won a competition to trial the panel. Five patients relapsed early, and 10 patients relapsed late.

Parameter	Number (n)
Non-pleomorphic ILC tumour samples	15
Average patient age (years)	52.2
Disease relapse status at 15-year follow-up	
Relapse	15
No relapse	0
Onest of relapse:	
< 3 years of primary diagnosis	5
> 6 years after primary diagnosis	10
Tumour size	
≤ 20 mm	2
> 20 mm, ≤ 50 mm	11
> 50 mm	2
Number of involved lymph nodes	
0	7
1 – 3	6
4 – 9	2
≥ 10	0
Sites of metastases	
Bone	11
Skin	4
Liver	3
Pleura	3
Abdominal (not liver)	1
Lung	1
Brain	1
Brachial Plexopathy	1
Leptomeninges	1
Other / Unknown	3
Tumour grade	
II	14
III	1
Oestrogen (ER) Status	
ER+	15
ER-	0
Progesterone (PR) Status	
PR+	12
PR-	3
HER2 Status	
HER2+	0
HER2-	15

Table 3.10: Clinical features of KHP non-pleomorphic ILC RNA cohort

I annotated the H&Es identifying areas of tumour content of at least 50% and a minimum tumour area of 11 mm^2 was selected per case, as per the HTG protocol (Figure 3.15). These were then used to select the tumour areas for sequencing on the corresponding unstained tissue section.

Overall using an unadjusted p-value threshold of $p < 0.001$ and log_2FC +/-1, 21 genes were identified as significantly differentially expressed between early and late relapsing patients. Fifteen genes were more highly expressed in early relapse and 6 genes more highly expressed in late relapse (Figure 3.16, Table 3.11). There was no overlap between the 21 differentially expressed genes and those associated with OS and MFS in the pleomorphic KHP cohort.

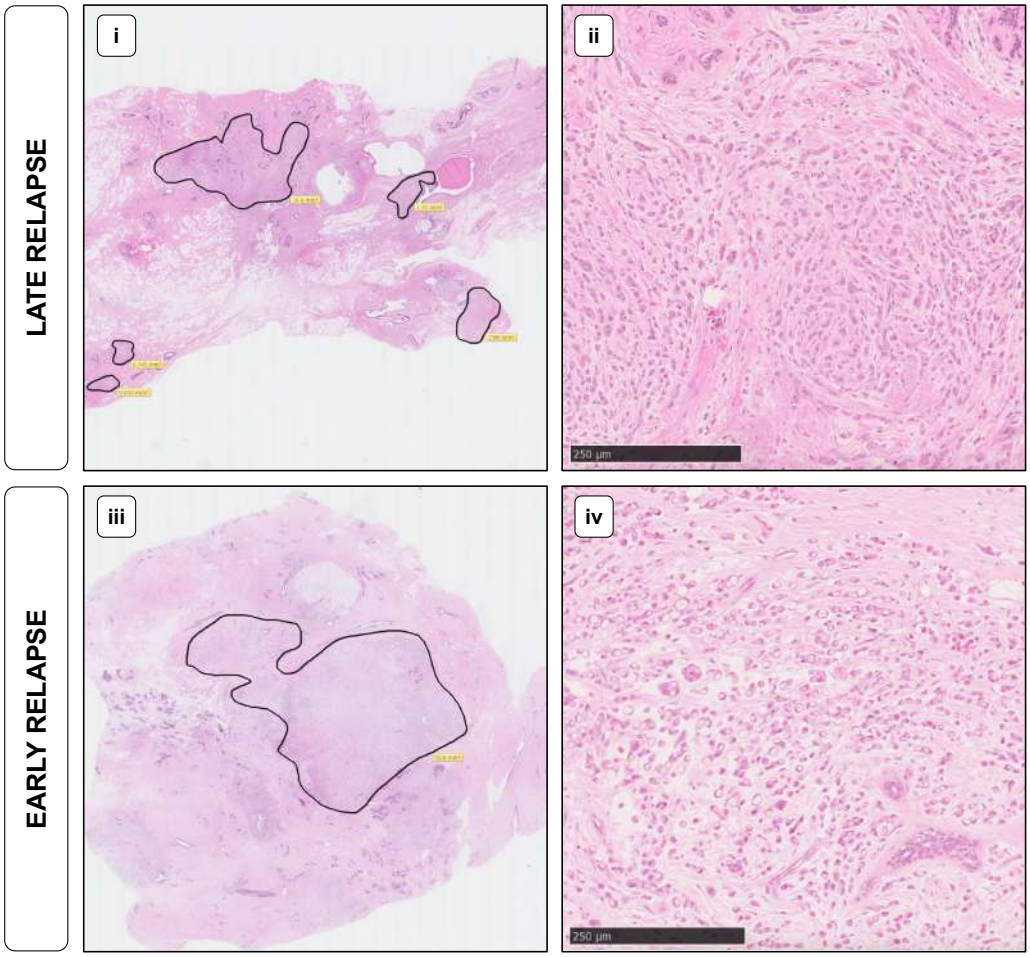


Figure 3.15: Tumour area selection for HTG RNA sequencing: Top panel shows i) low power image of representative late relapse case tumour area selection and ii) higher power (x10) view of selected tumour area. Bottom panel shows iii) low power image of representative early relapse case tumour area selection and ii) higher power (x10) view of selected tumour area.

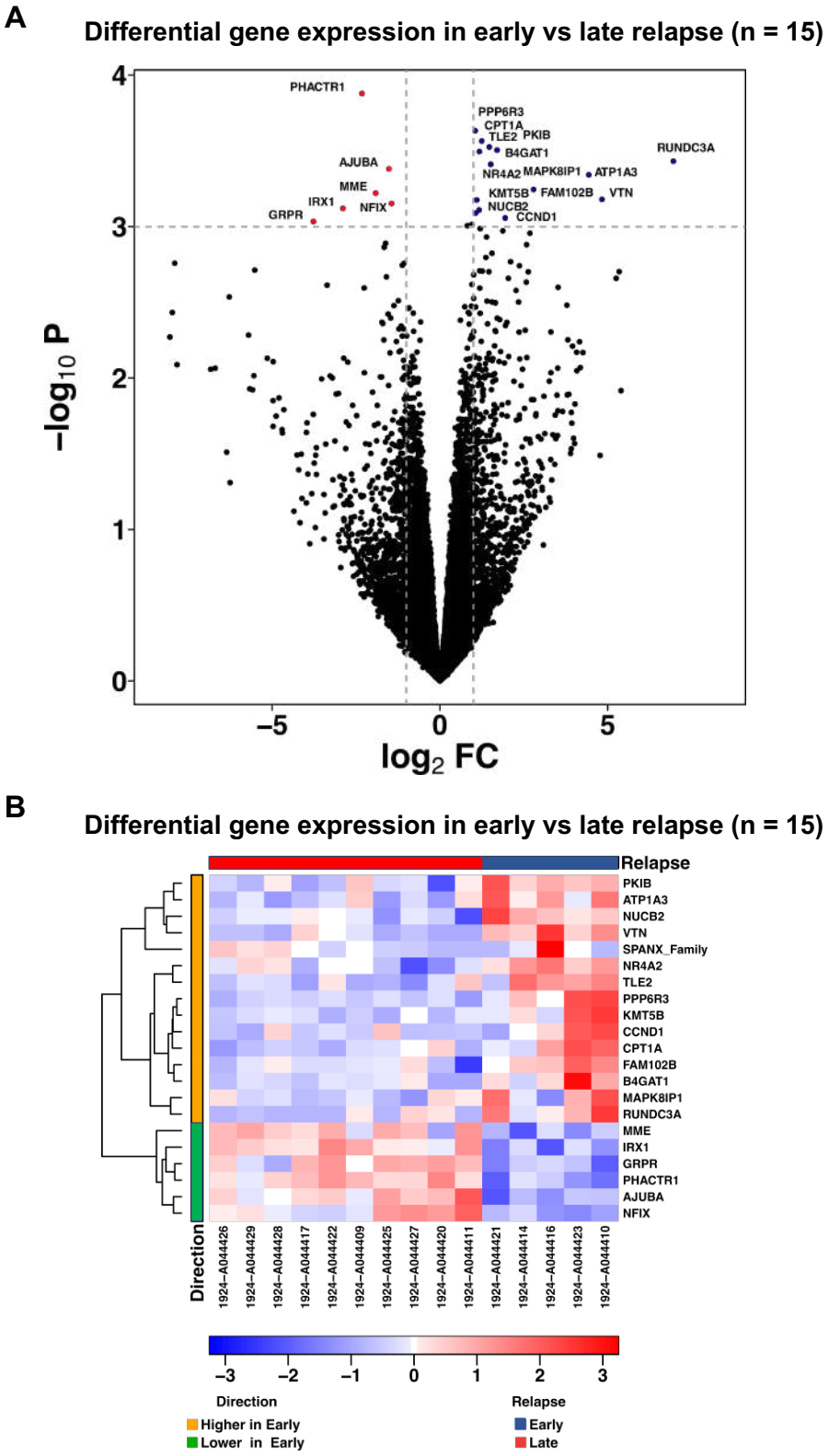


Figure 3.16: Twenty-one genes are differentially expressed in early vs late relapsing non-pleomorphic ILC: A) Volcano-plot and B) heatmap showing 15 genes more highly expressed in early relapse and 6 genes more highly expressed in late relapse (unadjusted $p < 0.001$ and log_2FC +/-1, red data points show genes more highly expressed in late relapse and blue data points show genes more highly expressed in early relapse).

	Gene	logFC	p-value
Higher expression in late relapse	GRPR	-3.769	0.00093
	IRX1	-2.891	0.00076
	PHACTR1	-2.324	0.00013
	MME	-1.915	0.00060
	AJUBA	-1.517	0.00042
	NFIX	-1.441	0.00070
Higher expression in early relapse	PPP6R3	1.054	0.00023
	NUCB2	1.065	0.00082
	KMT5B	1.097	0.00067
	FAM102B	1.169	0.00078
	TLE2	1.177	0.00032
	CPT1A	1.245	0.00027
	PKIB	1.471	0.00030
	NR4A2	1.514	0.00039
	B4GAT1	1.704	0.00031
	CCND1	1.945	0.00088
	MAPK8IP1	2.790	0.00057
	ATP1A3	4.437	0.00046
	VTN	4.829	0.00066
	RUNDC3A	6.957	0.00037
	SPANX_Family	8.406	0.00087

Table 3.11: Fifteen genes more highly expressed in early relapsing non-pleomorphic ILCs

3.37.1 Early relapse is associated with cell-cycle, interferon, TNF-α and ER signalling pathways in non-pleomorphic ILC

Cancer hallmarks was used for gene set enrichment analysis (GSEA) via the 'fgsea' library in R [377]. This identified an enrichment of pathways associated with cell cycle (G2M checkpoint / E2F targets), interferon gamma (IFN-γ), interferon alpha (IFN-α), TNF-α and oestrogen signalling pathways in early vs late relapse (using an adjusted p-value < 0.05). Pathways enriched in late relapse included epithelial – mesenchymal transition, myogenesis, apical junction and coagulation (Table 3.12).

	Pathway	Adjusted p-value	Normalised enrichment score
Higher in late relapse	Epithelial-mesenchymal transition	0.00200	-2.70
	Myogenesis	0.00200	-1.66
	Apical junction	0.00660	-1.51
	Coagulation	0.01215	-1.57
Higher in early relapse	Estrogen response - late	0.01033	1.51
	G2M checkpoint	0.00200	1.80
	E2F targets	0.00200	1.84
	Interferon gamma response	0.00200	1.88
	Interferon alpha response	0.00200	2.35
	Estrogen response - early	0.00246	1.72
	TNF-α signalling via NF-κB pathway	0.00430	1.59

Table 3.12: Eleven pathways associated with onset of relapse in non-pleomorphic ILC using Cancer Hallmarks database.

3.38 Pleomorphic ILCs show higher expression of genes involved in cell differentiation and androgen biosynthesis

Having identified transcriptomic differences between patients with differing clinical outcomes in separate KHP pleomorphic and non-pleomorphic cohorts, an assessment was next made to establish any differences between pleomorphic and non-pleomorphic ILCs at the gene expression level irrespective of clinical outcome, in an independent cohort; TCGA. This well-annotated dataset was used as it provides histological images of the tumour cases for review. Given that ILC histological subtype information was not available for the cases, digital pathology technology with QuPath, was used to determine which cases were pleomorphic and non-pleomorphic based on an assessment of tumour cell nuclear size. Pleomorphic nuclei were identified as 75 – 250 μm^2 in size, thus > 4 times the size of an average lymphocyte, in keeping with the WHO histological criteria for pleomorphic ILC [73, 78] (Figure 3.17). This **approach identified 36 pleomorphic and 143 non-pleomorphic cases.**

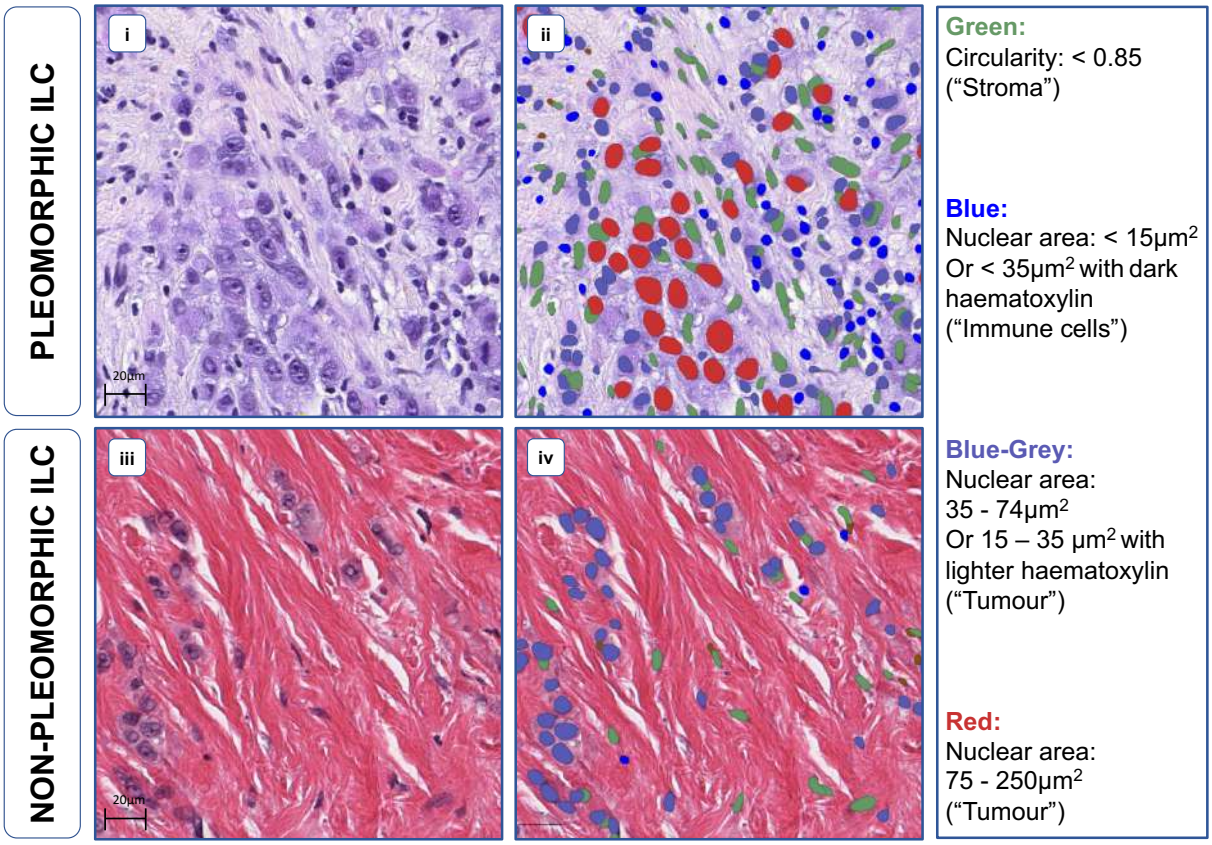


Figure 3.17: Digital pathology technology identifies pleomorphic ILCs from TCGA: Top panel: i) H&E and ii) corresponding QuPath image identifying cell types based on nuclear size and circularity in a pleomorphic ILC case containing red nuclei ≥ 75 µm indicating nuclear pleomorphism iii) H&E and iv) corresponding QuPath image showing a non-pleomorphic case with an absence of red nuclei suggesting a lack of nuclear pleomorphism

Having identified the pleomorphic ILCs from the TCGA dataset, differential gene expression analysis was performed in the pleomorphic (n = 36) vs non-pleomorphic (n = 143) cases. Top differentially expressed genes were identified if they fulfilled the criteria of an adjusted p-value threshold < 0.05 and log_2FC +/-1.5. Overall there were 37 significantly differentially expressed genes between the ILC histological subtypes (Figure 3.18, Table 3.13). None of the 37 differentially expressed genes were included in the significant genes associated with OS and MFS in the KHP pleomorphic cohort.

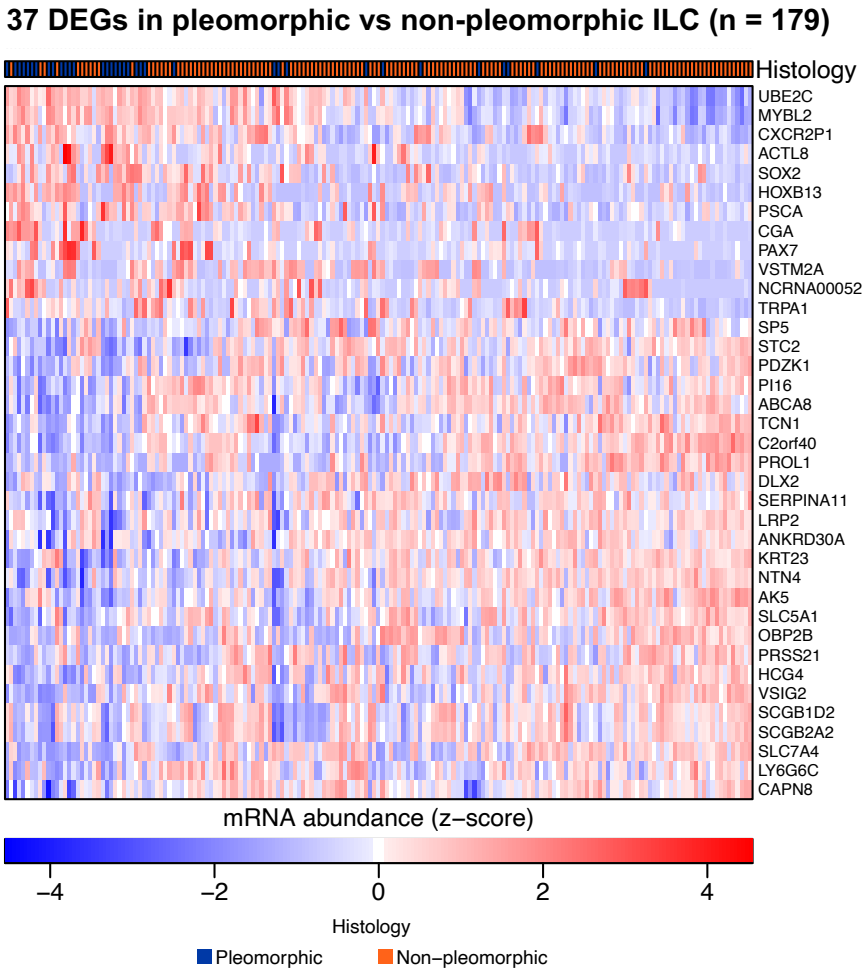


Figure 3.18: Pleomorphic ILCs show higher expression of 12 genes: Heatmap showing 37 significantly differentially expressed genes (DEGs) in pleomorphic (n = 36) and non-pleomorphic (n = 143) ILCs from the TCGA dataset identified histologically using QuPath.

	Gene	logFC	adjusted p-value
Higher expression in pleomorphic ILC	UBE2C	1.566	6.88E-07
	MYBL2	1.694	8.36E-07
	ACTL8	1.835	1.59E-06
	PSCA	2.157	0.00045
	CXCR2P1	1.584	0.00259
	SOX2	1.633	0.00543
	TRPA1	1.716	0.00899
	CGA	1.821	0.00959
	PAX7	1.532	0.01013
	NCRNA00052	1.531	0.01441
	HOXB13	1.934	0.01651
	VSTM2A	2.357	0.01846
Higher expression in non-pleomorphic ILC	HCG4	-1.636	4.36E-07
	NTN4	-1.513	1.56E-06
	STC2	-2.075	5.95E-06
	CAPN8	-1.718	0.00012
	LRP2	-2.156	0.00020
	LY6G6C	-1.689	0.00028
	SERPINA11	-1.876	0.00028
	ABCA8	-1.549	0.00031
	KRT23	-1.617	0.00050
	C2orf40	-1.768	0.00062
	ANKRD30A	-1.886	0.00068
	AK5	-1.529	0.00073
	SP5	-1.522	0.00086
	SLC7A4	-2.014	0.00096
	TCN1	-2.252	0.00185
	PRSS21	-1.608	0.00247
	VSIG2	-1.698	0.00325
	SCGB2A2	-2.744	0.00345
	PI16	-1.670	0.00454
	SLC5A1	-1.652	0.00554
	SCGB1D2	-2.340	0.01236
	DLX2	-1.558	0.01808
	PROL1	-1.828	0.02020
	PDZK1	-1.654	0.02071
	OBP2B	-1.837	0.03110

Table 3.13: Differentially expressed genes in pleomorphic vs non-pleomorphic ILC from TCGA: (blue = genes more highly expressed in pleomorphic ILC, red = genes more highly expressed in non-pleomorphic ILC).

Enrichr [376] was used to study the pathways associated with the twelve genes which were more highly expressed in the pleomorphic ILCs. This identified pathways associated with cell differentiation, G_2M checkpoint and androgen biosynthesis (Table 3.14).

Pathway	Adjusted p-value
Cell differentiation pathway	0.0029
Glycoprotein hormones	0.0472
Mineralocorticoid biosynthesis	0.0472
Thyroxine biosynthesis	0.0472
Androgen biosynthesis	0.0472
Hormone ligand-binding G-protein coupled receptors	0.0472
Peptide hormone biosynthesis	0.0472

Table 3.14: Seven pathways enriched in pleomorphic ILC from 'BioPlanet 2019' database.

To control for hormone status given that pleomorphic ILCs show a higher rate of ER- and HER2+ disease, and to determine if nuclear size and pleomorphism alone are reflective of underlying differentially expressed genes. an assessment was next made of the differences between ER+/HER2- pleomorphic (n = 17) and non-pleomorphic (n = 71) ILCs at the gene expression level. This identified 18 genes which were significantly differentially expressed between ER+/HER2- pleomorphic and non-pleomorphic ILC (Figure 3.19, Table 3.15). There was no overlap with these genes and those included in the significant genes associated with OS and MFS in the KHP pleomorphic cohort.

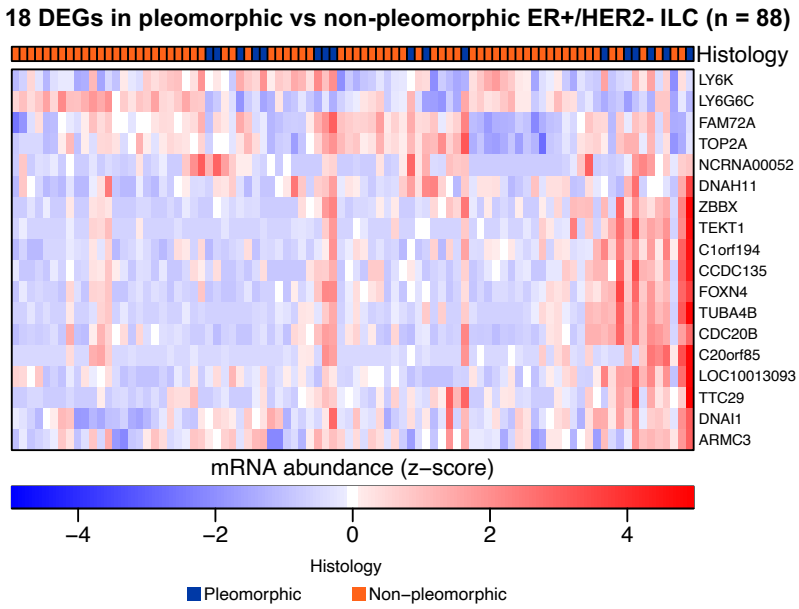


Figure 3.19: ER+/HER2- pleomorphic ILCs show differences from non-pleomorphic ILC at the gene expression level: Heatmap showing 18 differentially expressed genes (DEGs) in ER+/HER2- pleomorphic (n = 17) and non-pleomorphic (n = 71) ILCs from the TCGA dataset identified histologically using QuPath.

Differential gene expression analysis was also completed in HER2+ pleomorphic ILC (n = 7) and non-pleomorphic HER2+ ILC (n = 17). However no genes that fulfilled the statistical thresholds (adjusted p-value < 0.05, log_2FC +/-1.5) due to lack of statistical power.

Gene	logFC	adjusted p-value
CCDC135	1.54	0.0192
TTC29	1.60	0.0204
LOC100130933	1.54	0.0210
FAM72A	1.53	0.0210
TOP2A	1.62	0.0270
ZBBX	1.60	0.0294
NCRNA00052	2.42	0.0328
CDC20B	3.23	0.0330
DNAH11	1.62	0.0331
FOXN4	1.67	0.0349
C1orf194	1.70	0.0358
DNAI1	1.76	0.0369
TUBA4B	1.70	0.0383
ARMC3	1.96	0.0386
TEKT1	1.84	0.0447
C20orf85	2.01	0.0453
LY6K	1.97	0.0453
LY6G6C	-1.98	0.0296

Table 3.15: Eighteen differentially expressed genes in ER+/HER2- pleomorphic vs non-pleomorphic ILC from TCGA: (blue = genes more highly expressed in pleomorphic ILC, red = genes more highly expressed in non-pleomorphic ILC).

3.4 Discussion

The main aims of this chapter were to identify drivers of clinically aggressive ILCs at the genomic and transcriptomic levels. An assessment of any differences based upon ILC histological subtype was also made, given that pleomorphic ILCs are characterised by clinically aggressive disease. Histological assessment of the cohort was first made and the use of QuPath highlighted the benefit of digital pathology approaches, in confirming the initial histological classification, showing higher nuclear size and variability in nuclear size among pleomorphic cases.

At the genomic level, targeted sequencing of a cohort of pleomorphic and non-pleomorphic ILCs enabled the identification of alterations which firstly occur more frequently in pleomorphic ILC, as a distinct histological subtype. These included a higher frequency of *TP53* and *FAT1* mutations and a lower frequency of *PIK3CA* mutations. *TP53* mutations have previously been shown in numerous studies to occur at higher frequency in pleomorphic ILC [100, 101, 103, 293]. When the rate of *TP53* mutations was compared only in ER+/HER2- pleomorphic vs non-pleomorphic ILCs statistical significance was lost, suggesting that higher rates observed in pleomorphic ILC are driven by the higher rates of triple-negative and HER2+ disease rather than nuclear pleomorphism alone.

Less is known about *FAT1* mutations in the context of pleomorphic ILC. *FAT1 (*FAT Atypical Cadherin 1) is a member of the cadherin superfamily, a group of integral membrane proteins. It has previously been associated with cancer cell proliferation with loss of function mutations associated with progression of cancers of the head and neck [378]. In breast cancer loss of *FAT1* has been associated with resistance to CDK4/6 therapy [379]. Specifically in ILC, *FAT1* mutations have been shown to occur at higher frequency in metastatic lesions compared to primary disease [291]. This may suggest that the acquisition of *FAT1* mutations in primary tumour subclones results in greater metastatic potential of these subclones and clinically aggressive behaviour, which is consistent with the finding of higher rates of *FAT1* mutations in pleomorphic ILC, an aggressive variant, which is significantly associated with earlier disease relapse in our cohort and wider studies.

In addition, *HER2* amplifications and *CDK12* alterations (amplifications and missense mutations) were observed only in the pleomorphic cases. A higher frequency of *HER2* amplifications and mutations in pleomorphic ILC has been consistently reported in previous studies [100-103], yet this is not the case for *CDK12* alterations. Cyclin-dependent kinase 12

(*CDK12*) is a key transcription-associated cyclin-dependent kinase (CDK) with diverse roles in regulating gene transcription, translation, cell cycle progression and cell proliferation, DNA damage response and RNA splicing [380]. In breast cancer it displays both pro-tumoral and anti-tumour effects among different tumour types. For example in HER2+ disease it promotes tumour growth [381] whilst in TNBCs, *it* has tumour suppressor functions [382, 383]. Whilst the presence of *CDK12* alterations was not of prognostic significance in the KHP cohort, they significantly co-occurred with *HER2* alterations. Interestingly this has also been demonstrated in a previous study which used IHC to assess the occurrence and distribution of *CDK12* protein expression in independent breast cancer cohorts, correlating expression with genomic status and clinical outcome [382]. The study showed that *CDK12* was positively correlated with *HER2* positivity, although was not an independent predictor of worse outcome, consistent with the findings from the KHP cohort. The co-occurrence of these alterations results from the fact that *CDK12* has been shown to map to the smallest region within the *HER2* amplicon [384, 385].

Whilst *CDK12* alterations were not prognostic, *FGFR1* alterations were significantly associated with the development of disease relapse among the pleomorphic patients. *FGFR1* (Fibroblast Growth Factor Receptor 1) is a member of the Fibroblast Growth Factor family, which consists of 4 receptors and 18 ligands [386]. *FGFR1* amplifications occur in approximately 10% of breast cancers and have been associated with poor prognosis [387]. A study assessing *FGFR1*-amplified breast cancers, showed that *FGFR1* expression is strongly correlated with *FGFR1* copy number and that the overexpression of FGFR1 leads to the enhanced ligand-dependent as well as ligand-independent signalling [388]. In addition, *FGFR1* amplification was shown to occur more frequently in luminal B type breast cancers [388]. In the KHP cohort 60% of *FGFR1* altered tumours were ER+ and PR- and hence consistent with a luminal B phenotype. Furthermore *FGFR1* amplification has been shown to mediate endocrine therapy resistance in ER+ breast cancer [389] providing rationale for FGFR1 blockade. A recent phase IIa clinical trial (RADICAL; NCT01791985) assessed the use of AZD4547 (a selective inhibitor of FGFR 1, 2 and 3) with an aromatase inhibitor in 52 unselected ER+ metastatic breast cancer patients with resistance to endocrine therapy [390]. It showed some level of activity, with an objective response rate of 5% meeting the study end-point and 6 differentially-expressed genes were identified which differentiated those deriving benefit from the addition of AZD4547 [390]. Going forward larger trials are required to establish the benefit of *FGFR1* inhibition in *FGFR1* amplified breast cancers, particularly in rare subtypes such as pleomorphic ILC, for whom this alteration appears to be associated with disease relapse in the KHP cohort.

Whilst *FGFR1* alterations were associated with disease relapse in pleomorphic ILC, there were no other prognostic genes within both the pleomorphic and non-pleomorphic patients. Therefore an assessment of the transcriptomic landscape of clinically aggressive ILCs was next made to better understand the drivers of aggressive disease biology at the gene expression level. Whilst previous studies have identified unique gene expression subtypes in ILC [32, 277], they have not consistently shown prognostic significance or categorised patients based on ILC histological subtype. Therefore an assessment of pleomorphic ILCs at the gene expression level was completed using the KHP pleomorphic cohort enabling a better understanding of the transcriptomic landscape of this clinically aggressive ILC subtype. Importantly, whilst bulk RNA gene expression data is usually characterised by low tumour purity for ILC, due to the high stromal component of ILCs, tissue needle macrodissection prior to sequencing ensured a high tumour content within the cohort, ideal for studying pleomorphic ILC tumour cell specific signal. Survival analysis and the study of significant OS and MFS genes through the use of a random forest model, enabled the generation of prognostic risk scores within the KHP pleomorphic ILC discovery cohort which further validated as predictors of OS in matched purified and deconvolved ILC validation cohorts; TCGA, METABRIC and SCAN-B.

Differential gene expression analysis identified significantly up-regulated genes in early relapsing pleomorphic ILC patients, including *CSPG4, TG* and *CBX2*. *CBX2* (Chromobox 2), encodes a component of the polycomb multiprotein complex, and has previously been shown to be upregulated in breast cancer, with high expression associated with worse survival outcomes [391] which is consistent with higher expression in early relapse within the pleomorphic ILCs. Higher expression of genes such as *ATP8*, *MUC19* and *IGFBP5* were associated with the later onset of disease recurrence among relapsing pleomorphic ILC patients. Consistent with the finding of higher *IGFBP5* expression being associated with later onset of relapse, a study using IHC for IGFBP5 in a cohort of 153 ER+ breast cancers, showed that low levels were associated with tamoxifen resistance and shorter overall survival times [392].

Finally, digital pathology approaches enabled the identification of pleomorphic ILCs in the TCGA dataset and differential gene expression analysis and pathway enrichment analysis identified cell-cycle pathways and androgen biosynthesis as pleomorphic-enriched pathways. Studies suggest that triple-negative ILCs have unique transcriptomic profiles and are characterised by increased androgen-receptor signalling [68], and therefore increased androgen signalling in pleomorphic ILC vs non-pleomorphic ILC may reflect underlying higher rates of triple-negative disease in pleomorphic vs non-pleomorphic histology.

Overall the work in this chapter has characterised and identified differences between pleomorphic and non-pleomorphic ILC at the genomic level, further identifying an association with *FGFR1* alterations and clinically aggressive disease in pleomorphic ILC. In addition it has identified transcriptomic features associated with aggressive ILCs and survival analysis in the KHP pleomorphic cohort has generated a prognostic gene expression risk predictor which validates in larger independent ILC cohorts.

Chapter 4: Evaluation of the immune landscape and prognostic associations in ILC

4.1 Background

Over the last decade there has been growing interest in understanding and targeting the immune system as a therapeutic strategy in a range of solid malignancies. In particular, the use of immunotherapy has revolutionised the treatment options and clinical outcomes in multiple tumour types, including non-small cell lung cancer [393], malignant melanoma [394] and cancers of the head and neck [395]. Despite this, immunotherapy has proven to be less clinically effective in breast cancer. Only a small proportion of patients appear to derive some degree of benefit from immunotherapies, and these are predominantly in heavily pre-treated patients or TNBCs [244-247, 249]

Despite some clinical benefit being provided by immunotherapy in subsets of TNBC patients, ER+ disease, which includes most ILCs, accounts for the majority (70%) of breast cancer cases [93] and there is therefore a need to better understand the immune microenvironment in ER+ disease and specifically ILC and the possible role for immune-based treatments in these patients. There are multiple ongoing phase III trials assessing the efficacy of ICIs in the ER+ breast cancer population including CHECKMATE 7FL, KEYNOTE-B49 and KEYNOTE-756 [257-259] as well as the phase II GELATO trial specifically in ILC patients [260]. These will help elucidate the potential role of immunotherapy in ER+ disease as currently there are no approved immunotherapies for these patients. However alongside this, there is a fundamental need to better understand the nature of the immune microenvironment in ILC, particularly in clinically aggressive tumours such as pleomorphic ILC.

ER+/HER2- breast cancers such as ILC are generally considered ‘immune-cold’ tumours meaning they are characterised by a lower level of immune infiltrate compared to TNBCs and HER2+ tumours [396]. Importantly, the presence and abundance of neoantigen-specific T lymphocytes have been strongly implicated as playing a critical role in generating an effective response to ICIs. TNBCs are characterised by a more inflammatory tumour microenvironment compared to ER+ breast cancer, with a higher number of stromal TILs, CD8+ T lymphocytes and a higher level of expression of immune-related genes [397-399].

Additional factors explain why ER+ breast cancers respond less well to ICIs compared to TNBCs. TMB is a continuous variable and studies show that it can range from 0.001Mut/Mb

to over 1000Mut/Mb) across and within adult solid malignancies [400-402]. Whilst the TMB of breast cancer is classed as intermediate, with a median mutation rate of 2.63Mut/Mb [401, 403]. there is a relatively lower TMB in ER+ breast cancer compared to TNBCs [404, 405]. A higher TMB is associated with a higher number of neoantigens within the tumour microenvironment, or 'neo-antigen load' [14]. The recognition of tumour-specific neoantigens by T-cell receptors results in the initiation of an immune response driving cytotoxic T lymphocytes into the vicinity of the tumour promoting tumour cell death [14].

A histological assessment of the level of stromal TILs has been shown to hold prognostic and predictive value in the context of TNBC and HER2+ disease where increased TILs are associated with improved DFS and with higher response rates to neoadjuvant chemotherapy [313, 314, 334-338]. Some studies also suggest an association between a high level of TILs and response to anti-HER2 therapy in HER2+ breast cancers [406-408]. The quantification of stromal TILs therefore represents a useful and informative additional piece of information in a histology report, providing an indication of lymphocyte-based tumour immunogenicity at no additional diagnostic cost. However the prognostic relevance of TILs in ER+ breast cancer including ILC is less clear [338]. ER+ tumours are generally characterised by lower TILs than TNBC and ER-/HER2+ disease, and a large retrospective analysis interestingly showed that ER+ ILCs have a lower level of TILs compared to ER+ IC-NST [337]. Moreover, although showing a higher TMB, a low level of stromal TILs is seen in ILC metastases compared to the primary tumour, although a higher level has been identified in the previously used 'mixed non-classic' histology [282]. There have been no significant differences in the level of TILs identified between the different ILC metastatic sites [290].

Two main studies have previously characterised TILs using large retrospective ILC cohorts and identified that a small subset of cases have a relatively high level of stromal TILs [282, 339]. When associations with clinical outcome were assessed, high TIL scores (defined as > 10% [282] and > 5% [339] in the two studies respectively) were associated with younger patient age, lymph node involvement and more proliferative tumours in the first study [282], although in multivariate analysis, stromal TILs were not an independent predictor of worse outcome. In the second study high stromal TILs were associated with a range of poor prognostic features and worse OS and DFS [339]. When ILC histological subtypes were considered, the first study identified a significantly lower level of TILs in alveolar ILC ($p = 0.02$) and higher levels in 'mixed non-classic' ILC ($p < 0.001$) [282]. A distinct pleomorphic group was not defined in these studies. Overall, the results of these studies would therefore suggest that the role of TILs in ILC is different to that observed in TNBC and HER2+ disease, being instead associated with more clinically aggressive tumours. However the levels of stromal TILs

in pleomorphic ILC and their prognostic significance has not been explicitly assessed and there is a need to characterise the immune landscape at the histological level particularly in pleomorphic ILC and to better understand whether stromal TILs hold prognostic significance in this clinically aggressive ILC subtype.

4.2 Chapter Aims

Overall the purpose of this chapter is to:

3) Assess the abundance of stromal TILs in ILC and assess whether this differs between pleomorphic and non-pleomorphic cases, and according to hormone receptor status.

4) Assess the correlation between stromal TILs and genomic alterations in ILC.

5) Establish any associations between the level of stromal TILs and clinical outcome in pleomorphic and non-pleomorphic ILC.

4.3 Results

4.31 ILCs are characterised by a low level of stromal TILs

To gain a global picture of the immunogenicity of ILC, stromal TILs were quantified in the entire KHP ILC cohort at the histological level. The cohort consisted of a combination of pleomorphic (n = 63) and non-pleomorphic cases (n = 100). The median TIL score for the entire cohort (n = 163) was 2.5% (IQR: 1.5% - 4.5%). A previous study quantifying TILs in primary ILC assigned cases to low (≤ 5%), intermediate (> 5 and ≤ 10%) and high (> 10%) TIL groups [282]. In the KHP ILC cohort the majority (81.6%) of cases fell in the 'low' TIL group, 14.1% in the 'intermediate' and 4.3% in the 'high' TIL groups respectively (Figure 4.1).

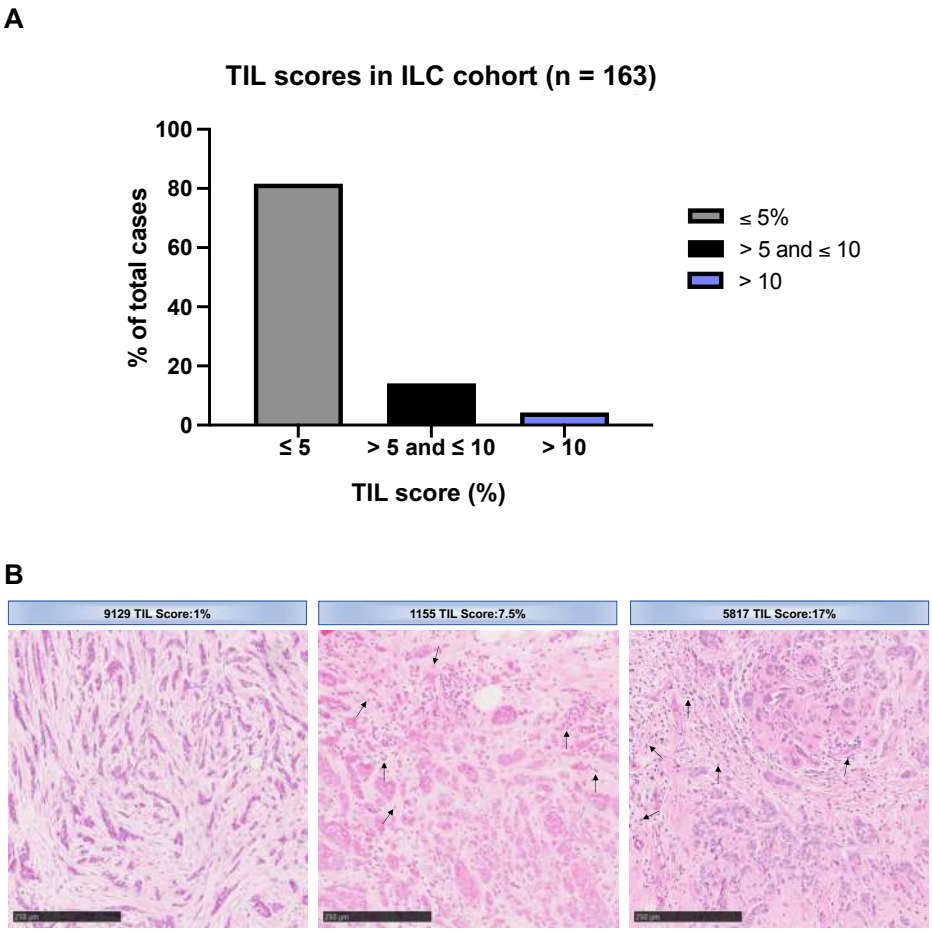


Figure 4.1: ILCs are characterised by low levels of stromal TILs: A) Bar chart showing the relative proportions of ILC cases within each TIL score category (n = 163: 81.6%, 14.1% and 4.29% of cases fall into the low, intermediate and high TIL score groups respectively) B) Representative H&E sections showing cases with TIL scores in the low (case 9129 scoring 1%), intermediate (case 1155 scoring 7.5%) and high (case 5817 scoring 17%) TIL score categories (stromal TILs are depicted by arrows, magnification x 10).

4.32 Pleomorphic ILC is characterised by a higher level of stromal TILs

Next stromal TIL scores were compared between pleomorphic (n = 63) and non-pleomorphic ILCs (n = 100). Heterogeneity with respect to the degree of immune cell infiltration was observed in both pleomorphic (Figure 4.2A) and non-pleomorphic histological subtypes. The median TIL score in the pleomorphic group was 4% (IQR: 2.0% - 6.5%) and in the non-pleomorphic ILC group was 2% (IQR:1.0% - 3.375%). Overall, there was a significantly higher level of stromal TILs in the pleomorphic vs non-pleomorphic ILC cases ($p < 0.0001$, Mann-Whitney U test, Figure 4.2B). A higher proportion of pleomorphic cases fell in the 'intermediate' and 'high' TIL groups [282] compared to the non-pleomorphic cases (Figure 4.2D, Table 4.1).

TIL score (%)	Pleomorphic (% of cases)	Non-pleomorphic (% of cases)	Entire cohort (% of cases)
less than equal 5	68.3	90.0	81.6
>5 less than equal 10	25.4	7.0	14.1
>10	6.3	3.0	4.3

Table 4.1: Summary of the proportion of ILC cases falling in each stromal TIL score category

Within the pleomorphic group there were higher numbers of ER- and HER2+ cases compared to the non-pleomorphic group (pleomorphic group: ER- cases: n = 8, HER2+ cases: n = 8, non-pleomorphic group: ER- cases: n = 0, HER2+ cases: n = 3). To control for hormone status, a further analysis was completed assessing the level of stromal TILs in ER+, HER2- pleomorphic (n = 37) vs non-pleomorphic ILCs (n = 83). Patients for whom ER and HER2 status was unavailable were excluded from this analysis. This again showed a significantly higher level of stromal TILs in pleomorphic vs non-pleomorphic tumours ($p = 0.0119$, Mann-Whitney U test, Figure 4.2C). A linear regression model with TIL levels as the outcome further demonstrated that pleomorphic ILCs were associated with a significantly higher level of TILs compared to non-pleomorphic ILCs ($p = 0.0165$, $\beta = 2.12$), after adjusting for clinical covariates such as tumour stage (T1, T2, T3), hormone status (1 = HER2+ or TNBC, 0 = others), and lymph node status (1 = positive, 0 = negative). Interestingly in mixed tumours containing classic and pleomorphic areas heterogeneity in the immune infiltrate was observed with denser infiltrates generally observed in the pleomorphic tumour regions (Figure 4.3).

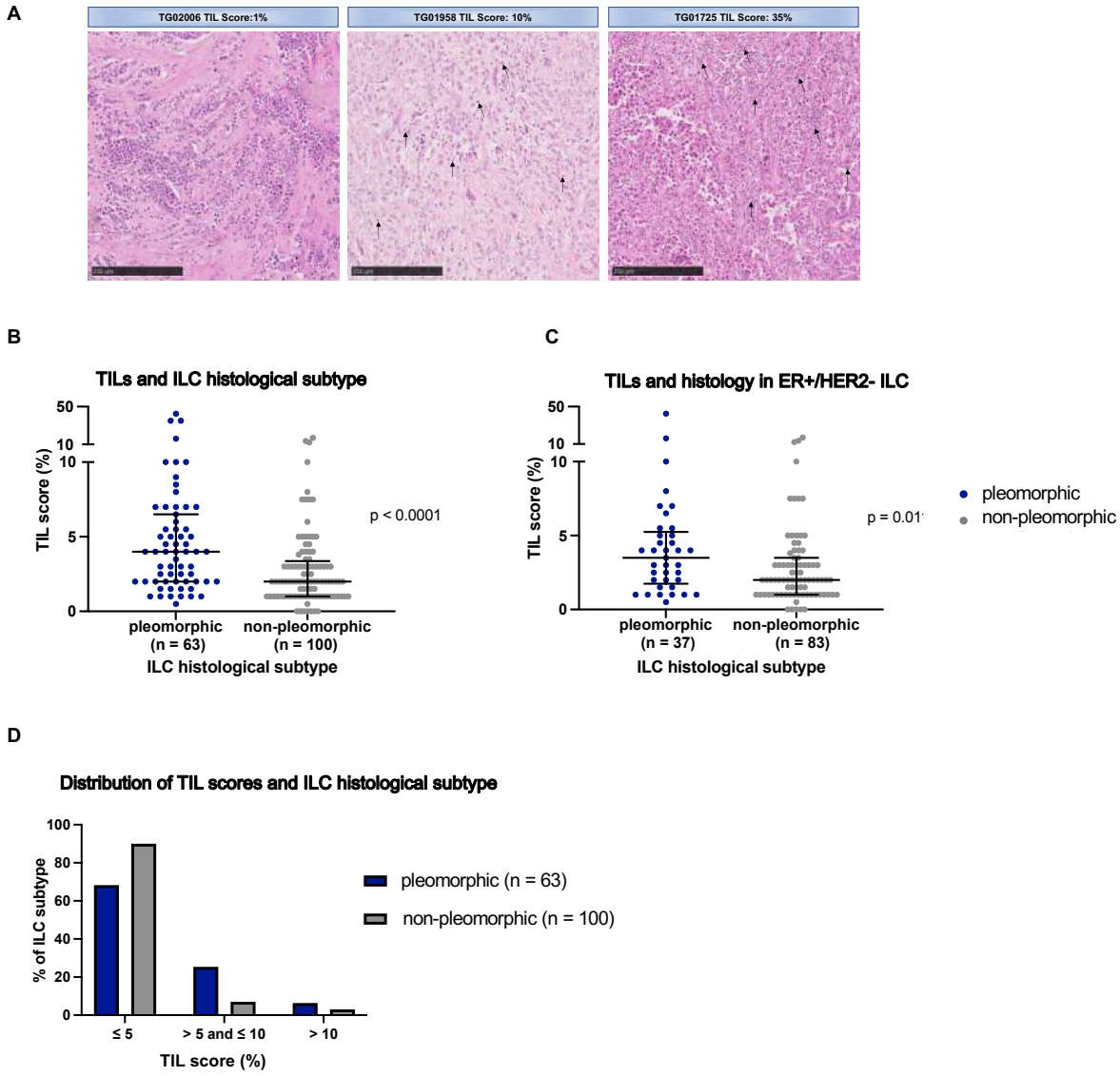


Figure 4.2: Pleomorphic ILC is more immunogenic than non-pleomorphic ILC at the histological level: A) Representative H&E sections from pleomorphic ILC cases showing low, intermediate and high TIL scores (stromal TILs are depicted by arrows, magnification x10) B) Scatter plot showing individual TIL scores for pleomorphic (n = 63) and non-pleomorphic ILC (n = 100) showing significantly higher TIL scores in the pleomorphic group ($p < 0.0001$, Mann-Whitney U test) C) Scatter plot showing individual TIL scores for ER+ HER2- pleomorphic (n = 37) vs non-pleomorphic ILC (n = 83) showing significantly higher TIL scores in the pleomorphic group ($p = 0.0119$, Mann-Whitney U test) D) Bar chart showing relative proportions of pleomorphic (n = 63) and non-pleomorphic (n = 100) ILCs falling into each TIL score category showing higher proportions of pleomorphic ILCs in the intermediate and high TIL score groups.

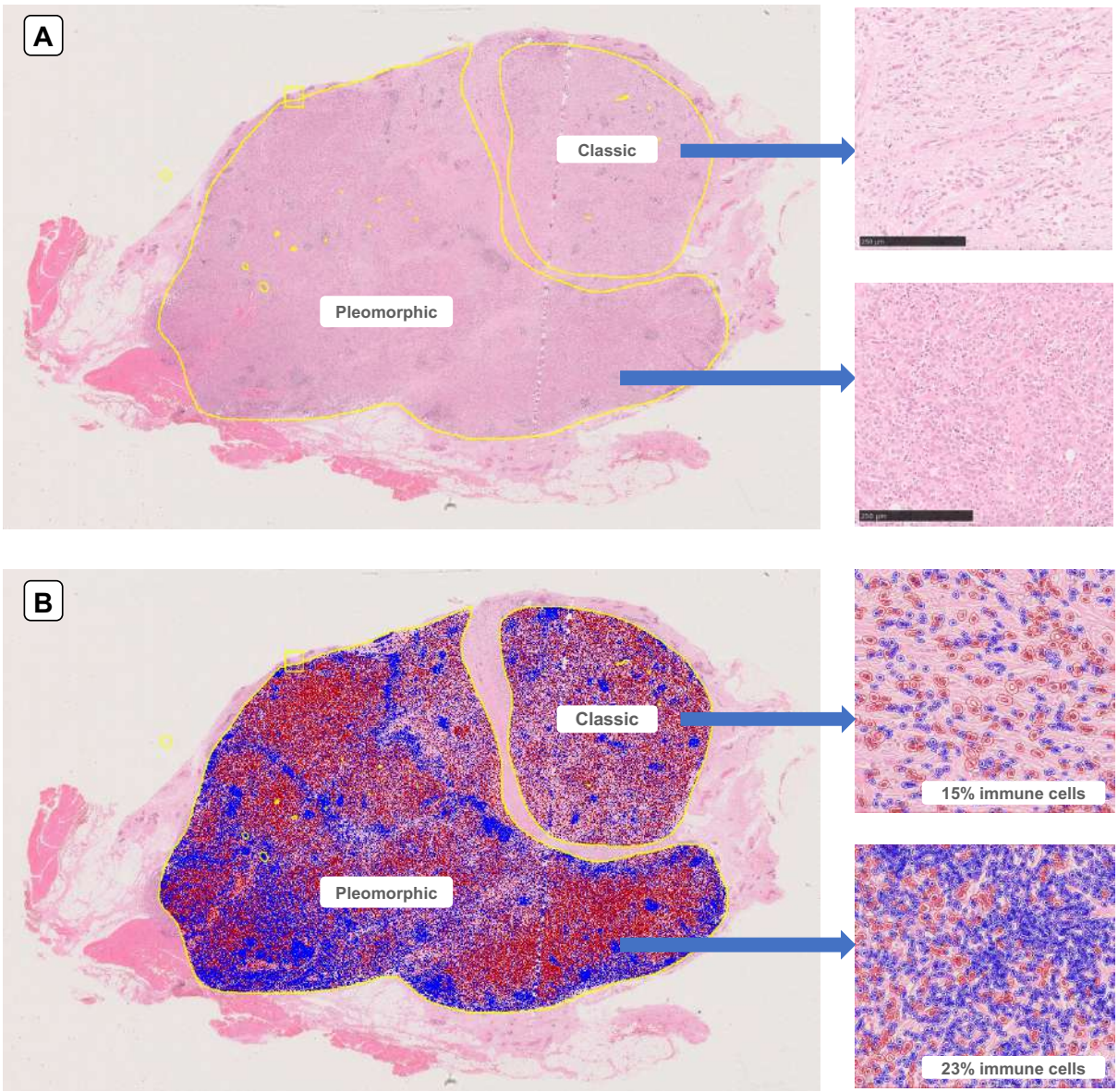


Figure 4.3: Heterogeneity of the immune infiltrate at the histological level in ILC: A) Representative mixed ILC case containing classic and pleomorphic ILC components at low and high power indicating the different areas B) QuPath image of the same case showing immune cells (blue) and tumour cells (red) indicating a higher density of immune cells in the pleomorphic area.

4.33 Classic and alveolar subtypes have lower TILs than pleomorphic ILC

Given that the non-pleomorphic cases were classified histologically as classic (n = 85), solid (n = 3), alveolar (n = 3) and mixed non-pleomorphic ILCs (n = 9), further analysis was completed to assess for differences in TIL scores between pleomorphic ILC and these additional histological subtypes as summarised in Table 4.2. Classic and alveolar ILCs showed significantly lower stromal TILs compared to the pleomorphic ILCs ($p < 0.001$ and $p = 0.0141$) whilst no significant differences were observed between the solid and mixed non-pleomorphic ILCs and the pleomorphic group. However this analysis was limited by small numbers. Overall no significant differences in TIL scores were identified between classic ILC and the other non-pleomorphic ILC subtypes (Table 4.3), in contrast to previous reports that found lower levels of TILs in the alveolar subtype compared to classic ILC [282].

ILC histological subtype	Number of cases	Median TIL score (%)	Interquartile range (%)	TIL score comparison to pleomorphic ILC (n = 63) p - value (Mann-Whitney U test)
Pleomorphic	63	4.0	2.0 - 6.5	NA
Classic	85	2.0	1.0 - 3.5	< 0.0001
Solid	3	3.0	1.0 - 3.0	0.249
Alveolar	3	1.0	0.0 - 2.0	0.0141
Mixed without pleomorphic component	9	3.0	1.0 - 5.75	0.374

Table 4.2: Summary of comparisons of stromal TIL scores between non-pleomorphic ILC subtypes and pleomorphic ILC

ILC histological subtype	Number of cases	Median TIL score (%)	Interquartile range (%)	TIL score comparison to classic ILC (n = 85) p - value (Mann-Whitney U test)
Alveolar	3	1.0	0.0 - 2.0	0.127
Solid	3	3.0	1.0 - 3.0	0.915
Mixed without pleomorphic component	9	3.0	1.0 - 5.75	0.462

Table 4.3: Summary of comparisons of stromal TIL scores between non-pleomorphic ILC subtypes and classic ILC

4.34 The level of stromal TILs is not influenced by hormone status in ILC

Given that stromal TIL levels are known to be higher in TNBC and HER2+ disease, the association between hormone status and immune infiltrate in pleomorphic and non-pleomorphic ILC was assessed. ER status information was available for 59 of 63 pleomorphic patients (51 (86.4%) ER+ and 8 (13.6%) ER-). The median TIL score in ER+ patients was 4.0% compared to 5.75% in the ER- group. There was no significant difference in stromal TIL scores between the ER+ and ER- groups (p = 0.1843, Mann-Whitney U test, Figure 4.4A). All the non-pleomorphic cases were ER+ (n = 100). HER2 receptor status was available for 52 pleomorphic patients and of these 8 (15.4%) were HER2+. The median stromal TILs score in the HER2+ group was 6% compared to 4% in the HER2- group. There was no significant difference in stromal TIL scores according to HER2 status (p = 0.1059, Mann-Whitney U test, Figure 4.4B). Within the non-pleomorphic group HER2 status information was available for 86 patients and of these 3 cases (3.49%) were HER2+. These were all classed histologically as classic ILCs. There was no difference in the level of TILs between the HER2+ and HER2- classic ILC patients (p = 0.2499, Mann-Whitney U test).

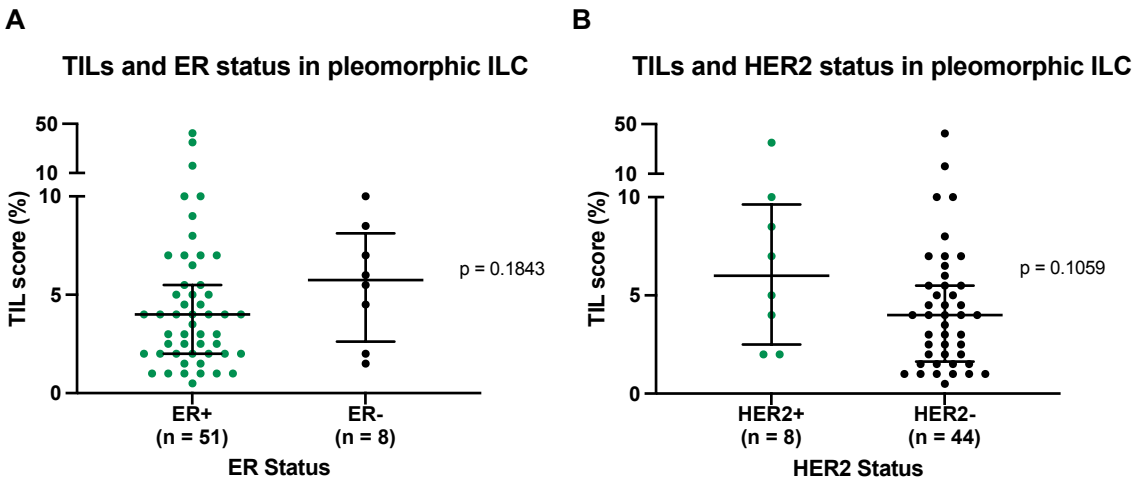


Figure 4.4: Stromal TILs are not associated with hormone status in pleomorphic ILC: A) Scatter plot showing TIL scores for pleomorphic ILC patients with ER+ (n = 51) vs ER- (n = 8) disease showing no difference in TIL scores between these two groups (p = 0.1843, Mann-Whitney U test) B) Scatter plot showing TIL scores for pleomorphic ILC patients with HER2+ (n = 8) vs HER2- (n = 44) disease showing no difference in TIL scores between these two groups (p = 0.1059, Mann-Whitney U test).

4.35 *TP53* and *ARID1A* mutations are associated with the level of stromal TILs in pleomorphic ILC

Having found no differences in the level of stromal TILs according to hormone status in the cohort, we next sought to establish any associations between the level of stromal TILs and the presence of frequently occurring genomic alterations in pleomorphic (n = 52) and non-pleomorphic ILC (n = 43). We previously identified a significantly higher rate of *TP53* mutations as well as higher TILs in pleomorphic vs non-pleomorphic ILC. Within the pleomorphic group, *TP53* mutant tumours (n = 12) were associated with a significantly higher level of stromal TILs compared to *TP53* wild-type cases (n = 40) (p = 0.0358, Mann-Whitney U test, Figure 4.5A, C, Table 4.4). Among non-pleomorphic ILC, *TP53* alterations were less common and there was no difference in the level of stromal TILs in patients with *TP53* alterations (n = 2 mutations, n = 1 deletion) and *TP53* wild-type cases (n = 40) (p = 0.895, Mann-Whitney U test, Table 4.4).

When *ARID1A* mutation status was considered, significantly lower stromal TILs were present in *ARID1A* mutant (n = 8) compared to wild-type (n = 44) pleomorphic ILCs (p = 0.0118, Mann-Whitney U test, Figure 4.5B, C, Table 4.4). *ARID1A* mutations were less common in non-pleomorphic ILC and there was no difference in stromal TILs between *ARID1A* mutant (n = 2) and wild-type (n = 41) cases (p = 0.2735, Mann-Whitney U test, Table 4.4).

A further assessment was made of any association between the level of stromal TILs and genomic alterations affecting *CDH1*, *HER2* and the *PIK3CA/AKT1/PTEN* signalling pathway which showed no differences in the level of stromal TILs according to these alterations in both pleomorphic and non-pleomorphic ILC (Table 4.4).

	Comparison Group: (altered vs wild-type)	No of cases	Median TIL score (%)	Interquartile range (%)	p - value
Pleomorphic ILC (n = 52)	*CDH1*	32 vs 16	4.00 vs 3.75	2.00 - 5.875 vs 2.125 - 7.00	0.8100
	PIK3CA/AKT1/PTEN	21 vs 31	5.00 vs 4.00	2.75 - 6.75 vs 1.50 - 7.00	0.1320
	HER2	10 vs 42	3.25 vs 4.00	1.375 - 8.875 vs 2.00 - 6.625	0.7440
	TP53	12 vs 40	5.75 vs 3.25	4.00 - 9.25 vs 2.00 - 5.50	0.0358
	ARID1A	8 vs 44	2.00 vs 4.25	1.125 - 3.635 vs 2.50 - 7.00	0.0118
Non-pleomorphic ILC (n = 43)	*CDH1*	30 vs 13	2.00 vs 2.00	1.375 - 3.625 vs 1.00 - 6.25	0.8480
	PIK3CA/AKT1/PTEN	25 vs 18	2.00 vs 2.00	1.00 - 4.25 vs 1.375 - 5.625	0.5660
	TP53	3 vs 40	2.00 vs 2.00	1.00 - 5.00 vs 1.125 - 4.375	0.8950
	HER2	2 vs 41	2.00 vs 2.00	2.00 - 2.00 vs 1.00 - 4.75	0.8680
	ARID1A	2 vs 41	4.00 vs 2.00	3.00 - 5.00 vs 1.00 - 4.25	0.2740

Table 4.4: TILs and genomic alterations in pleomorphic and non-pleomorphic ILC: Summary table showing associations between the frequency of genomic alterations (mutations and CNAs) and the level of stromal TILs in pleomorphic and non-pleomorphic ILC (p - value is from the Mann-Whitney U test, with blue box indicating significant results).

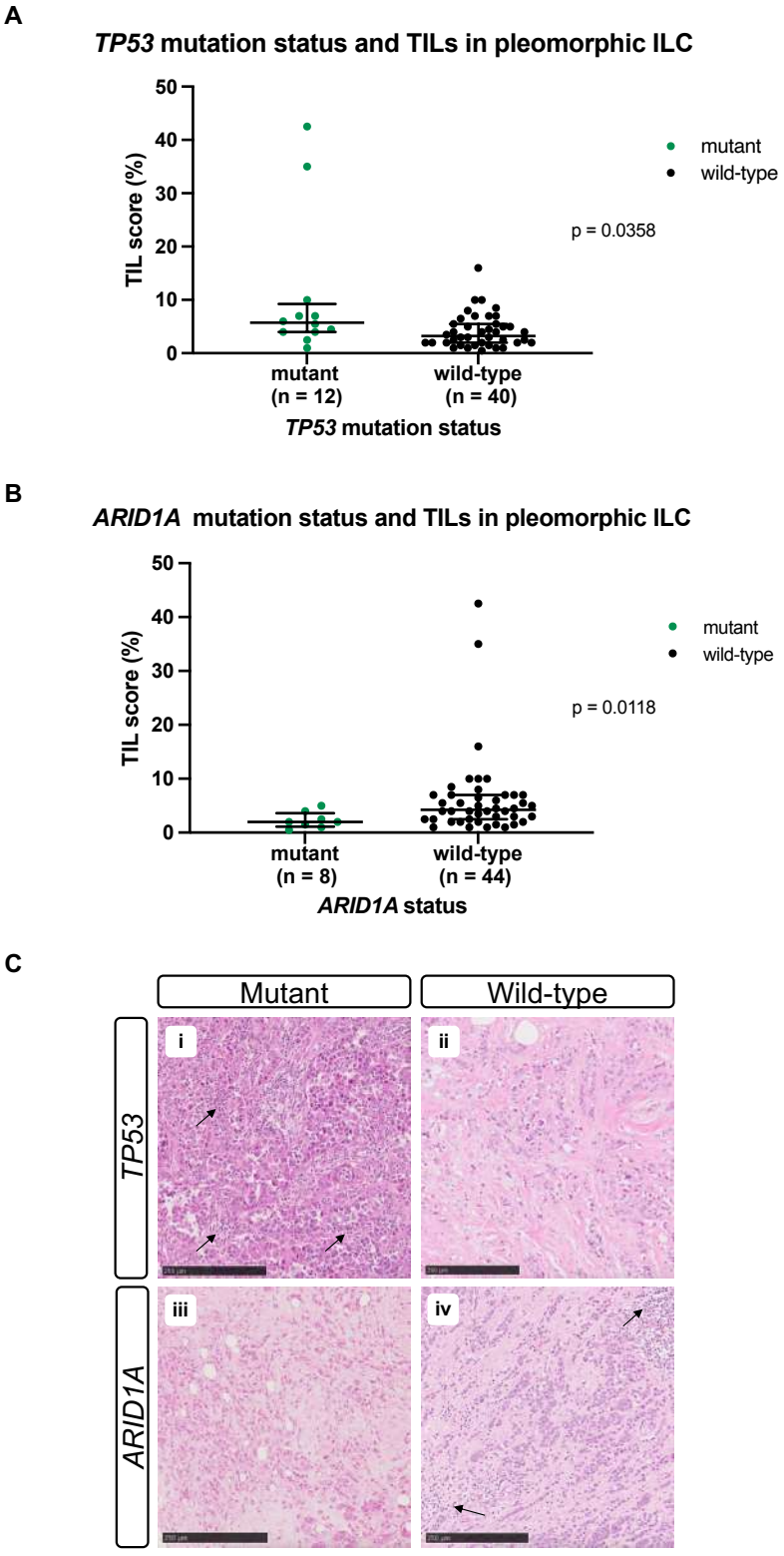


Figure 4.5: *TP53* and *ARID1A* mutations are associated with the level of stromal TILs in pleomorphic ILC: A) Scatter plot showing significantly higher stromal TILs in *TP53* mutant (n = 12) vs wild-type (n = 40) pleomorphic ILCs (p = 0.0358, Mann-Whitney U test) B) Scatter plot showing significantly lower stromal TILs in *ARID1A* mutant (n = 8) vs wild-type (n = 44) pleomorphic ILCs (p = 0.0118, Mann-Whitney U test) C) Representative H&E images showing: top panel i) *TP53* mutant case TG01725 with high stromal TILs (35%) compared to ii) *TP53* wild-type case 17058347 with low stromal TILs (1.5%), bottom panel: iii) *ARID1A* mutant case 17051250 showing low stromal TILs (1%) and iv) *ARID1A* wild-type case 17011789 showing higher stromal TILs (10%) (stromal TILs are depicted by arrows).

4.36 There is no correlation between tumour mutational load and TILs

Having assessed the association between stromal TILs and specific frequently occurring genomic alterations in ILC, an assessment was next made to establish whether the total mutational load correlates with the density of immune infiltrate in ILC. The level of infiltration by TILs has been hypothesized to reflect mutational load [270, 409, 410]. To better understand any association between the number of mutations in each tumour and stromal TILs, a Spearman's rank test was performed in the pleomorphic and non-pleomorphic cases where mutational data was available (n = 96). This failed to show any significant correlation between the two variables in the ILC cohort (Spearman's rank correlation coefficient (ρ) = -0.014, p = 0.896, Figure 4.6).

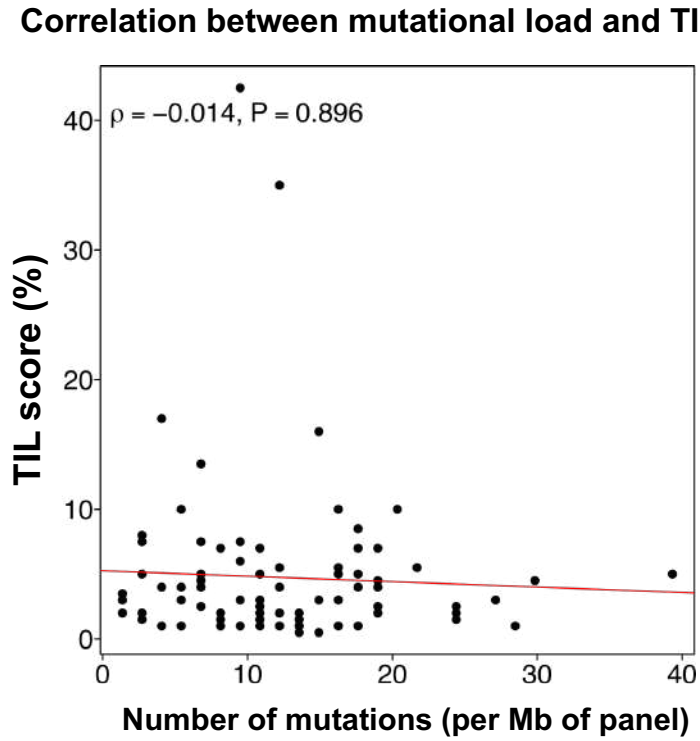


Figure 4.6: There is no correlation between mutational load and stromal TILs in ILC: Scatter plot showing the correlation between the number of mutations and stromal TILs in each ILC tumour (Spearman's rank correlation coefficient (ρ) = -0.014, p = 0.896, n = 96)

4.37 Stromal TILs are not associated with clinical outcome in the cohort

Given that previous studies have associated a high level of stromal TILs with poor prognostic factors [282, 339] and worse survival outcomes in ILC in one study [339], an assessment was made of the association between the level of stromal TILs and clinical outcome in the KHP cohort. Univariable survival analysis using a cox-proportional hazards model was performed with stromal TILs used as an independent variable. For analysis of TIL scores, each individual score was categorised as either ≤ 5% or > 5%. Given that pleomorphic ILCs show higher stromal TILs compared to non-pleomorphic cases, a separate analysis was performed in pleomorphic and non-pleomorphic patients. Patients with *de novo* metastatic disease were removed prior to survival analysis and the follow-up time was truncated at 15 years, with patients with metastatic events occurring after 15 years being censored. Overall there was no difference in MFS between the patients with stromal TIL scores ≤ 5% vs > 5% in both pleomorphic (n = 61, p = 0.609, HR: 0.78, CI: 0.29 - 2.05, Figure 4.7A) and non-pleomorphic ILCs (n = 93, p = 0.372, HR: 0.61, CI: 0.21 - 1.79, Figure 4.7B). There was also no difference in OS between the patients with stromal TIL scores ≤ 5% vs > 5% in both pleomorphic (n = 63, p = 0.205, HR: 0.55, CI: 0.21 - 1.39, Figure 4.7C) and non-pleomorphic ILCs (n = 100, p = 0.383, HR: 0.68, CI: 0.29 - 1.62, Figure 4.7D).

As previously mentioned, 9 patients (2 pleomorphic and 7 non-pleomorphic) presented with *de novo* metastatic disease and were excluded from survival analysis. To analyse the immune infiltrate in these patients and other early relapsing ILCs compared to late relapsing patients, a comparison of stromal TIL scores in early (< 3 years) vs late (> 6 years) relapse was made, excluding patients without disease relapse. This showed no difference in stromal TIL scores between early relapsing (including *de novo* metastatic patients) and late relapsing patients in both pleomorphic (n = 21) and non-pleomorphic (n = 40) patients (p = 0.9965 and p = 0.3732 respectively, Mann-Whitney U test, Figure 4.7E, F).

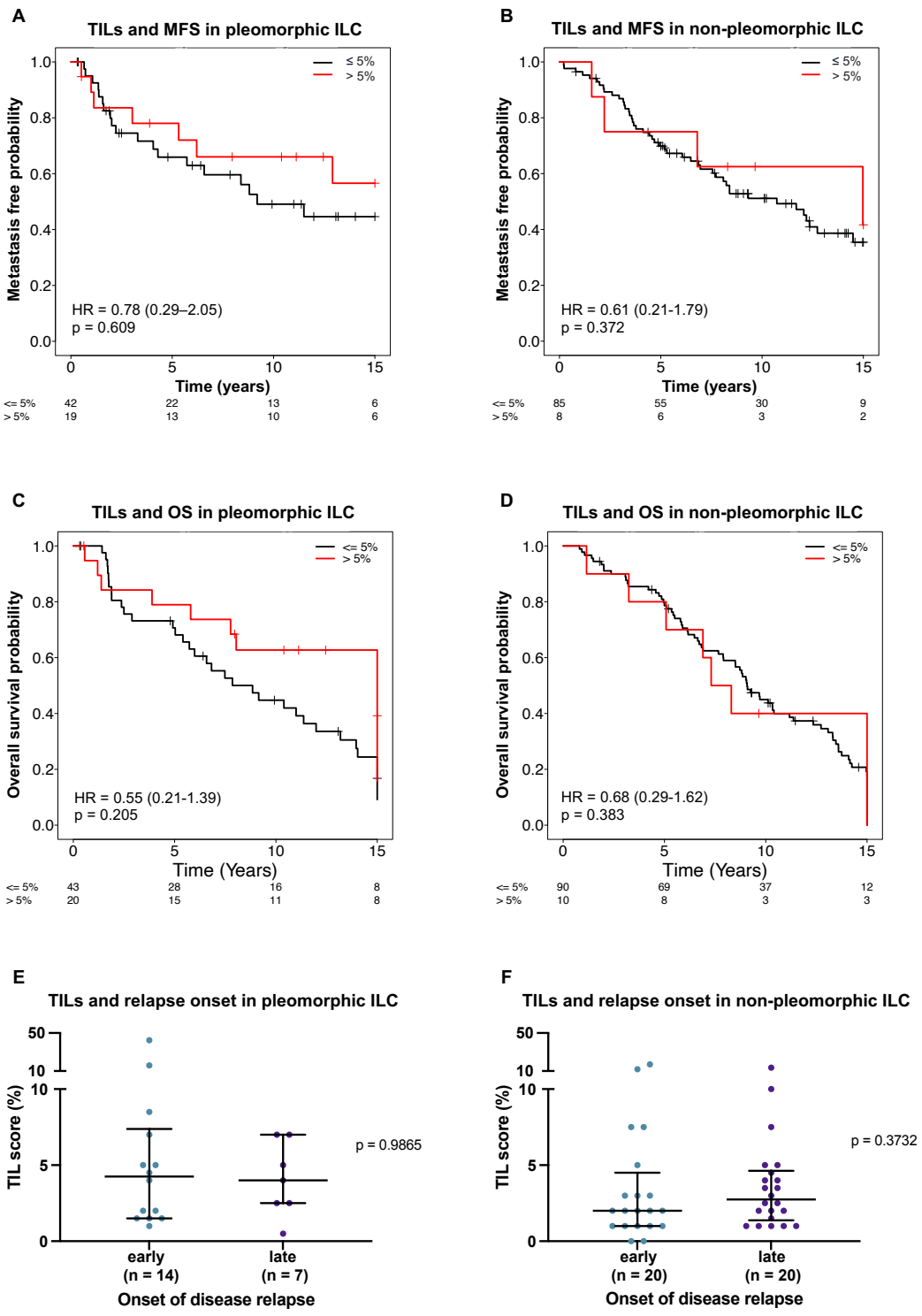


Figure 4.7: Stromal TILs are not associated with clinical outcome in the KHP cohort: A) KM graph showing the association between TILs and MFS in pleomorphic ILC using TILs groups ≤ 5% and >5% (n = 61, p = 0.609, HR: 0.78, CI: 0.29 - 2.05) B) KM graph showing the association between TILs and MFS in non-pleomorphic ILC (n = 93, p = 0.372, HR: 0.61, CI: 0.21 - 1.79) C) KM graph showing the association between TILs and OS in pleomorphic ILC (n = 63, p = 0.205, HR: 0.55, CI: 0.21 - 1.39) D) KM graph showing the association between TILs and OS in non-pleomorphic ILC (n = 100, p = 0.383, HR: 0.68, CI: 0.29 - 1.62) E) Scatter plot showing TILs in early vs late relapsing pleomorphic ILCs (n = 21, p = 0.9865, Mann-Whitney test) F) Scatter plot showing TILs in early vs late relapsing non-pleomorphic ILCs (n = 40, p = 0.3732, Mann-Whitney test).

4.4 Discussion

Characterising a cohort of ILCs histologically with a particular focus on pleomorphic ILC reveals new insights into the immune landscape of this histologically rare, poorly understood and clinically aggressive ILC subtype.

Consistent with previous studies [282, 339], when using stromal TILs as a measure of tumour immunogenicity, this work shows that ILCs are generally a weakly immunogenic tumour type with the majority of the combined pleomorphic and non-pleomorphic cohort falling into the low TIL score group. However a minority of all cases fell into the high TIL score group suggesting that a subset of ILCs is more immunogenic. Interestingly a lower proportion of ILCs fell into the 'high' (> 10%) TIL score group (4.3%) compared to a previous study which classified 15% of ILCs as having stromal TILs > 10%. This may be reflective of differences in the proportions of the different ILC histological subtypes within the two different cohorts but may also reflect some degree of inter-observer variability in TIL scoring by different pathologists in different centres. Nevertheless, it is important to recognise the wider context, that 'immune-hot' or 'lymphocyte-predominant' breast cancers (LPBCs), which are generally TNBCs, are defined by stromal TIL scores > 50% [411] and therefore even the small proportion of ILCs with TIL scores > 10% are still relatively weakly immunogenic compared to LPBCs and indeed none of the ILCs studied in the KHP ILC cohort showed stromal TILs > 50%.

Previous studies assessing TILs classified ILCs based upon architectural features e.g. classic, solid, alveolar, and mixed, however at the cytological level pleomorphic cases were not explicitly identified or studied [282, 339]. The results presented in this chapter provide new insights into the immune landscape of pleomorphic ILC and show that it is associated with a significantly higher level of stromal TILs compared to non-pleomorphic ILC and this holds true after having controlled for pleomorphic-enriched hormonal phenotypes (ER-/HER2+ patients).

When the other ILC histological subtypes were considered, a previous large study identified a significantly lower level of stromal TILs in alveolar ILC compared to classic ILC [282]. Whilst this association was not observed in the KHP cohort, the presence of only 3 alveolar cases in the cohort resulted in a lack of power to detect significant differences. This was also true for the other uncommon subtypes in the cohort including solid and mixed non-pleomorphic tumours and therefore small numbers of the rarer ILC subtypes represents a limitation of this part of the study.

TNBCs and HER2+ breast cancers are characterised by a higher level of stromal TILs compared to ER+ tumours [396]. Consistent with the fact that the vast majority of non-pleomorphic ILCs are of a luminal A phenotype, in the KHP ILC cohort all the non-pleomorphic cases were ER+ and just 3 cases were HER2+. Consistent with the higher rates of TNBCs and HER2+ disease in pleomorphic ILC, higher numbers of these cases were observed in the pleomorphic group. Interestingly within the pleomorphic group there was no association between hormone status and the level of immune infiltrate. A previous study quantifying stromal TILs in 614 ILCs found that ER- negative tumours (but not HER2+ tumours) were associated with a higher level of stromal TILs [282]. However the study did not identify pleomorphic cases, and since pleomorphic ILC shows higher rates of ER- tumours [77], it is possible that here ER- tumours were more indicative of pleomorphic histology which itself is associated with higher stromal TILs rather than specifically ER status, which is what has been identified in this work. A larger number of ER- and HER2+ cases would help further with establishing any association between hormone status and TILs in pleomorphic ILC since the KHP cohort is limited by relatively small numbers of these cases.

Whilst hormone status did not show an association with stromal TILs in the KHP ILC cohort, analysis at the genomic level identified that *TP53* mutations are associated with a denser stromal lymphocytic infiltrate in pleomorphic ILC. Interestingly *TP53* mutations have previously been shown to promote immunogenicity in breast cancer [412] although not directly at the histological level and not previously in ILC or pleomorphic ILC, a rare, clinically aggressive subtype, where higher rates of *TP53* alterations are observed. A recent study compared the level of enrichment of 26 immune signatures, which reflected the activities of a diverse range of immune cells and their associated pathways, between *TP53* mutant and *TP53* wild-type breast cancers using a range of different PAM-50 breast cancer cases from the TCGA and METABRIC datasets [244]. The *TP53*-mutant tumours exhibited significantly stronger immune signatures in almost all of the 26 signatures analysed [412]. The findings from the KHP ILC cohort are therefore consistent with this and suggest that in pleomorphic ILC *TP53* mutations promote immunogenicity which is reflected at the histological level.

Aside from individual genomic alterations, when the total mutational load was considered, whilst pleomorphic ILCs had a significantly higher mutational load and higher stromal TILs compared to non-pleomorphic ILC, there was no direct correlation between tumour mutational load and the level of stromal TILs in the ILC cohort. A previous study assessed tumour mutational load and correlation with TILs in a cohort of 62 advanced breast cancers and found a significant positive correlation between these two variables [413]. However the study cohort only contained 6 ILCs with the remainder of cases consisting mainly of ER+ and triple-negative

IC-NSTs [413]. The findings through sequencing of the larger KHP ILC cohort (n = 95) and correlation with stromal TILs suggest that in ILC, as a special histological breast cancer subtype, the level of stromal TILs cannot be considered to accurately represent the underlying level of mutational load of the tumour.

As previously discussed, there are several ongoing phase III trials assessing the efficacy of ICIs in the ER+ breast cancer population include CHECKMATE 7FL, KEYNOTE-B49 and KEYNOTE-756 [257-259]. Furthermore specifically in ILC, the phase II GELATO trial is assessing the efficacy of atezolizumab in heavily pre-treated patients with advanced disease [260]. Studying the associations between baseline stromal TILs and clinical response to checkpoint inhibition specifically in the ILC patients within these trials will help elucidate the clinical relevance and utility of stromal TILs assessment in ILC as an adjunct to more frequently used assessments of tumour immunogenicity such as PD-L1 expression, as assessed through IHC.

Overall the characterisation of the immune landscape in ILC at the histological level in this chapter, with a particular focus on pleomorphic cases, has provided new insights into the relatively enhanced immunogenicity of this rare histological ILC subtype, as well as associations with genomic alterations. The work has further highlighted that the level of stromal TILs assessed at the gross histological level does not appear to hold prognostic significance in both pleomorphic and non-pleomorphic ILC.

Chapter 5: Mapping the heterogeneity of the immune infiltrate in pleomorphic ILC

5.1 Background

Based upon the quantification of stromal TILs as a measure of tumour immunogenicity, the majority of ILCs are weakly immunogenic, or 'immune-cold' with only small subsets of ILC patients, including pleomorphic ILCs showing relatively higher stromal TILs. However, despite this there appears to be evidence that at least a subset of ILCs is more immunogenic aside from stromal TILs alone. For example, 'immune-related' gene-expression subtypes have been identified in ILC patients from the RATHER and TCGA cohorts (n = 106 and n = 144 respectively) [32, 277].

There is an increased recognition that given the complexity and heterogeneity of the tumour immune microenvironment, considering stromal TILs alone is a relatively simplistic approach. The tumour immune microenvironment represents a highly dynamic network of interactions between various lymphoid and myeloid lineage cell subpopulations of the innate and adaptive immune systems, including T and B lymphocyte subsets, dendritic cells, and macrophages [308, 414, 415]. The abundance of immune subpopulations and their interactions can either create an inflammatory microenvironment which favours suppression of the tumour or an immuno-suppressive microenvironment which promotes tumour progression.

In a recent study, the phenotypic, transcriptional and functional diversity of TILs was investigated in a cohort of primary untreated ER+ ILCs (n = 87) and IC-NSTs (n = 94) using flow cytometry, multiplexed IHC and single cell RNA sequencing technologies [340]. This identified that macrophages, rather than T lymphocytes, were the predominant immune subpopulation infiltrating the tumour bed in ILC and the most transcriptionally diverse immune subpopulation between the ILC and IC-NST [340]. Tumours with abundant macrophages but minimal stromal TILs would be considered 'immune-cold' based on traditional assessments of histological immunogenicity quantifying stromal TILs, although this may not be accurate given that macrophages represent a key additional immune subpopulation with diverse roles in the complex tumour immune microenvironment. However macrophages and their prognostic significance has not been characterised in pleomorphic ILC and there is a need to better understand their presence and prognostic associations in this clinically aggressive tumour type and in particular the significance of the presence of M1-like and M2-like macrophages.

Whist M1-type macrophages contribute to an inflammatory immune microenvironment favouring tumour suppression, M2-type macrophages contribute to an immuno-suppressive microenvironment [332].

Another key element of the tumour immune microenvironment to consider is spatial heterogeneity. Immunity typically functions through intercellular contacts and through short-distance cytokine communications and therefore understanding the spatial location of immune cells with respect to tumour cells and other immune subpopulations can offer a framework for understanding the unique ILC disease biology and for identifying potential predictive biomarkers associated with clinical outcome [416]. Immune cells themselves may be arranged individually dispersed throughout the tumour or alternatively as clusters. Tertiary lymphoid structures (TLSs) are ectopic lymphoid aggregates, located in non-lymphoid tissues [26]. They are composed of T cell rich areas, B cells and supporting dendritic cells and are responsible for driving an antigen-specific immune response. They typically develop in inflamed tissues, being associated with chronic inflammation, autoimmune conditions, and malignancy [26]. They recruit and activate TILs therefore acting as a site of immune response activation within the tissue microenvironment [28]. In the context of cancer, they have attracted interest as being associated with improved outcomes in some cancers, including in TNBC and HER2+ breast cancer [29, 30]. In addition, they have been associated with improved responses to immunotherapies and there is growing interest in the possibility of targeting TLSs to enhance treatment responses [27]. However, the presence of TLSs has not been studied in pleomorphic ILC. Immune-related spatial analysis is emerging as a new type of cancer research for studying tumour immune heterogeneity from a spatial perspective.

Given the diverse functions of TILs within the tumour immune microenvironment there is a need to further characterise TILs into the various immune subpopulations in ILC, and to establish their prognostic significance particularly in pleomorphic ILC, a histologically more immunogenic yet understudied ILC subtype. With the development of immune-related spatial analysis technologies, there is a need to better understand the spatial organisation of immune cells and any prognostic associations in pleomorphic ILC. Moreover given the growing evidence that macrophages are a more dominant immune subpopulation than TILs in ILC, with diverse functions depending on whether they have been polarised to an M1-like or M2-like phenotype, there is a need to study their presence further in ILC and establish any prognostic associations as well as any differences in the levels of macrophages between pleomorphic and non-pleomorphic ILC.

5.2 Chapter Aims

Overall the purpose of this chapter is to:

1) Evaluate the prognostic associations of the immune microenvironment in aggressive ILCs.

2) Assess the spatial significance of the various immune subpopulations.

3) Functionally assess the role of immune subpopulations associated with aggressive ILCs.

5.3 Results

5.31 Pleomorphic ILCs show heterogeneity in the composition and spatial organisation of the immune infiltrate

The abundance of immune subpopulations was assessed at the protein level in a high-plex unbiased manner using the NanoString Digital Spatial Profiling (DSP) platform. Review of the NanoString tumour H&Es (Figure, 5.1 & 5.2) and immunofluorescent (IF) images showing the selected morphology markers, guided the selection of regions of interest (ROIs) (Figure 5.3). ROIs were selected based on CD45 and Pan-Cytokeratin (PanCk) morphology marker staining with immune rich tumour regions selected. Eighty-four ROIs were represented in 7 NanoString slides (Table 5.1). One of the original 20 tumours (TG02210 on slide 7) was excluded due to poor tissue adherence to the slide. Visualisation and quantification of morphology markers demonstrated heterogeneity and variability in immune cell densities, immune cell subtypes, and tumour cell content across each ROI, within and between tumours (Figure 5.4).

Seventy-one protein markers (excluding controls) were measured across each ROI (Supplementary Table 3). Normalized counts of each target relative to non-specific counts (background) were evaluated. Some non-immune markers were captured in the analysis despite selecting quantification of the immuno-oncology proteins in CD45+ve areas only. This occurred if there was overlap between immune cells and epithelial cells within the selected immune compartment. Given that protein expression was being characterised exclusively in immune cells, the non-immune markers were excluded from the analysis. Quantification of the 71 immuno-oncology protein markers further demonstrated immune heterogeneity across ROIs (Figure 5.5A). Principal component analysis showed that ROIs clustered based on patient but not slide location (Figure 5.5B, 5.5C).

	Study ID	ER	PR	HER2	Stromal TILs score (%)	lymph nodes (n)	Size (mm)	Grade	Age at diagnosis (years)	Onset of relapse
SLIDE 1	1355	1	0	0	7	2	22	2	62.5	2
	17011789	1	1	0	10	0	40	3	47.9	3
	17013592	1	1	0	5	6	40	3	57.2	1
SLIDE 2	17014363	1	0	1	7	22	40	3	53.9	1
	17015028	1	1	0	8	0	60	3	79.7	3
	17018625	1	1	0	7	8	50	2	47.3	3
SLIDE 3	17021357	0	0	0	5.5	NA	30	NA	53.5	3
	17023306	0	0	1	8.5	33	30	3	58.1	1
	17013387	1	0	NA	7	1	60	3	67.3	2
SLIDE 4	17025981	1	1	1	5	2	50	3	88.1	1
	17047324	1	1	0	16	33	40	2	66.6	1
	17051586	0	0	0	7	0	10	2	66.1	2
SLIDE 5	17055238	1	1	0	7.5	0	40	2	80.5	3
	17055645	0	0	0	6	0	10	2	43.4	2
	17057784	0	0	0	4.5	11	40	2	60.8	3
SLIDE 6	TG01139	1	1	0	5.5	3	40	2	49.4	3
	TG01725	NA	NA	1	35	1	11	3	42.7	3
	TG01958	0	NA	0	10	5	2	2	34.3	3
SLIDE 7	TG02166	1	1	0	5	0	40	3	56.1	NA
	TG02210	1	1	0	6.5	0	13	3	57.5	3

Table 5.1: Clinical details and outcome data for pleomorphic ILC DSP cohort: (For hormone status: NA = unavailable information, 0 = negative, 1 = positive. For onset of relapse: 1 = early relapse < 3 years, 2 = late relapse > 6 years, 3 = no relapse, NA = relapse but onset unknown).

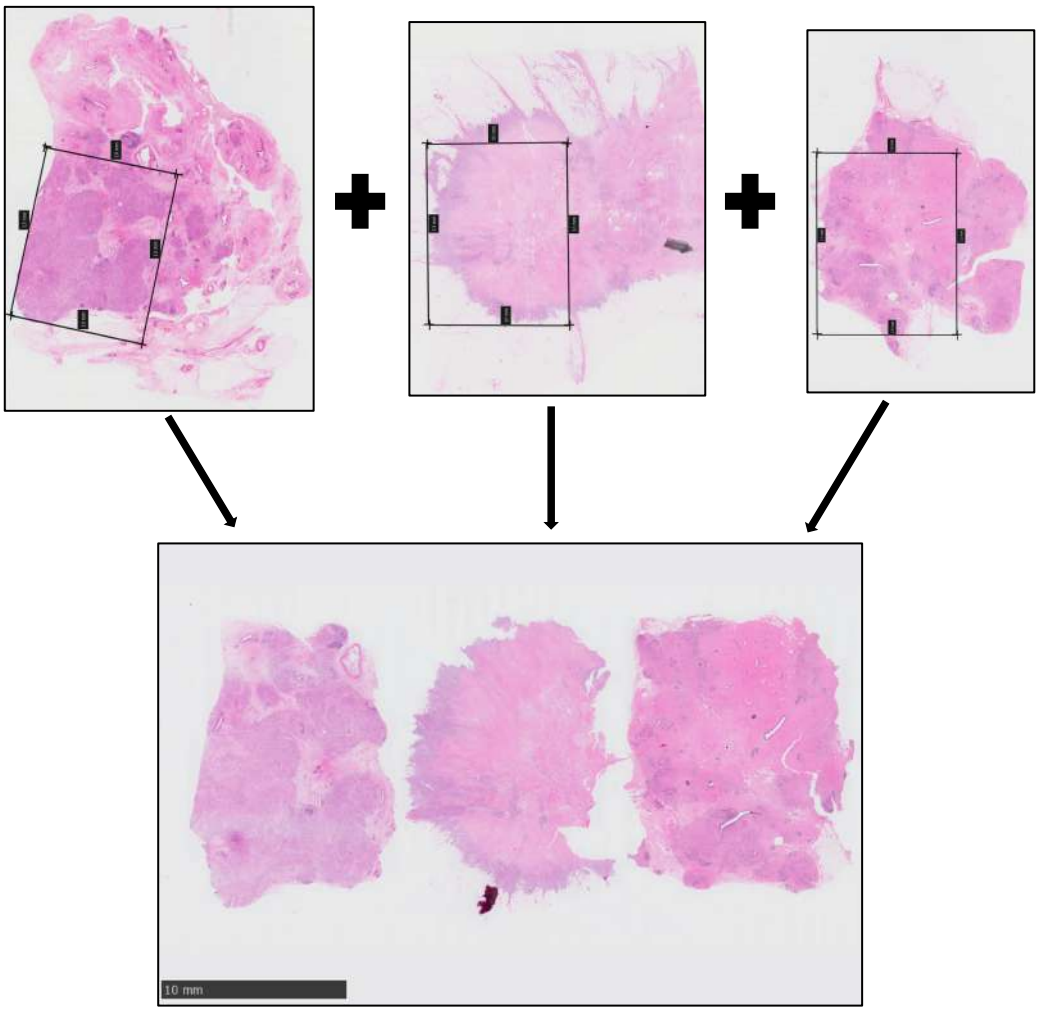


Figure 5.1: Selection of 10 x 13mm tumour regions of pleomorphic ILCs on the NanoString DSP slides: Representative slide showing three 10 x 13mm areas selected from individual tumour cases (left to right: 17014363, 17015028, 17018625) for slide 2 to capture the most immunogenic areas. Multiple selected tumour areas were then orientated on each NanoString DSP slide.

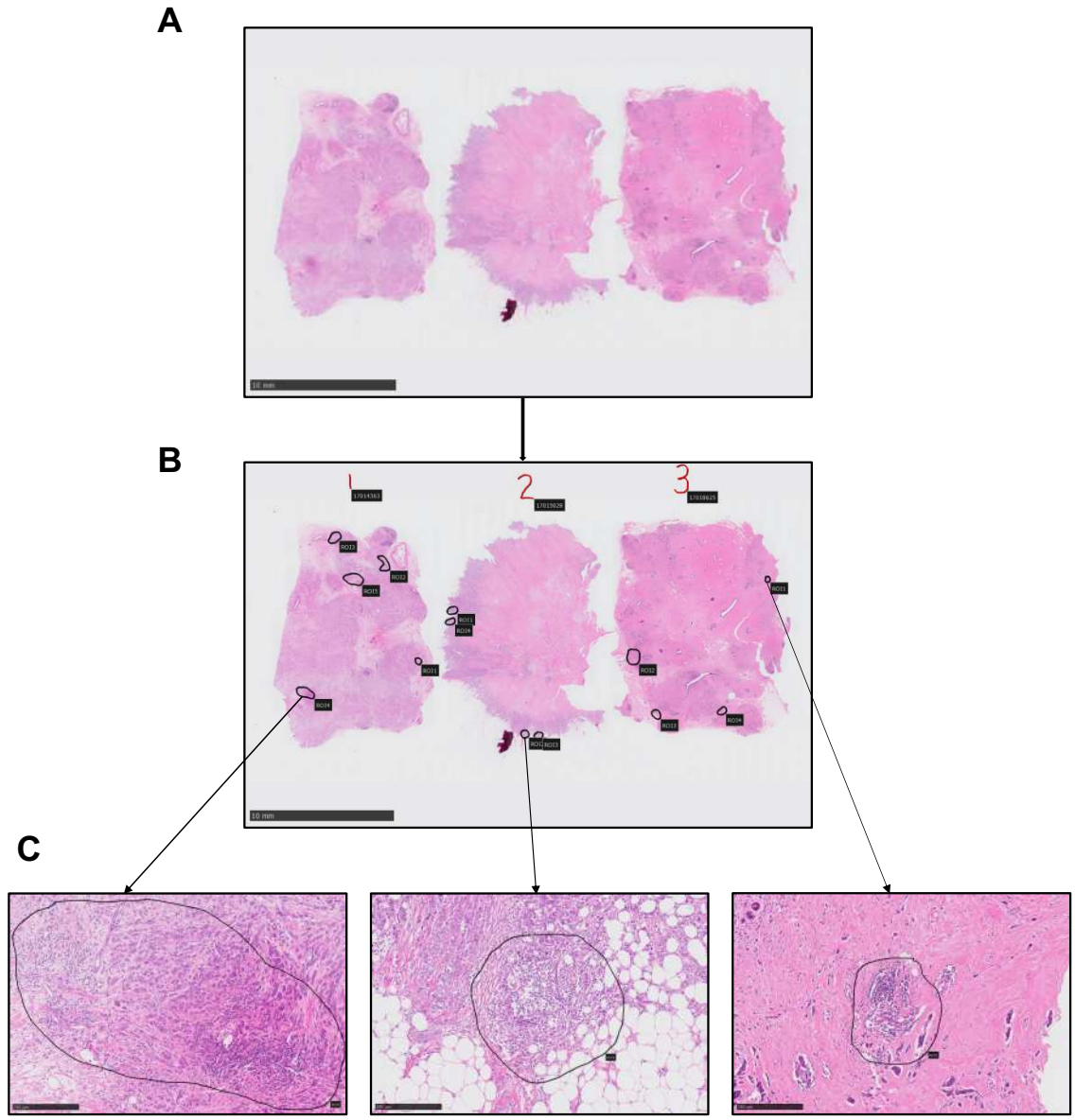


Figure 5.2: Review of tumour H&Es identifies immune rich areas for NanoString DSP: A) Representative example showing three individual tumour cases from slide 2 (left to right: 17014363, 17015028, 17018625) on a single NanoString DSP slide B) Histological review of the tumour sections was used to select the areas with the most immune cells as potential ROIs prior to visualization with morphology markers C) High power images of individual immune rich areas selected for each tumour case.

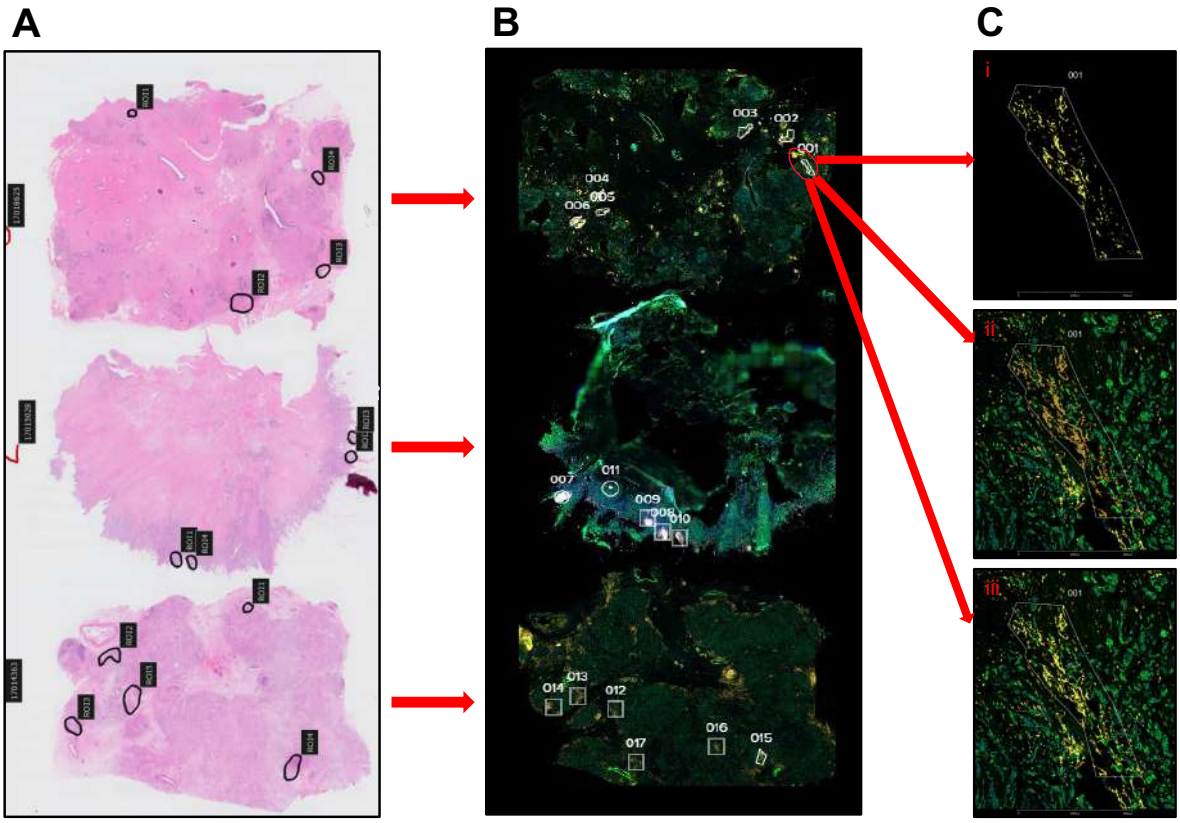


Figure 5.3: Morphology marker Immunofluorescence (IF) imaging enables selection of immune rich tumour regions: A) Representative H&E showing three individual tumour cases from slide 2 (superior to inferior: 17018625, 17015028, 17014363) on a single NanoString DSP slide with immunogenic areas identified by black circular annotations prior to live image viewing B) Live IF image showing morphology markers (PanCk = green, CD45 = yellow, CD3 = red, DNA = blue) and all ROIs selected upon review of the live image (additional ROIs were selected during review of the live IF images and immune rich areas identified on H&E did not always directly correlate exactly with the most immunogenic areas on live image review, since this was a subsequent tumour tissue section, and there were also subtle differences in tissue orientation between the H&E and live IF image).

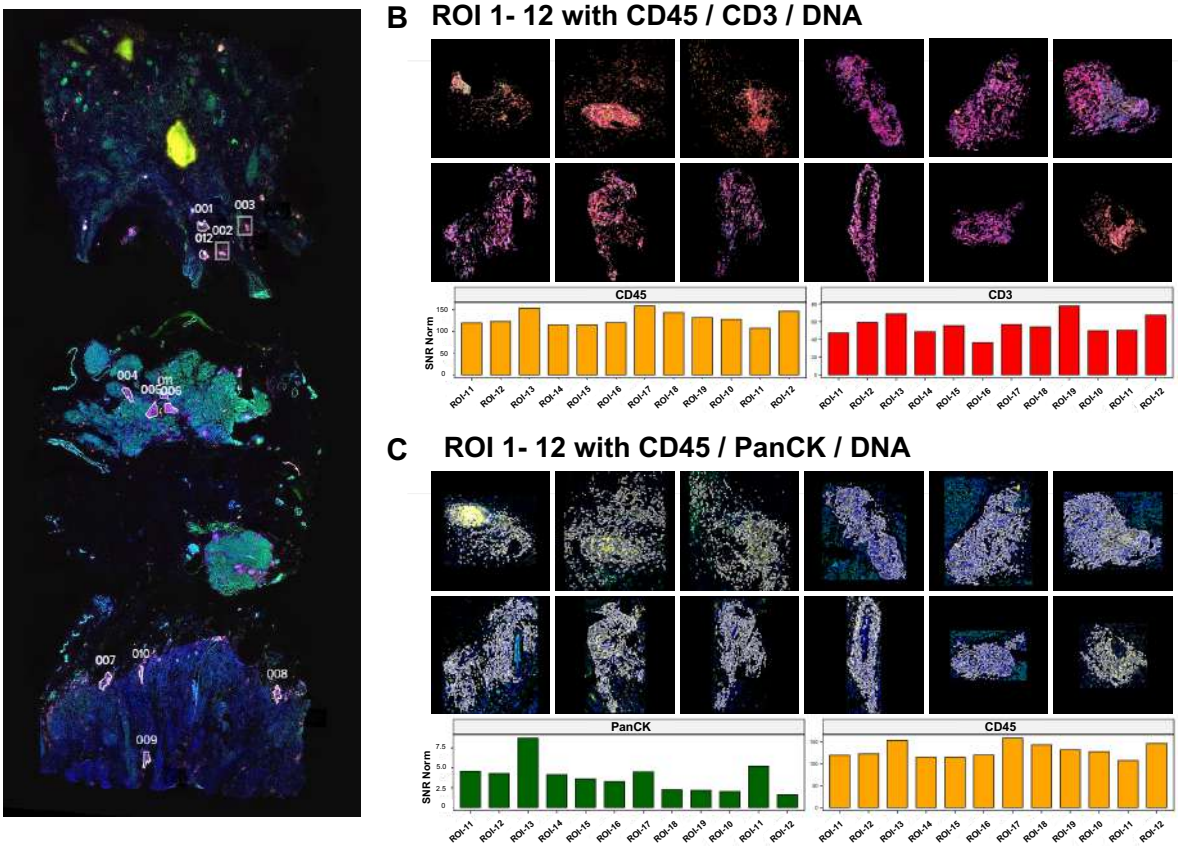


Figure 5.4: Representative IF NanoString DSP imaging demonstrates immune heterogeneity across selected ROIs: A) IF image of slide 1 (Superior to inferior: 17013592, 17011789, 1355) showing morphology markers: PanCK (green), CD45 (yellow), CD3 (red), DNA (blue) and 12 selected ROIs B) IF images showing individual ROIs with CD45, CD3 and DNA channels and underlying bar charts showing normalised morphology marker counts across each individual ROI for CD45 and CD3 C) IF images showing individual ROIs with PanCK, CD45 and DNA channels and underlying bar charts showing normalised morphology marker counts across each individual ROI for PanCK and CD45.

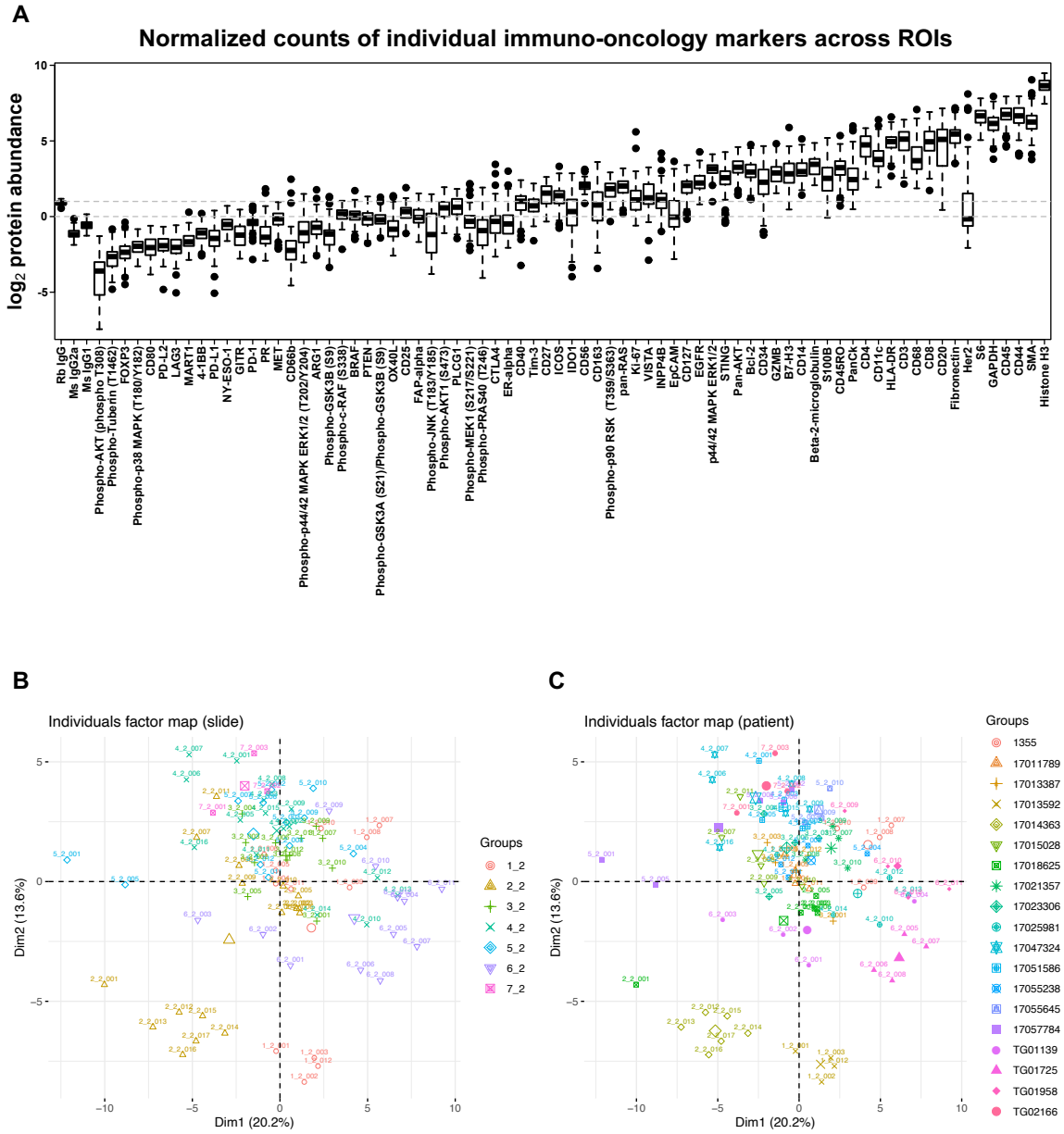


Figure 5.5: Heterogeneity in the expression of immuno-oncology proteins is observed across ROIs: A) Normalised counts of immuno-oncology protein markers across selected ROIs B) Principal component analysis assessing for slide effect: individuals factor map assessing clustering of ROIs based upon slide location (individual datapoints represent ROIs with ROIs colour-coded based upon slide location) C) Principal Component analysis assessing for patient effect: individuals factor map assessing clustering of ROIs based upon patient (individual datapoints represent ROIs with ROIs are colour-coded based upon patient).

5.32 CD11c+ cells are associated with improved outcome in pleomorphic ILC

A linear mixed model was used to assess differential protein expression in i) relapse (n = 10) vs no-relapse (n = 9) and ii) early (n = 5) vs late (n = 4) relapsing patients. Given that this was a small discovery cohort, relaxed statistical thresholds were applied (un-adjusted $p < 0.1$, LFC +/- 0.585) which found a significant enrichment of CD11c+ cells in patients who did not relapse vs those that did ($p = 0.0533$, LFC = -0.657, Figure 5.6A). CD11c is widely established as a marker of dendritic cells but is also expressed on monocytes, M1-like macrophages, granulocytes and memory B cells [417, 418]. In the early vs late relapse comparison CD11c+ cells were not significantly differentially expressed between the two patient groups ($p = 0.348$, LFC = -0.257, Supplementary Table 4).

No other significantly differentially expressed immuno-oncology proteins were identified between relapsing and non-relapsing pleomorphic ILC patients (Table 5.2). Non-immune markers e.g. HER2 were not studied further given that CD45+ areas only were sequenced and non-immune signal reflects overlap between immune and epithelial cells.

5.33 CD68+ cells are associated with early onset of disease relapse in relapsing patients

Comparing early (n = 5) vs late (n = 4) relapsing patients, a significantly higher level of CD68+ cells was observed in patients who relapsed early ($p = 0.0553$, LFC = 1.608, Figure 5.6B). CD68 is a pan-macrophage marker expressed on both M1-like and M2-like macrophages. CD68 levels were not significantly differentially expressed in the relapse vs no-relapse comparison ($p = 0.731$, LFC = 0.172). When CD163, a marker of M2-like macrophages was considered, there was no difference in CD163 levels in the relapse vs no relapse group ($p = 0.767$, LFC = 0.151) and in the early vs late relapse group ($p = 0.321$, LFC = 0.842). No other immune proteins were significantly differentially expressed between the two groups (Table 5.3). Next an average M2/M1 ratio was calculated for each patient based upon the levels of CD163 (M2-like) to CD80 (M1-like) protein expression across all the ROIs from each patient. There was no difference in the M2/M1 ratios between relapse (n = 10) and no-relapse (n = 9) ($p = 0.7802$, Mann-Whitney U test), and between early (n = 5) and late relapse (n = 4) ($p = 0.7302$, Mann-Whitney U test, Supplementary Table 5).

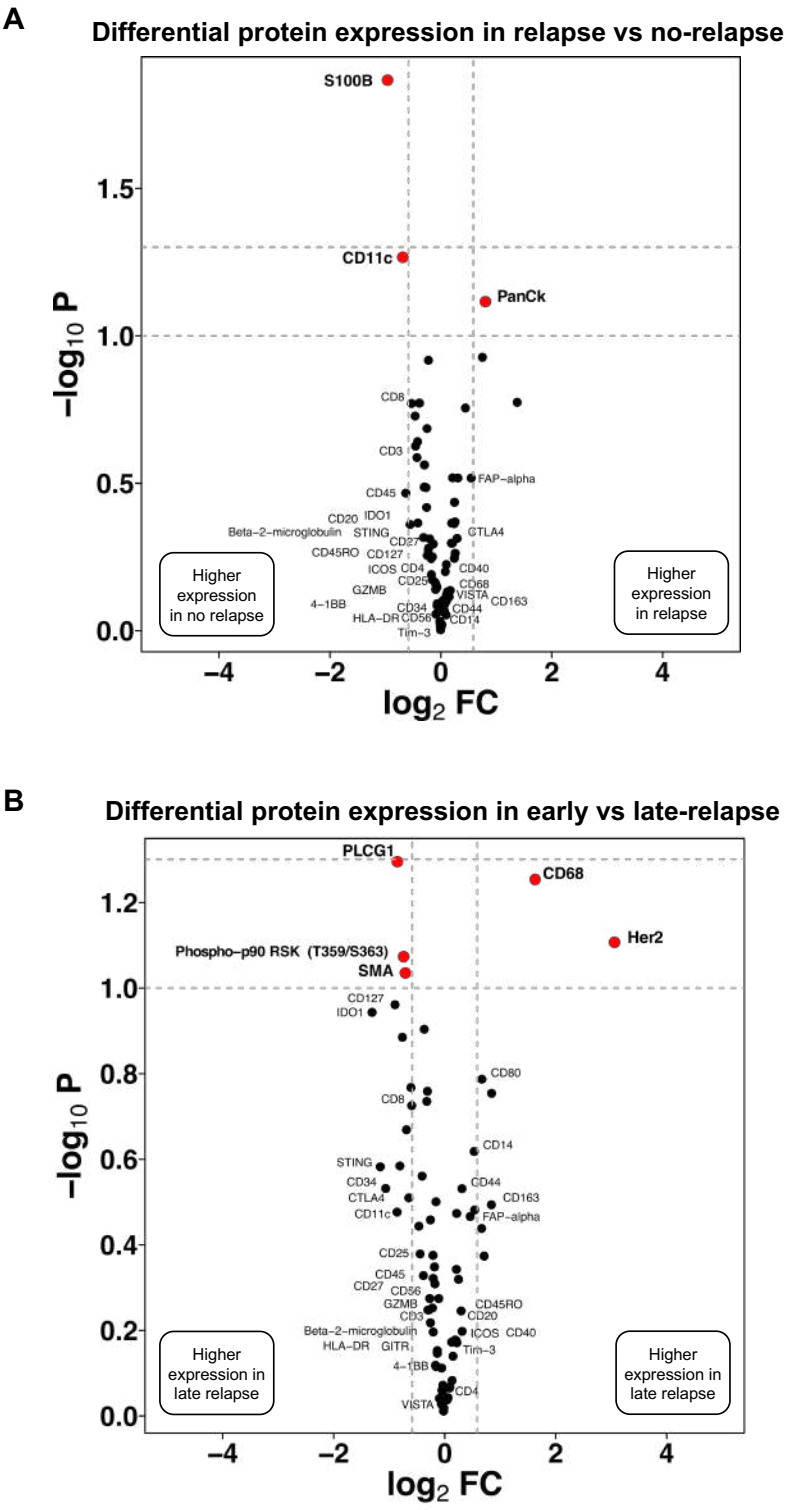


Figure 5.6: CD11c+ cells are enriched in no-relapse patients and CD68+ cells are enriched in early relapsing patients: A) Volcano-plot showing differential expression of proteins from the NanoString DSP panel in no-relapse (n = 9) vs relapse (n = 10). CD11c+ cells were more highly expressed in no-relapse (p = 0.0533, LFC = -0.657 B) Volcano-plot showing differential expression of proteins from the NanoString DSP panel in early (n = 5) vs late (n = 4) relapse. CD68+ cells were more highly expressed in patients who relapsed early (p = 0.0553, LFC = 1.608) (red dots show statistically significant proteins). Non-immune significantly differentially expressed proteins e.g. S100B were excluded from further analysis. Non-significant non-immune related protein labels are excluded for visualisation purposes.

Relapse vs no relapse			
Protein	**Cell type / function**	**Logfold change**	**p-value**
CD11c	Dendritic cells / M1 macrophages / monocytes / memory B cells	-0.657	0.053
CD8	Cytotoxic T cells	-0.529	0.170
CD3	All T lymphocytes	-0.430	0.259
IDO1	Suppression of effector T cells and natural killer (NK) cells, activation of regulatory T cells and myeloid derived suppressor cells (MDSCs)	-0.414	0.431
CD20	B lymphocytes	-0.554	0.436

Table 5.2: Function of 5 differentially expressed immune protein markers in pleomorphic ILC tumours of patients with and without disease relapse [419-426]

Early vs late relapse			
Protein	**Cell type / function**	**Logfold change**	**p-value**
CD68	Macrophages	1.608	0.055
CD127	Memory and effector T cells	-0.895	0.109
IDO1	Suppression of effector T cells and natural killer (NK) cells, activation of regulatory T cells and myeloid derived suppressor cells (MDSCs)	-1.308	0.114
CD8	Cytotoxic T cells	-0.595	0.188
CD14	Monocytes, granulocytes and macrophages	0.530	0.241

Table 5.3: Function of 5 differentially expressed immune protein markers in pleomorphic ILC tumours of patients with early and late relapse [422, 426-429]

5.34 *TP53* mutations are associated with higher levels of CD56+ cells

Next an assessment was made to establish any correlation between the presence of the frequently observed genomic alterations in pleomorphic ILC and the levels of the major immune subpopulations. These immune subpopulations of interest included T cells (CD3+), cytotoxic T cells (CD8+), helper T cells (CD4+), B cells (CD20+), macrophages (CD68+) and natural killer cells (CD56+). Targeted sequencing data was available for 18 of the 19 patients included in the study. The levels of these immune subpopulations were compared between i) *CDH1* mutant vs wild-type patients ii) *PIK3CA/PTEN/AKT* mutant vs wild-type iii) *HER2* amplified/mutant vs wild-type and iv) *TP53* mutant vs wild-type. This showed significantly higher normalised counts of CD56+ cells in *TP53* mutant (n = 6) vs wild-type (n = 12) patients (p = 0.0135, Mann-Whitney U test). There were no other significant differences between the levels of the major immune cell populations and subpopulations (Table 5.4).

Comparison Group	No of cases	Median SNR normalised expression counts	p - value (Mann-Whitney U test)
CDH1 altered vs wild-type **CD3**	13 vs 5	4.949 vs 4.570	0.6331
CDH1 altered vs wild-type **CD8**	13 vs 5	5.019 vs 4.359	0.3359
CDH1 altered vs wild-type **CD4**	13 vs 5	4.747 vs 4.751	> 0.9999
CDH1 altered vs wild-type **CD20**	13 vs 5	4.414 vs 5.416	0.5028
CDH1 altered vs wild-type **CD68**	13 vs 5	3.385 vs 3.908	0.3873
CDH1 altered vs wild-type **CD56**	13 vs 5	1.973 vs 2.092	0.3359
PIK3CA/PTEN/AKT altered vs wild-type **CD3**	11 vs 7	4.833 vs 5.320	0.1509
PIK3CA/PTEN/AKT altered vs wild-type **CD8**	12 vs 7	4.486 vs 5.081	0.4252
PIK3CA/PTEN/AKT altered vs wild-type **CD4**	13 vs 7	4.310 vs 5.204	0.0853
PIK3CA/PTEN/AKT altered vs wild-type **CD20**	14 vs 7	4.414 vs 4.437	0.4789
PIK3CA/PTEN/AKT altered vs wild-type **CD68**	15 vs 7	3.404 vs 3.209	0.7242
PIK3CA/PTEN/AKT altered vs wild-type **CD56**	16 vs 7	2.008 vs 2.035	0.5962
HER2 altered vs wild-type **CD3**	3 vs 15	5.219 vs 4.930	0.3603
HER2 altered vs wild-type **CD8**	3 vs 15	4.655 vs 5.019	> 0.9999
HER2 altered vs wild-type **CD4**	3 vs 15	5.204 vs 4.310	0.2500
HER2 altered vs wild-type **CD20**	3 vs 15	4.437 vs 4.134	0.5735
HER2 altered vs wild-type **CD68**	3 vs 15	4.502 vs 3.385	0.4265
HER2 altered vs wild-type **CD56**	3 vs 15	2.253 vs 2.008	0.4265
TP53 altered vs wild-type **CD3**	6 vs 12	5.134 vs 4.907	0.5532
TP53 altered vs wild-type **CD8**	6 vs 12	5.094 vs 4.731	0.8201
TP53 altered vs wild-type **CD4**	6 vs 12	4.868 vs 4.251	0.3355
TP53 altered vs wild-type **CD20**	6 vs 12	3.886 vs 4.431	0.8916
TP53 altered vs wild-type **CD68**	6 vs 12	3.943 vs 3.282	0.0831
TP53 altered vs wild-type **CD56**	6 vs 12	2.185 vs 1.861	0.0135*

Table 5.4: Associations between genomic alterations and the normalised counts of immune cell populations in the NanoString DSP cohort: This identified significantly higher normalised counts of CD56+ cells in *TP53* mutant (n = 6) vs wild-type (n = 12) patients (p = 0.0135, Mann-Whitney U test) (* denotes p value < 0.05).

5.35 The spatial organisation of immune cells is not associated with clinical outcome in the KHP pleomorphic cohort

One of the key advantages of NanoString DSP technology is that it provides morphological context rather than quantification data alone. Therefore having assessed differential expression of the various immune proteins between patients with varying clinical outcomes, an assessment was next made of the spatial organisation of the immune cells and any prognostic associations.

During H&E review and ROI selection, heterogeneity with respect to the spatial organisation of immune cells was noted (Figure 5.7A). Some tumours were characterised by well-developed and organised lymphoid structures in association with tumour cells whilst other showed a significantly more diffuse infiltrate characterised by single immune cells dispersed throughout the tumour and stroma. To further understand the prognostic significance of the degree of spatial clustering of immune cells, each individual ROI was analysed in turn using the IF images containing the morphology markers and the proportion of the total CD45+ cells that were forming clusters or immune cell aggregates was quantified. Given that each patient's tumour contained multiple ROIs, an average cluster proportion score was calculated for each patient (Table 5.5). The cluster proportions were then compared between patients with and without disease relapse. The hypothesis was that patients without disease relapse would have higher cluster scores than relapsing patients given that TLSs have previously been associated with improved outcomes in solid tumours [29, 30]. However there was no difference in the degree of clustering of immune cells between relapsing (n = 10) and non-relapsing patients (n = 9) ($p = 0.9556$, Mann-Whitney U test, Figure 5.7B).

Previous studies have shown that the size and degree of maturation of TLSs can be of prognostic significance in cancer [20-25]. TLSs in the tumour microenvironment can vary from small aggregates of B lymphocytes to more organised lymphoid structures containing germinal centres [26]. There is no widely used cut-off for the numbers of immune cells required to define a TLS, although a recent study assessing TLSs in lung cancer proposed a methodology for automated detection and quantification of TLSs using H&Es which concluded that a minimum number of 45 lymphocytes was required to define a TLS [430] (although mature TLSs with germinal centres can contain several hundred cells). In the NanoString cohort few well-established TLSs with germinal centres were present, yet there was still heterogeneity in the size of the immune clusters which were in association with tumour cells. Therefore to assess any association between immune cell cluster size and clinical outcome in the pleomorphic

cohort, the number of cells in the largest immune cell cluster was visually quantified for each patient and assigned to a cluster group (Table 5.5). Overall there was no difference in the size of the largest immune cell cluster between relapsing and non-relapsing patients ($p = 0.9798$, Mann-Whitney U test, Figure 5.7C). Given that a previous study determined that 45 lymphocytes were required to define a TLS [430], using the immune cluster size cut-offs, 8/19 (42%) of patients included in the NanoString DSP cohort had TLSs (immune clusters with > 50 cells). There was no significant difference in the frequency of TLSs between relapsing and non-relapsing patients ($p > 0.9999$, Fisher's exact test, Table 5.6). When the largest cluster size category of > 100 immune cells was considered, there was no difference in the frequency of large clusters between relapsing and non-relapsing patients ($p > 0.9999$, Fisher's exact test, Table 5.7).

	Study ID	Onset of relapse	Average proportion of CD45+ cells forming clusters across ROIs (%)	Number of cells in largest cluster
SLIDE 1	1355	2	50	2
	17011789	3	90	4
	17013592	1	60	3
SLIDE 2	17014363	1	60	2
	17015028	3	95	4
	17018625	3	0	1
SLIDE 3	17021357	3	80	2
	17023306	1	40	3
	17013387	2	0	1
SLIDE 4	17025981	1	30	2
	17047324	1	90	4
	17051586	2	95	4
SLIDE 5	17055238	3	50	2
	17055645	2	80	4
	17057784	3	0	1
SLIDE 6	TG01139	3	90	2
	TG01725	3	30	2
	TG01958	3	90	4
SLIDE 7	TG02166	NA	0	1

Table 5.5: Clustering of immune cells and relapse status in each NanoString DSP patient: For onset of relapse: 1 = early relapse < years, 2 = late relapse > 6 years, 3 = no relapse, NA = relapse but onset unknown. For number of cells in largest cluster: 1 = no clusters, 2 = 20 - 50, 3 = 50 - 100, 4 = 100).

	clusters > 50 cells	clusters < 50 cells	Total (n)
relapse (n)	4	6	10
no-relapse (n)	4	5	9
Total (n)	8	11	19

Table 5.6: Contingency table showing frequency of TLSs in relapse vs no-relapse: No difference in the frequency of TLSs was observed between relapsing and non-relapsing patients ($p > 0.9999$, Fisher's exact test).

	clusters > 100 cells	clusters < 100 cells	Total (n)
relapse (n)	3	7	10
no-relapse (n)	3	6	9
Total (n)	6	13	19

Table 5.7: Contingency table showing frequency of immune clusters > 100 cells in relapse vs no-relapse: No difference in the frequency of clusters with > 100 cells was observed between relapsing and non-relapsing patients ($p > 0.9999$, Fisher's exact test).

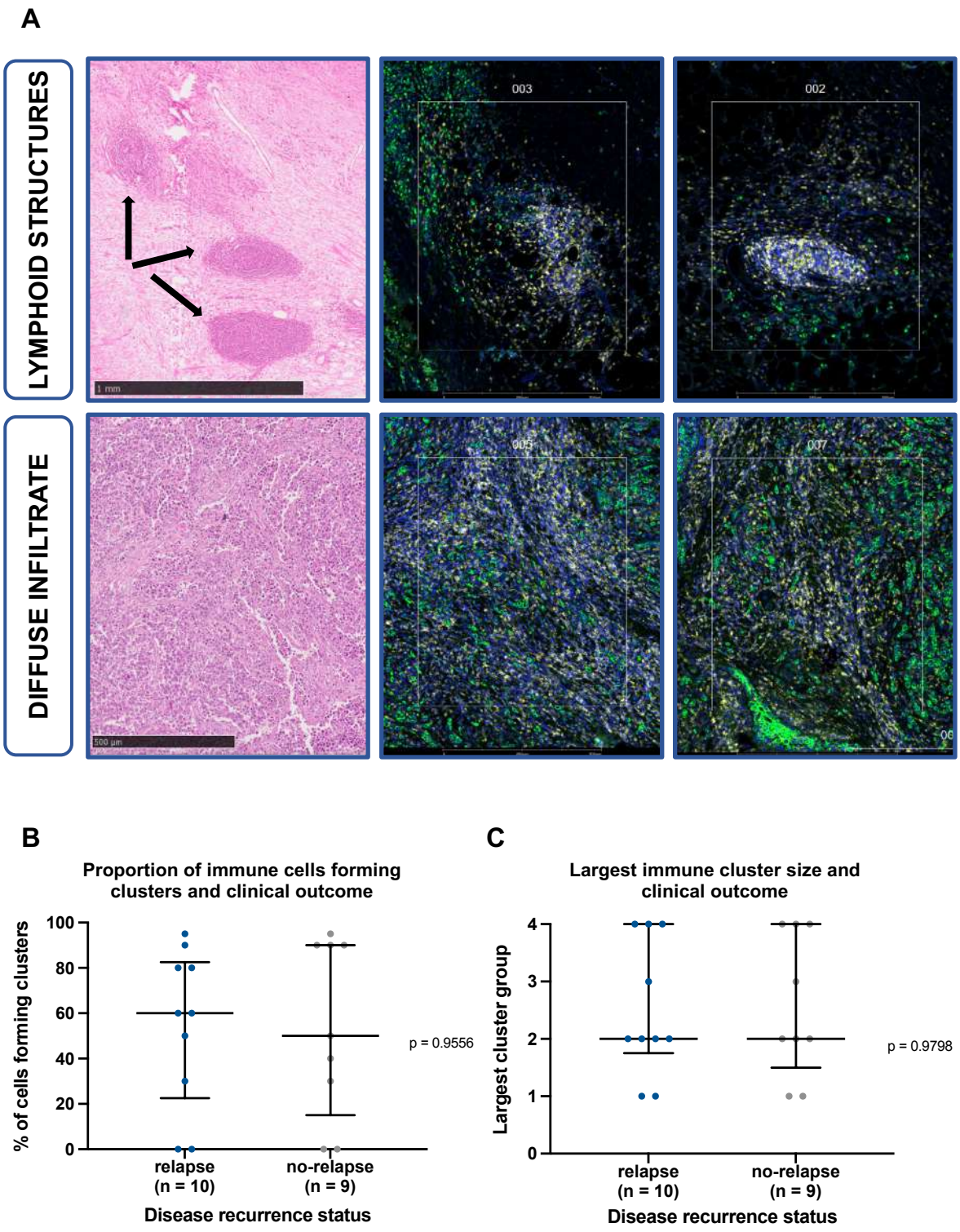


Figure 5.7: The spatial organisation of immune cells is not associated with disease relapse status: A) Top panel: H&E of case 17013592 with arrows showing organised lymphoid structures / aggregates (depicted by arrows) and IF images of two corresponding selected ROIs from the case. Bottom panel: H&E of case TG01725 showing a scattered and diffuse infiltrate of immune cells throughout the tumour and IF images of two corresponding ROIs from the case (PanCK = green, CD45 = yellow, CD3 = red, DNA = blue) B) Scatterplot showing the average % of CD45+ cells forming clusters across ROIs per patient C) Scatterplot showing the number of cells in the largest immune cluster per patient: 1 = no clusters, 2 = 20 - 50, 3 = 50 - 100, 4 = 100).

5.36 Pleomorphic ILCs contain more macrophages than non-pleomorphic ILCs

Having identified a significantly higher level of CD68+ cells in pleomorphic ILC patients with early disease relapse from the NanoString DSP cohort, an independent extended cohort of both pleomorphic and non-pleomorphic ILCs was used to validate this finding. CD68 IHC was performed in the KHP cohort in pleomorphic (n = 54) and non-pleomorphic (n = 78) ILCs. Unlike the Salgado stromal TILs scoring methodology which is a well-established standardised TILs scoring approach in breast cancer, there is no specific CD68 scoring methodology for breast tumours.

The TILs scoring methodology focused only on stromal TILs as opposed to stromal and intra-tumoral TILs as the assessment of stromal TILs is more reproducible. Intra-tumoral TILs are typically present at lower numbers and in fewer breast cancer cases and can be more challenging to identify on basic H&E [308]. Moreover the scoring of intra-tumoral TILs does not provide further prognostic information compared to stromal TILs alone. However when assessing CD68 in the cohort, given that there is no standardised methodology for CD68 staining and little is known of the level of macrophage infiltration in ILC, particularly pleomorphic ILC, it was deemed appropriate to capture and quantify all CD68+ cells in the IHC analysis, therefore both intra-tumoral and stromal macrophages.

QuPath digital pathology technology [341] was used to quantify the level of CD68 staining in selected representative ROIs of each case (Figure 5.8). Quantification of CD68 staining across the cohort showed heterogeneity in the level of macrophage infiltration. Overall when the entire CD68 ILC cohort was considered, pleomorphic ILCs showed significantly higher CD68 scores compared to non-pleomorphic cases (n = 132, p = 0.0002, Mann-Whitney U test, Figure 5.9A), which was maintained when just focussing on ER+/HER2- tumours (n = 33) vs ER+/HER2- non-pleomorphic (n = 64) ILCs (p = 0.0422, Mann-Whitney U test, Figure 5.9B-C). Of note, there was no significant difference in CD68 scores between ER+ and ER- pleomorphic ILCs (p = 0.1275, Mann-Whitney U test, Figure 5.10A). Additionally no significant difference was observed between CD68 scores in HER2+ (n = 7) and HER2- (n = 39) pleomorphic ILCs (p = 0.0758, Mann-Whitney U test, Figure 5.10B).

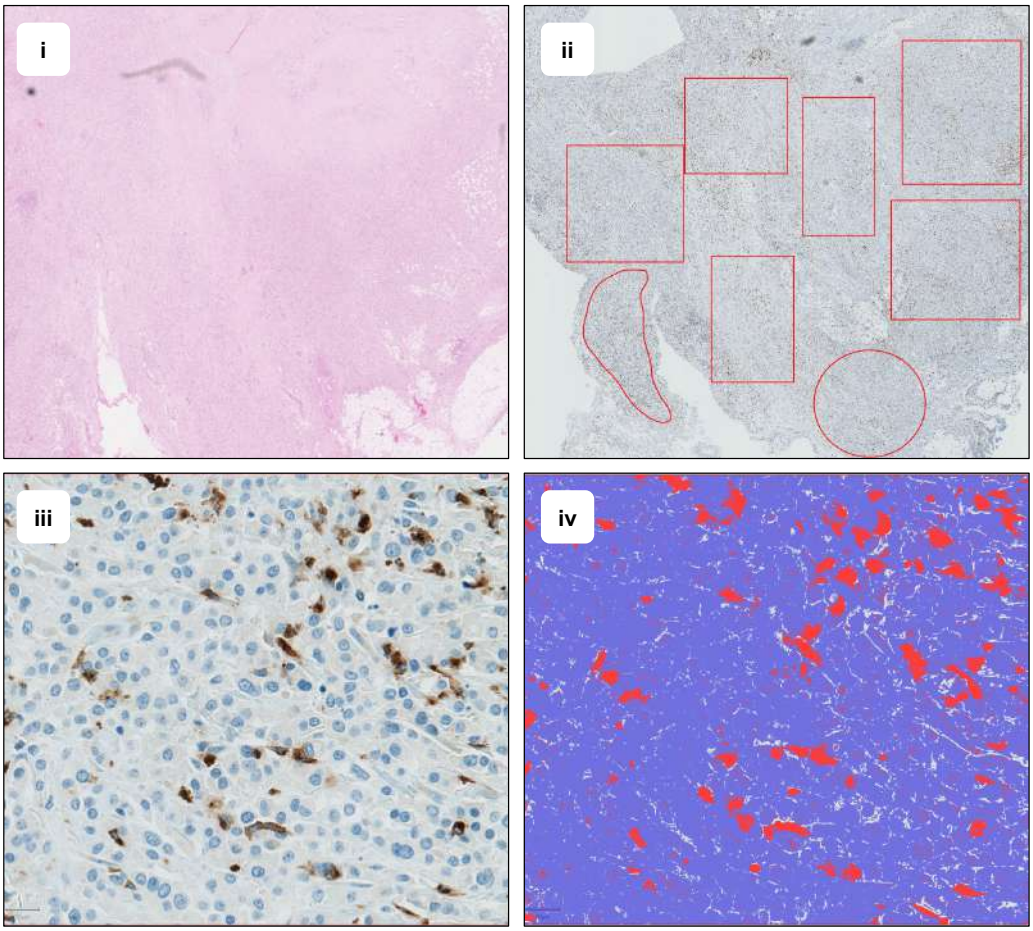


Figure 5.8: Digital pathology quantification of CD68 staining for case 245: i) Low power (x4) H&E of a representative ILC tumour case and ii) corresponding CD68 IHC showing QuPath annotations of representative regions of interest (ROIs) iii) High power (x20) view within an ROI showing CD68 staining and iv) corresponding QuPath image which uses a pixel classifier to identify areas of positive CD68 staining.

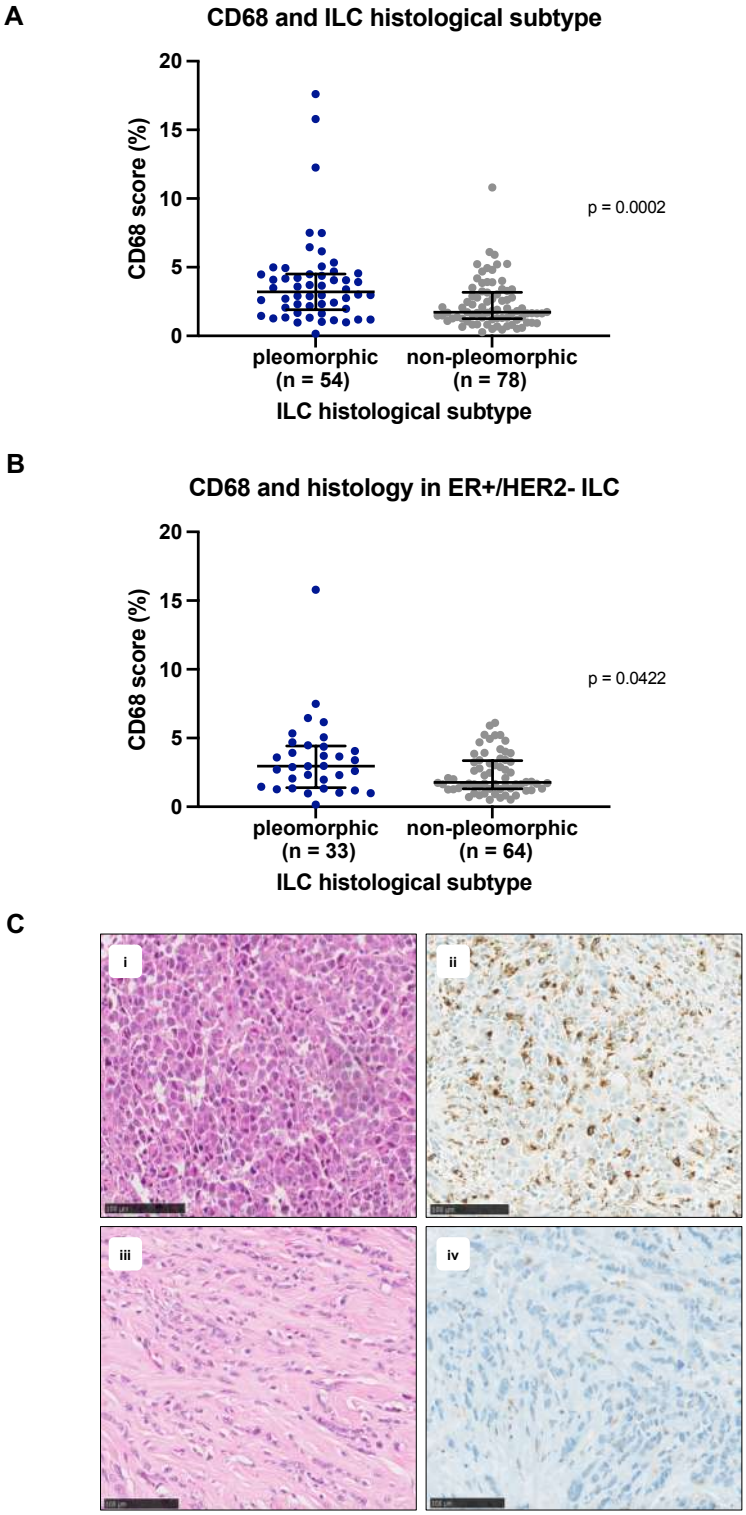


Figure 5.9: Pleomorphic ILCs show a higher level of macrophages compared to non-pleomorphic cases: A) Scatterplot showing significantly higher CD68 scores in pleomorphic vs non-pleomorphic cases (p = 0.0002, Mann-Whitney U test) and B) in ER+/HER2- pleomorphic vs non-pleomorphic cases (p = 0.042, Mann-Whitney U test) C) i) H&E of pleomorphic case TG01725 and ii) corresponding CD68 IHC showing strong CD68 staining (17.6%) iii) H&E of non-pleomorphic case 3589 and iv) corresponding CD68 IHC showing mainly absent staining (1.33%).

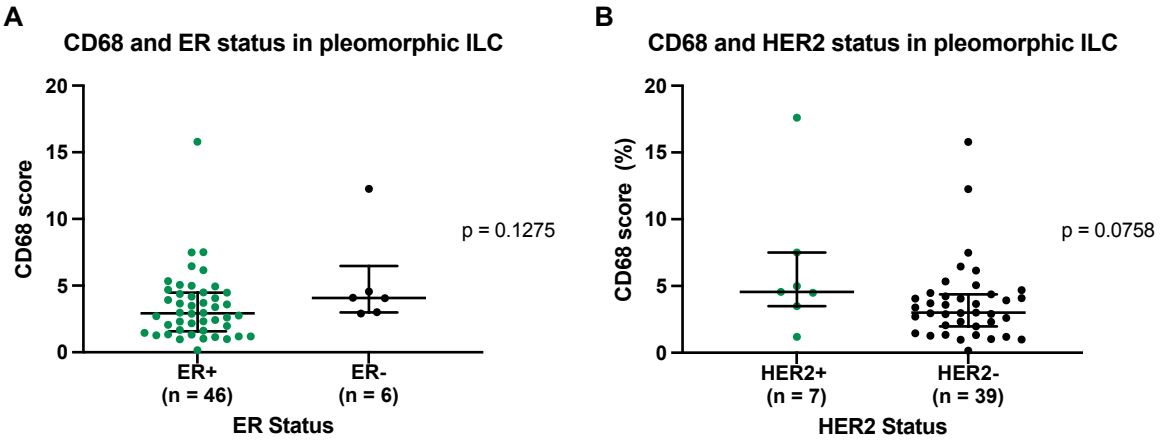


Figure 5.10: CD68 levels do not significantly differ according to hormone status in pleomorphic ILC: A) Scatterplot showing CD68 scores in ER+ (n = 46) vs ER- (n = 6) cases (p = 0.1275, Mann-Whitney U test) B) Scatterplot showing CD68 scores in HER2+ (n = 7) vs HER2- (n = 39) pleomorphic ILCs (p = 0.0758, Mann-Whitney U test).

5.37 CD68 levels are positively correlated with stromal TILs in ILC

Previous analysis identified higher levels of both stromal TILs and macrophages in pleomorphic ILC. These immune subpopulations are distinct and morphologically different with lymphocytes being derived from the lymphoid cell-lineage, appearing as small cells (8 -10 microns) which possess a large nucleus containing dense hetero-chromatin and minimal cytoplasm containing few mitochondria and ribosomes [315]. Macrophages are derived from the myeloid lineage and are round to oval in shape (10 – 30 microns) with an eccentrically located, oval / indented nucleus and much more abundant cytoplasm, which may appear as 'foamy' due to the presence of numerous secondary lysosomes [330, 331].

To better understand any association between the presence of stromal TILs and macrophages, a Spearman's rank test was performed in the entire CD68 KHP cohort (n = 132). This showed a statistically significant weakly positive correlation between CD68 levels and stromal TILs in the cohort (Spearman's rank correlation coefficient (ρ) = 0.378, p = 7.9 x 10^{-6}, Figure 5.11A). Even when pleomorphic ILC alone was considered, patients with high TILs generally showed a high level of infiltration by macrophages whilst patients with low stromal TILs generally had low levels of macrophages (Figure 5.11B).

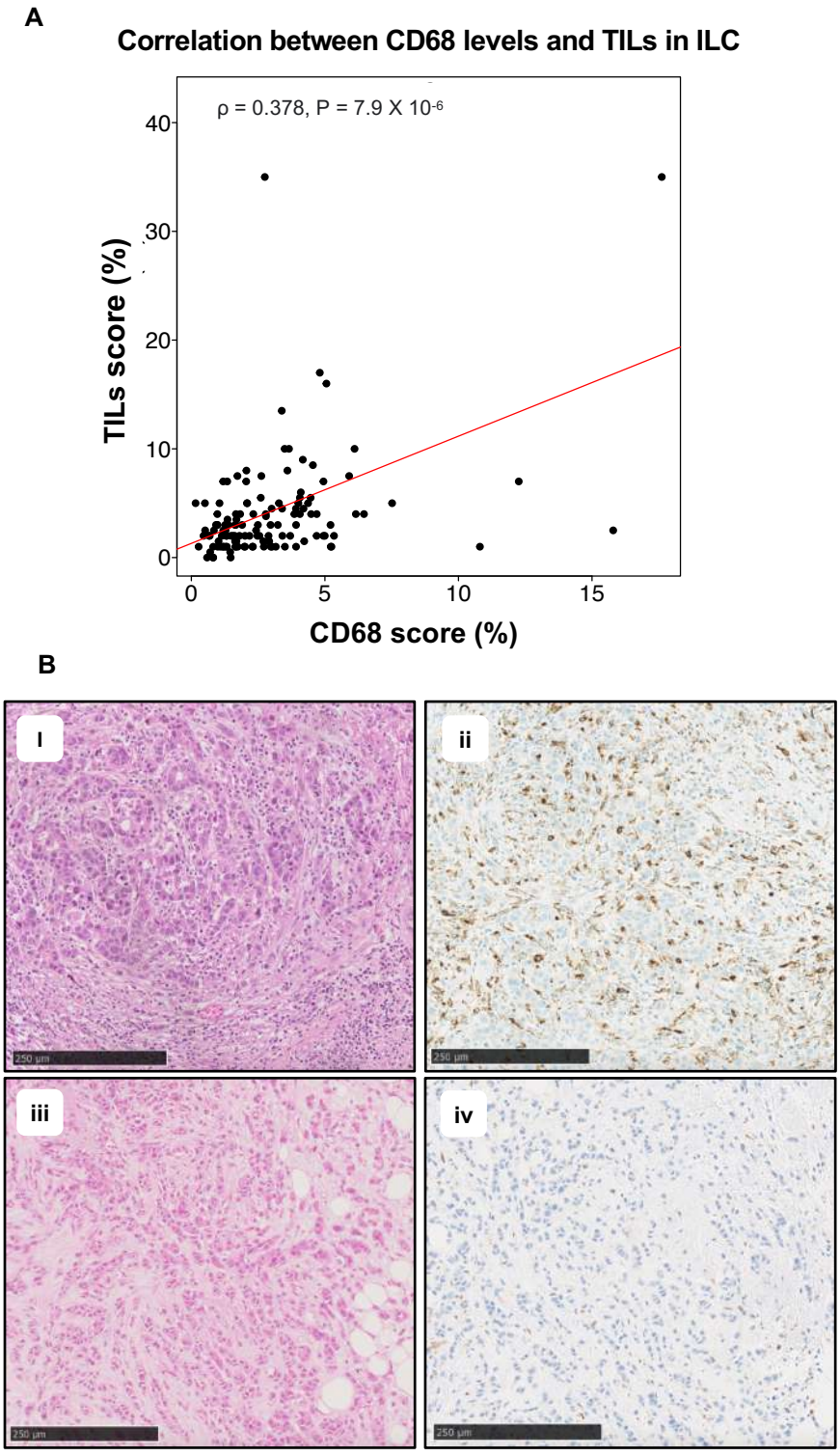


Figure 5.11: CD68 levels are positively correlated with stromal TILs in ILC: A) Scatterplot showing correlation between CD68 and stromal TILs levels in entire KHP lobular cohort (n =132, Spearman's rank correlation coefficient (ρ) = 0.378, p = 7.9 x 10^{-6}) B) I) H&E of pleomorphic ILC case TG011725 with high stromal TILs (TILs score: 35%) and ii) high levels of macrophages (CD68 score = 17.6) iii) H&E of pleomorphic case 17051250 with low stromal TILs (1%) and iv) minimal CD68 infiltrate (CD68 score = 0.986).

5.38 High CD68 levels are not associated with poor clinical outcomes in the ILC validation cohort

Univariable survival analysis using a cox-proportional hazards model was performed on the pleomorphic and non-pleomorphic ILC KHP validation cohorts to identify any association between CD68 scores and MFS/OS (n = 132). Seven patients with *de novo* metastases were removed prior to survival analysis and the follow-up time truncated at 15 years. CD68 scores were split into tertiles and clinical covariates e.g. tumour stage, lymph node status and hormone status were adjusted for. Overall there was no significant association between CD68 levels and MFS in both pleomorphic ILC (n = 52, 34 - 66%: p = 0.473, HR: 0.64, CI: 0.19 - 2.17, > 66%: p = 0.341, HR: 0.58, CI: 0.19 - 1.77, Figure 5.12A) and non-pleomorphic ILC (n = 73, 34 - 66%: p = 0.358, HR: 1.38, CI: 0.69 - 2.74, > 66%: p = 0.812, HR: 0.9, CI: 0.36 - 2.21, Figure 5.12B). There was also no significant association between CD68 levels and OS in pleomorphic ILC (n = 54, 34 - 66%: p = 0.027, HR: 0.27, CI: 0.08 - 0.86, > 66%: p = 0.107, HR: 0.46, CI: 0.18 - 1.18, Figure 5.12C) and non-pleomorphic ILC (n = 78, 34 - 66%: p = 0.881, HR: 1.05, CI: 0.57 - 1.93, > 66%: p = 0.282, HR: 1.46, CI: 0.73 - 2.91, Figure 5.12D).

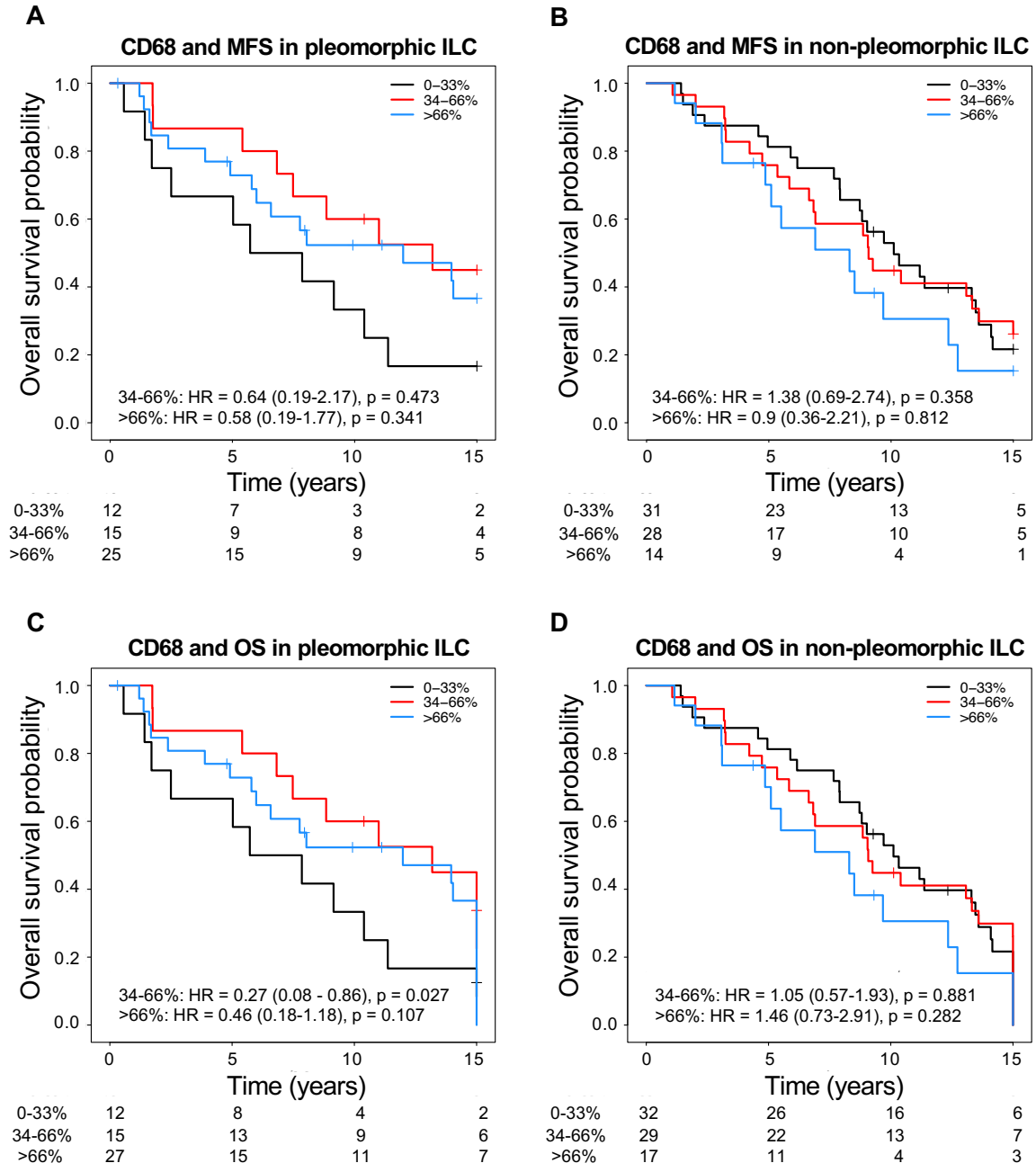


Figure 5.12: CD68 levels are not associated with survival outcomes in pleomorphic and non-pleomorphic ILC: A) KM graph showing the association between CD68 levels and metastasis free survival in pleomorphic ILC using tertiles (n = 52, 34 - 66%: p = 0.473, HR: 0.64, CI: 0.19 - 2.17, > 66%: p = 0.341, HR: 0.58, CI: 0.19 - 1.77) B) KM graph showing the association between CD68 levels and metastasis free survival in non-pleomorphic ILC (n = 73, 34 - 66%: p = 0.358, HR: 1.38, CI: 0.69 - 2.74, > 66%: p = 0.812, HR: 0.9, CI: 0.36 - 2.21) C) KM graph showing the association between CD68 levels and overall survival in pleomorphic ILC (n = 54, 34 - 66%: p = 0.027, HR: 0.27, CI: 0.08 - 0.86, > 66%: p = 0.107, HR: 0.46, CI: 0.18 - 1.18) D) KM graph showing the association between CD68 levels and overall survival in non-pleomorphic ILC (n = 78, 34 - 66%: p = 0.881, HR: 1.05, CI: 0.57 - 1.93, > 66%: p = 0.282, HR: 1.46, CI: 0.73 - 2.91).

5.39 High M2/M1 ratios are associated with relapse in pleomorphic ILC

Having identified that pleomorphic ILC is characterised by higher levels of macrophages compared to non-pleomorphic ILC, an assessment was next made to determine whether the ratio of M2-like to M1-like macrophages holds prognostic significance in pleomorphic ILC. Dual IHC staining for CD68 (green), a pan-macrophage marker (which stains both M1-like and M2-like macrophages) and CD163, an M2-like marker (brown) was performed on 18 relapsing and 17 non-relapsing pleomorphic ILCs. QuPath was used to quantify the level of staining for both markers (Figure 5.13A). The ratio of M2-like to M1-like macrophages in representative regions was calculated for each tumour (Table 5.8). Twenty-one samples (60%) showed higher levels of M1-like compared to M2-like macrophages and the remaining fourteen samples (40%) showed higher levels of M2-like compared to M1-like macrophages. Overall significantly higher M2/M1 ratios were present in relapsing compared to non-relapsing pleomorphic ILC patients ($p = 0.0449$, Mann-Whitney U test).

Univariable survival analysis using a cox-proportional hazards model was performed on 34 of the 35 patients for whom dual IHC was completed. One patient was excluded who presented with *de novo* metastatic disease. M2/M1 ratio was used as an independent variable to assess any association between M2/M1 ratio and MFS and M2/M1 ratios were split into tertiles. Clinical covariates such as tumour stage, lymph node status and hormone status were adjusted for. Overall, patients in the top tertile (67 - 100% of M2/M1 ratios) had significantly worse MFS compared to patients in the bottom tertile (0 - 33% of M2/M1 ratios) ($p = 0.007$, HR: 15.76, CI: 2.14 - 116.27, Figure 5.13B). The difference between the bottom and middle (34 - 66% of M2/M1 ratios) tertiles was not statistically significant ($p = 0.25$, HR: 2.78, CI: 0.49 - 15.82, Figure 5.13B).

Study ID	M1 %	M2 %	M2/M1	Time between diagnosis and relapse (years)
3117	2.63	0.93	0.35	3.03
9068	5.63	2.73	0.48	1.13
17013766	3.76	1.94	0.52	1.38
9894	2.95	2.16	0.73	0.00
17013509	2.45	1.84	0.75	9.20
17013387	3.17	2.39	0.75	5.32
1355G	4.57	3.56	0.78	12.92
3589	6.59	6.25	0.95	5.73
17023306	4.27	4.16	0.97	1.00
9350	6.61	6.82	1.03	1.94
17025981	2.93	3.49	1.19	1.07
2602 / 17024882	0.86	1.44	1.67	0.65
17058347	2.23	4.04	1.81	3.30
17056290	0.81	1.79	2.20	2.22
17058332	1.58	3.54	2.24	4.07
17014363	2.96	9.83	3.32	0.52
17051586	0.88	3.13	3.58	6.21
17013592	0.56	2.80	4.99	1.57
17055086	9.46	1.00	0.11	NA
17058082	1.92	0.47	0.24	NA
17006677	0.79	0.20	0.26	NA
17051250	0.28	0.09	0.31	NA
17041799	1.30	0.45	0.35	NA
17037739	4.88	2.05	0.42	NA
17007630	3.83	1.67	0.44	NA
17011975	1.70	0.99	0.58	NA
17057784	1.20	0.78	0.65	NA
17025943	2.44	1.84	0.76	NA
17004052	2.13	1.76	0.83	NA
17023387	12.24	11.08	0.91	NA
17022542	0.30	0.42	1.42	NA
17057789	3.26	4.77	1.46	NA
17056904	1.30	2.34	1.80	NA
17020041	4.31	9.88	2.29	NA
17054621	0.68	1.70	2.49	NA

Table 5.8: M1 and M2-like macrophage scores and M2/M1 ratios in 35 pleomorphic ILCs: The % scores represent the % of green (CD68) and brown (CD163) staining for M1-like and M2-like macrophages respectively. NA refers to patients without relapse

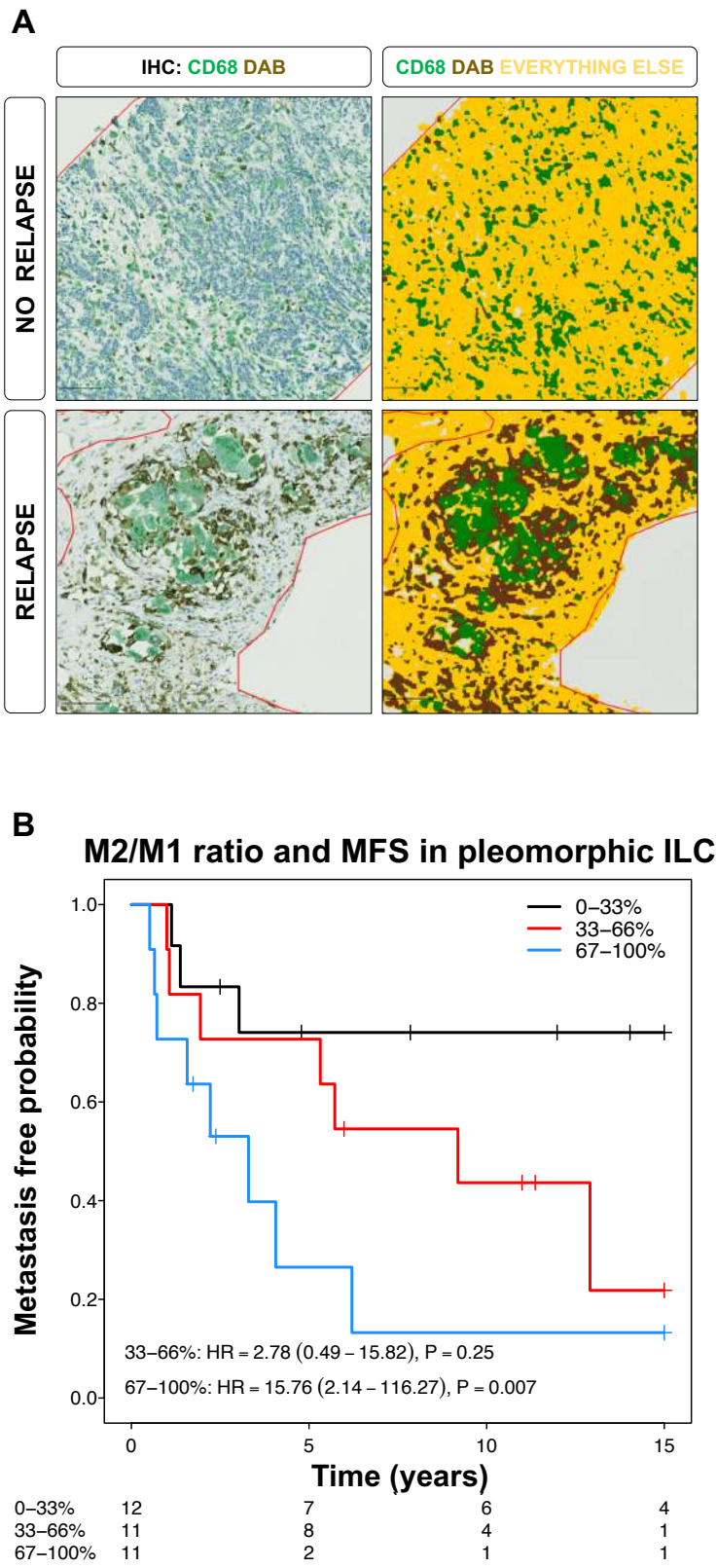


Figure 5.13: High M2/M1 ratios are associated with relapse in pleomorphic ILC: A) Top panel: representative no-relapse case 17055086 showing CD68 staining but almost absence of CD163 staining on IHC (left) and corresponding QuPath image. Bottom panel: representative relapse case 17058347 showing CD68 staining and a higher proportion of CD163 staining on IHC (left) and corresponding QuPath image. B) KM graph showing association between M2/M1 ratio scores and MFS in pleomorphic ILC (n = 34, 34 – 66%: p = 0.25, HR: 2.78, CI: 0.49 - 15.82, > 66%: p = 0.007, HR: 15.76, CI: 2.14 - 116.27).

5.40 ILC spheroids show higher growth rates in the presence of M2-like vs M1-like macrophages

Having identified that the presence of relatively higher levels of M2-like macrophages compared to M1-like macrophages is associated with disease recurrence in pleomorphic ILC, a 3D lobular co-culture experiment was next performed to further understand the growth rates of ILC cells using the cell line, MDA-MB-134, in the presence of M1-like and M2-like polarised macrophages when grown as spheroids. MDA-MB-134 cells are ER+/PR+/HER2-, characterised by the loss of E-cadherin and derived from a pleural effusion. They are commonly used as an ILC cell line [431].

Firstly, THP-1 cells, a human leukaemia monocytic cell line, were cultured in suspension in RPMI+10% FBS. MDA-MB-134 cells which had already been tagged with red fluorescent protein (RFP) and luciferase (Luc2) (MM134-RFP-Luc2) [356]. Cells were polarised into M1-like and M2-like macrophages by the addition of various cytokines and then grown as spheroids with MM134-RFP-Luc2 cells (see methods).

Firstly, when total spheroid size was considered, the spheroids containing M2-like macrophages were significantly larger than those containing M1-like macrophages ($p = 0.0344$, 2-way ANOVA test, Figure 5.14A, B). Next to focus on the tumour cells alone and ensure that the difference in spheroid size was not attributed solely to an increased growth of M2-like compared to M1-like macrophages, the intensity of red staining, reflecting the RFP-labelled tumour cells, was measured over time. This showed a significantly higher level of red intensity in the spheroids grown with M2-like compared to M1-like macrophages ($p = 0.001$, 2-way ANOVA test, Figure 5.14C), suggesting that the M2-like macrophage promote faster growth of the tumour cells.

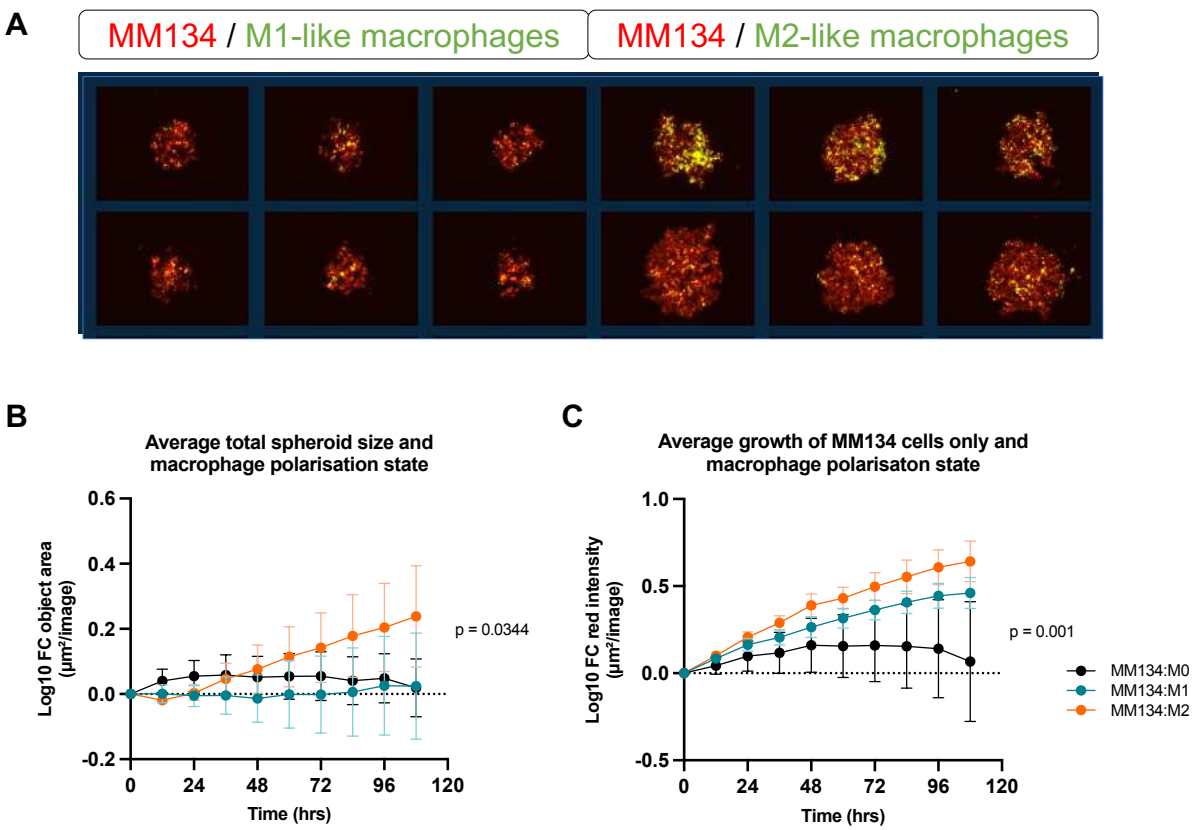


Figure 5.14: ILC spheroids show higher growth rates in the presence of M2-like vs M1-like macrophages: A: Representative Incucyte® images showing tumour cell (red) / macrophage (green) spheroids grown in the presence of M1-like (left) and M2-like macrophages (right) (MM134 = MM134-RFP-Luc2 cells) B) Growth curve summarising the pooled average of all experiments (n = 3) showing significantly higher spheroid size when MM134-RFP-Luc2 cells are grown with M2-like vs M1-like macrophages (p = 0.0344, 2-way ANOVA test) C) Growth curve showing pooled average of all experiments (n = 3) showing significantly higher red intensity, reflecting tumour cell specific growth, when MM134-RFP-Luc2 cells are grown with M2-like vs M1-like macrophages (p = 0.001, 2-way ANOVA test).

5.41 Immune-hot and immune-cold tumour cells differ at the transcriptomic level in pleomorphic ILC

Having studied the immune landscape with spatial context at the protein level using NanoString DSP, a whole transcriptome assessment was next made of the immune and tumour landscape with spatial context in pleomorphic ILCs (n =10) with the highest immune infiltrate (> 4% stromal TILs, Table 5.9). Slides were prepared with up to three tumours orientated on a single slide (Figure 5.15). Following slide preparation the unstained slides were subject to the NanoString whole transcriptome (WTA) workflow (Figure 5.16). During ROI selection 'Immune-hot' tumour regions, characterised by the presence of abundant immune cells and 'immune-cold' tumour regions, characterised by an absence of immune cells, were selected for each tumour. These regions were selected to also include cancer-associated fibroblasts (CAFs). Six of the 10 cases had previously been characterised at the protein level using NanoString DSP. The morphology markers PanCK, CD45 and *α-SMA* were used to identify tumours cells, immune cells, and CAFs respectively (Figure 5.17). A total of 76 ROIs were selected in 10 tumours across the 4 slides and each ROI was segmented into the separate PanCK+, *α*-SMA+ and CD45+ (if present) compartments producing distinct areas of interest (AOIs) (Figure 5.17). These distinct compartments/AOIs were then sequenced separately within each ROI producing distinct tumour cell, CAF, and immune cell transcriptomic data within each ROI. In total 182 distinct AOIs were sequenced.

	Study ID	ER	PR	HER2	Stromal TILs (%)	Lymph nodes (n)	Size (mm)	Grade	Age at diagnosis (years)	Onset of relapse
SLIDE 1	17054621	1	1	0	4	1	32	3	51.7	2
	17018625	1	1	0	7	8	50	2	47.3	3
SLIDE 2	1355	1	0	0	7	2	22	2	62.5	2
	17004052	1	1	0	4	0	20	2	80.8	3
	17013592	1	1	0	5	6	40	3	57.2	1
SLIDE 3	17011789	1	1	0	10	0	40	3	47.9	3
	17058145	1	1	0	4	3	60	3	59.6	2
SLIDE 4	3117	1	1	0	5.5	1	17	2	78.0	1
	17015028	1	1	0	8	0	60	3	79.7	3
	17047324	1	1	0	16	33	40	2	66.6	1

Table 5.9: Clinical details and outcome data for pleomorphic ILC NanoString WTA cohort: For ER, PR and HER2 0 = negative, 1 = positive. For onset of relapse: 1 = early relapse < years, 2 = late relapse > 6 years, 3 = no relapse

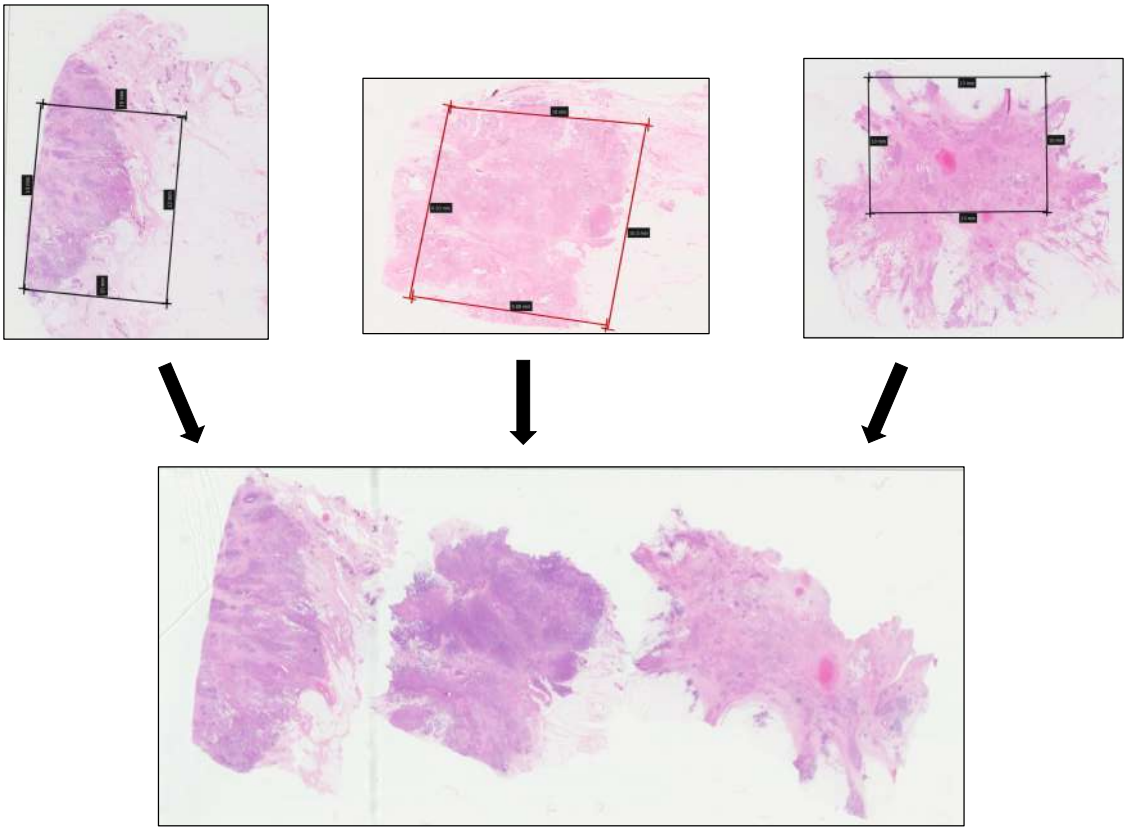


Figure 5.15: Selection of 10 x 13mm tumour regions from pleomorphic ILCs orientated on NanoString Whole Transcriptome (WTA) slides: Representative case showing three 10 x 13mm areas selected from individual tumour cases (left to right: 1355, 17004052, 17013592) for slide 2 to capture the most immunogenic areas. Multiple selected tumour areas were then orientated on each NanoString WTA slide.

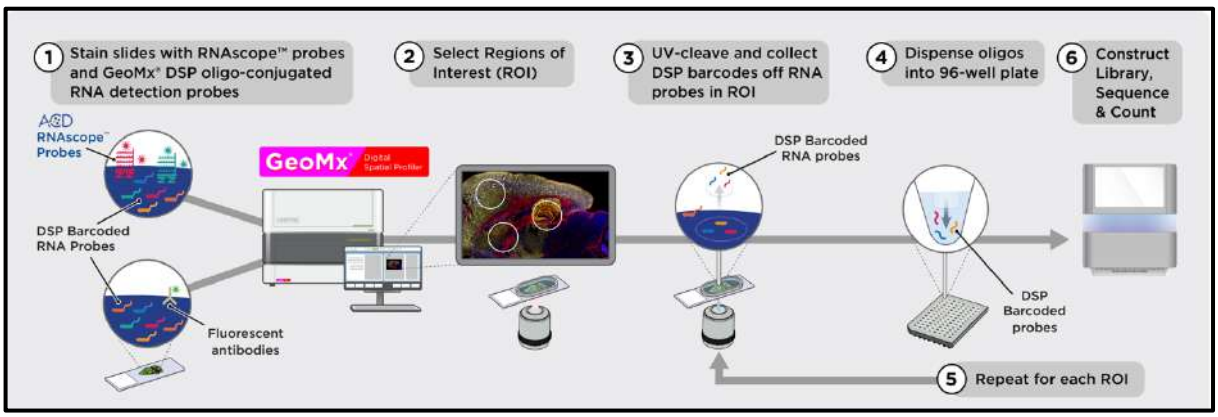


Figure 5.16: NanoString WTA workflow: Diagrammatic representation of the GeoMx WTA workflow summarising 6 keys stages. Taken from [432].

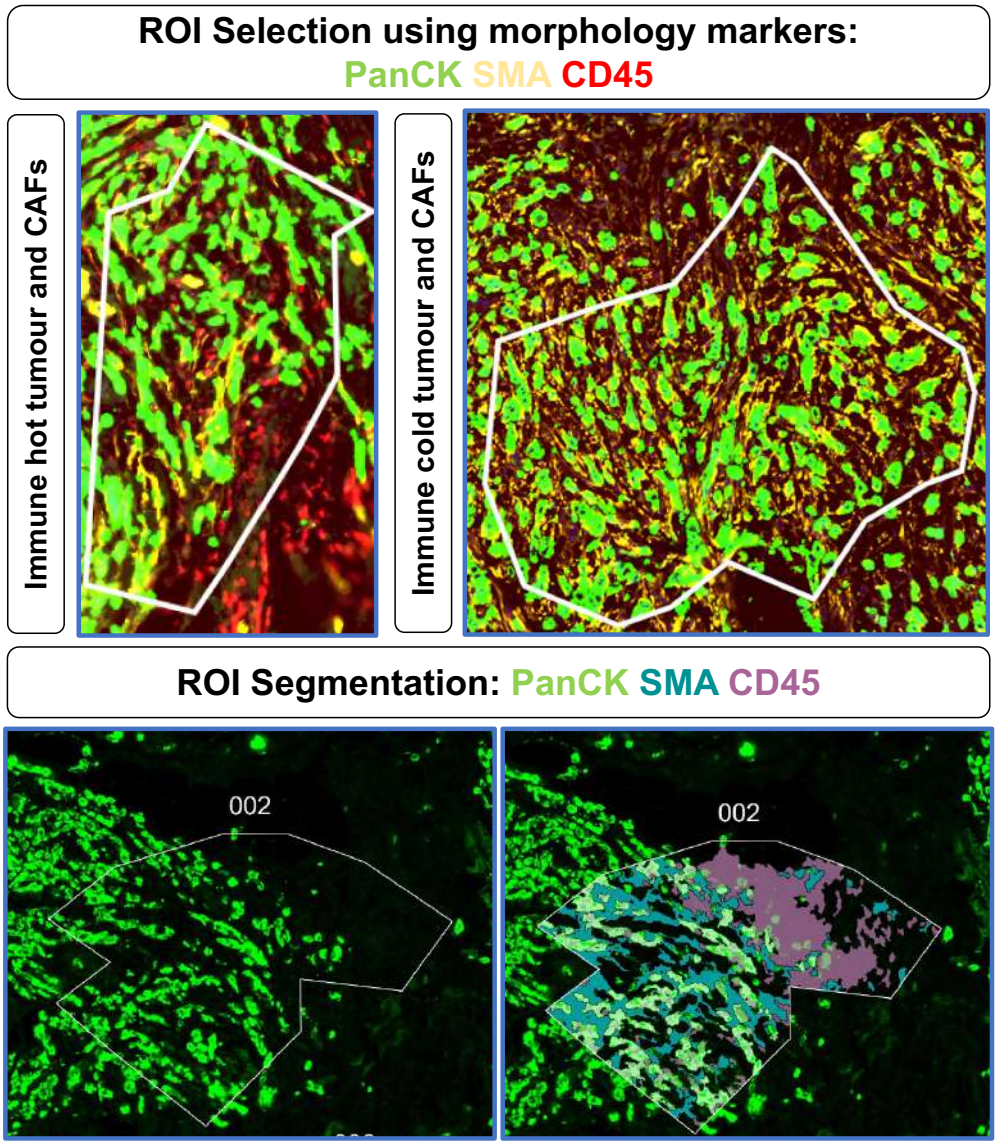


Figure 5.17: Selection and segmentation of immune-hot and immune-cold tumour regions of interest (ROIs): Top panel shows two distinct ROIs; left is an 'immune-hot' ROI characterised by the presence of CD45+ immune cells in addition to PanCK+ tumour cells and *α-SMA* CAFs and right showing an 'immune-cold' ROI characterized by the presence of PanCK+ tumour cells (green) and *α-SMA* CAFs. Bottom panel shows the segmentation of an individual 'immune-hot' ROI into distinct PanCK+ tumour, α-SMA+ CAF and CD45+ immune cell compartments known as 'areas of interest' (AOIs).

Comparing the transcriptomic profiles between immune-hot and immune-cold tumour cells identified 20 significantly differentially expressed genes (log_2FC+/- (1.25) and un-adjusted p < 0.001, Table 5.10).

	Gene	Coefficient	p - value	adj.p-val
↑ expression in immune-hot PanCK+ cells	GATA3	0.325	0.0004976	0.201
	GLUL	0.395	0.0002386	0.156
	IER3	0.473	0.0007120	0.205
	MRPS30	0.639	0.0001415	0.146
↑ expression in immune-cold PanCK+ cells	OR2T12	-0.391	0.0000009	0.011
	OR4S1	-0.346	0.0001653	0.146
	TBX18	-0.487	0.0001676	0.146
	CLCNKB	-0.390	0.0006174	0.201
	HPSE2	-0.384	0.0005953	0.201
	C18orf63	-0.425	0.0005324	0.201
	P2RX7	-0.493	0.0001109	0.137
	B3GALT2	-0.365	0.0000209	0.096
	LIN7A	-0.355	0.0009741	0.218
	KLK2	-0.425	0.0002209	0.156
	GDF11	-0.375	0.0006740	0.205
	HOXB13	-0.414	0.0009616	0.218
	NINJ2	-0.330	0.0000943	0.137
	SMCO4	-0.388	0.0002157	0.156
	ZNF530	-0.346	0.0000886	0.137
	WDR86	-0.379	0.0007747	0.205

Table 5.10: Twenty differentially expressed genes in immune-hot vs immune-cold tumour (PanCK+) cells.

Pathway enrichment (Enrichr [376]) identified pathways associated with interferon signalling and type-2 helper T lymphocytes in BioCarta 2016 (Table 5.11).

Pathway	Adjusted p-value
Chaperones modulate interferon signaling pathway	0.0188
GATA3 participate in activating the Th2 cytokine genes expression	0.0188

Table 5.11: Pathways associated with 20 genes differentially expressed in immune-hot vs immune-cold tumour (PanCK+) cells

5.42 *HOXB13* is associated with poor prognosis in ILC

Next, to assess the potential prognostic significance of the differentially expressed genes, their expression was assessed in well annotated ILC cohorts, METABRIC (n = 147) and SCAN-B (n = 386) and correlated with clinical outcome (excluding TNBC and HER2+ ILCs). Differential gene expression analysis was performed in METABRIC and SCAN-B to identify differentially expressed genes in the comparison groups; i) relapse vs no-relapse and ii) early (< 3 years) vs late (> 6 years) relapse. The prognostic significance of the 20 genes of interest identified from the NanoString WTA cohort were then assessed in METABRIC and SCAN-B. If the genes showed a $\log_2$FC+/-(1.25) and un-adjusted $p < 0.05$ in at least one of the two comparison groups in METABRIC or SCAN-B they were considered to be prognostically relevant.

This analysis identified that *HOXB13* which was expressed at significantly higher levels in immune-cold tumour cells in the NanoString WTA cohort, was associated with both relapse in the relapse vs no relapse comparison (LFC: 0.326, p = 0.007) and with early relapse in the early vs late relapse comparison (LFC: 0.996, p = 0.006) in SCAN-B (Table 5.12). Aside from *HOXB13*, of the 20 genes significantly differentially expressed in immune-hot and immune-cold tumour cells in the NanoString WTA cohort, there were no other prognostic genes identified in METABRIC and SCAN-B.

NanoString WTA cohort				SCAN-B relapse vs no relapse			SCAN-B early vs late relapse		
Gene	Coefficient	P.Value	adj.P.Val	logFC	P.Value	adj.P.Val	logFC	P.Value	adj.P.Val
OR2T12	-0.391	0.0000009	0.011	NA	NA	NA	NA	NA	NA
OR4S1	-0.346	0.0001653	0.146	NA	NA	NA	NA	NA	NA
TBX18	-0.487	0.0001676	0.146	-0.196	0.015	0.267	NA	NA	NA
CLCNKB	-0.390	0.0006174	0.201	NA	NA	NA	NA	NA	NA
HPSE2	-0.384	0.0005953	0.201	-0.301	0.001	0.144	-0.206	0.200	0.929
C18orf63	-0.425	0.0005324	0.201	NA	NA	NA	NA	NA	NA
P2RX7	-0.493	0.0001109	0.137	0.074	0.205	0.647	NA	NA	NA
B3GALT2	-0.365	0.0000209	0.096	-0.028	0.144	0.570	NA	NA	NA
LIN7A	-0.355	0.0009741	0.218	-0.229	0.138	0.565	-0.424	0.242	0.929
KLK2	-0.425	0.0002209	0.156	NA	NA	NA	0.242	0.344	0.929
GDF11	-0.375	0.0006740	0.205	NA	NA	NA	0.217	0.262	0.929
HOXB13	-0.414	0.0009616	0.218	0.326	0.007	0.218	0.996	0.006	0.757
NINJ2	-0.330	0.0000943	0.137	NA	NA	NA	NA	NA	NA
SMCO4	-0.388	0.0002157	0.156	NA	NA	NA	NA	NA	NA
ZNF530	-0.346	0.0000886	0.137	NA	NA	NA	-0.158	0.102	0.908
WDR86	-0.379	0.0007747	0.205	-0.264	0.013	0.263	NA	NA	NA
GATA3	0.325	0.0004976	0.201	NA	NA	NA	NA	NA	NA
GLUL	0.395	0.0002386	0.156	NA	NA	NA	NA	NA	NA
IER3	0.473	0.0007120	0.205	-0.236	0.123	0.543	NA	NA	NA
MRPS30	0.639	0.0001415	0.146	NA	NA	NA	NA	NA	NA

Table 5.12: Prognostic role of genes differentially expressed in immune-hot and immune-cold tumour regions assessed in SCAN-B: The table summarises the assessment of the prognostic significance of the 20 genes which were differentially expressed in immune-hot vs immune-cold tumour cells in the NanoString WTA cohort, in SCAN-B (n = 386). This identifies *HOXB13* as associated with relapse (LFC: 0.326, p = 0.007) and early relapse (LFC: 0.996, p = 0.006) in the separate comparison groups highlighted in red. NA = gene not detected due to low levels.

5.43 Immune-hot and immune-cold cancer-associated fibroblasts differ at the transcriptomic level in pleomorphic ILC

ILC has been shown to have a distinct tumour microenvironment characterised by a more pronounced growth of CAFs compared to IC-NST [311]. CAFs can represent a large component of the tumour stroma and have essential roles in facilitating crosstalk between tumour cells and other elements of the TME including immune cells [307]. To study CAFs in pleomorphic ILC at the transcriptomic level, and differences between CAFs that are closely associated with immune cells (immune-hot) and those which are devoid of immune cells (immune-cold), differential gene expression analysis was performed in the *α-SMA*+ cells in immune-hot vs immune-cold ROIs, adjusting for patient effect. This identified 13 significantly differentially expressed genes between the immune-hot and immune-cold CAFs, using $\log_2$FC+/-(1.25) and un-adjusted p < 0.001 (Table 5.13). Pathway analysis identified that immune related pathways such as interferon and TNF-*α* signalling, IL-6/JAK/STAT3 signalling, and the inflammatory response were enriched in immune-hot CAFs (Table 5.14).

	Gene	Coefficient	p-value	adj p-val
↑ expression in immune-hold SMA+ cells	**A2M**	0.654	0.0006694	0.613
	ICAM1	0.526	0.0003061	0.613
	TAP1	0.544	0.0001780	0.613
	CD4	0.444	0.0006737	0.613
	CXCL9	0.976	0.0000155	0.192
	B2M	0.504	0.0003740	0.613
	TGM2	0.591	0.0000802	0.497
	GOLT1B	0.412	0.0009577	0.613
	HEG1	0.476	0.0003719	0.613
	MPEG1	0.443	0.0004787	0.613
	HLA-DRA	0.657	0.0007907	0.613
↑ expression in immune-cold SMA+ cells	**KAZALD1**	-0.447	0.0007889	0.613
	KRTAP10-10	-0.346	0.0007389	0.613

Table 5.13: Thirteen differentially expressed genes in immune-hot vs immune-cold CAFs (*α*-SMA+ cells)

Pathway	Adjusted p-value
Interferon Gamma Response	0.0001
IL-6/JAK/STAT3 Signaling	0.0070
Interferon Alpha Response	0.0070
UV Response Up	0.0146
TNF-alpha Signaling via NF-kB	0.0165
Inflammatory Response	0.0165

Table 5.14: Pathways enriched in immune-hot CAFs (*α*-SMA+ cells)

5.44 *CXCL9* and *GOLT1B* are associated with poor prognosis in ILC

CXCL9, which showed higher expression in immune-hot CAFs, was associated with disease relapse in the relapse vs no-relapse comparison in SCAN-B (LFC: 0.497, p = 0.034, Table 5.15). *GOLT1B* was also more highly expressed in the immune-hot CAFs and showed a significant association with early disease relapse in SCAN-B (LFC: 0.407 p = 0.001, Table 5.15).

NanoString WTA cohort				SCAN-B relapse vs no relapse			SCAN-B early vs late relapse		
Gene	**Coefficient**	**P.Value**	**adj.P.Val**	**logFC**	**P.Value**	**adj.P.Val**	**logFC**	**P.Value**	**adj.P.Val**
A2M	0.654	0.000669	0.613	-0.201	0.058	0.428	NA	NA	NA
ICAM1	0.526	0.000306	0.613	-0.062	0.441	1.000	NA	NA	NA
TAP1	0.544	0.000178	0.613	0.318	0.005	0.200	-0.338	0.214	1.000
CD4	0.444	0.000674	0.613	NA	NA	NA	NA	NA	NA
CXCL9	0.976	0.000015	0.192	0.497	0.034	0.355	NA	NA	NA
B2M	0.504	0.000374	0.613	0.285	0.003	0.171	NA	NA	NA
TGM2	0.591	0.000080	0.497	NA	NA	NA	0.304	0.066	0.889
GOLT1B	0.412	0.000958	0.613	0.058	0.244	0.687	0.407	0.001	0.617
HEG1	0.476	0.000372	0.613	-0.113	0.197	0.638	-0.061	0.007	0.770
MPEG1	0.443	0.000479	0.613	NA	NA	NA	0.257	0.403	0.950
HLA-DRA	0.657	0.000791	0.613	-0.163	0.313	0.788	NA	NA	NA

Table 5.15: Prognostic role of genes differentially expressed in immune-hot and immune-cold CAFs assessed in SCAN-B: The table summarises the assessment of the prognostic significance of the 13 genes which were differentially expressed in immune-hot vs immune-cold CAFs from the NanoString WTA cohort, in SCAN-B (n = 386). This identified *CXCL9* as associated with relapse (LFC: 0.497, p = 0.034) and *GOLT1B* as associated with early relapse (LFC: 0.407, p = 0.001) in the separate comparison groups highlighted in red. NA = gene not detected due to low levels.

5.4 Discussion

Having previously identified in the KHP ILC cohort that the majority of ILCs are characterised by low stromal TILs and given the limited efficacy of T-centric immunotherapies in ER+ disease including ILC, there is a need for a change in strategy in targeting the immune system to improve clinical outcomes for ILC patients. To do this, there has been a fundamental need to better understand the nature of the immune landscape beyond stromal TILs in ILC and specifically in pleomorphic ILC, a rare, clinically aggressive, and understudied tumour type.

In this work, the use of NanoString DSP enabled the comprehensive characterisation of the immune microenvironment in a subset of histologically immunogenic pleomorphic ILCs at the protein level. This revealed the prognostic significance of two immune subpopulations, interestingly both of which are related to macrophages. Firstly, CD11c+ cells were significantly enriched in the tumours of patients who did not develop disease recurrence compared to relapsing patients. CD11c is a marker of multiple immune subpopulations including dendritic cells, M1-like macrophages, monocytes, and memory B cells [417, 418]. In further analysis within the relapsing patients, CD68, a pan-macrophage marker [433] which identifies both M1-like and M2-like macrophages was associated with more clinically aggressive tumours and early disease recurrence.

Consistent with the finding from NanoString DSP of higher CD68 levels in patients with early relapse, a previous study assessed the degree of macrophage infiltration in 100 breast cancers using CD68 IHC and correlated high levels of macrophages with poor prognostic features. These included larger tumour size and more advanced stage disease and high macrophage levels were also correlated with worse 5-year survival rates although it is unclear if the cohort included ILCs [434]. A further study assessed CD68, CD11c, and CD163 levels in a TMA of 367 breast cancers and correlated the level of all (CD68+), M1-like (CD11c+) and M2-like (CD163+) macrophages with a range of clinicopathological features and survival outcomes. They show that high levels of CD11c+ macrophages in the stroma were associated with significantly improved DFS as well as OS in breast cancer [435].

In addition to being expressed in macrophage subsets, CD11c is a dendritic cell marker, with the NanoString DSP findings suggesting that high levels of dendritic cells within the pleomorphic ILC tumour microenvironment help mediate an effective anti-tumour response preventing disease relapse. Dendritic cells are considered to have a key role in tumour-related immune reactions, acting as potent antigen-presenting cells. They recognise and process tumour antigens and present them to lymphocytes triggering a naïve T lymphocyte response,

and they are therefore critical in linking the innate and adaptive immune systems [436, 437]. Overall, given the limited size of the NanoString DSP cohort, further research and study of dendritic cells in larger ILC validation cohorts would be necessary to better understand any prognostic relevance of this immune cell population in ILC.

Interestingly the NanoString early vs late relapse comparison identified that a pan-macrophage marker, CD68, was expressed at significantly higher levels in early relapsing patient tumours. The assessment of CD68 levels in the wider KHP validation pleomorphic and non-pleomorphic cohorts firstly identified that pleomorphic ILC contains more macrophages than non-pleomorphic ILC which together with higher levels of stromal TILs, supports the notion that pleomorphic ILC is fundamentally a more immunogenic tumour than non-pleomorphic ILC, and this is after having controlled for differences in the frequency of ER- and HER2+ disease between the two ILC histological subtypes.

Whilst the total level of macrophages as assessed through CD68 IHC were not associated with clinical outcome in the KHP validation cohort, further analysis characterising the macrophages into M1-like and M2-like phenotypes, revealed that higher M2/M1 ratios were associated with significantly worse MFS in pleomorphic ILC, highlighting the relevance of the polarisation states of macrophages within the ILC tumour microenvironment and that the relative abundance of immunosuppressive M2-like compared to pro-inflammatory M1-like macrophages holds prognostic significance. A recent study assessing macrophages in ILC found that M1-like macrophages are more abundant than M2-like macrophages [340]. It is unclear if the cases in this study included rare ILC subtypes such as pleomorphic ILC, but the results from the pleomorphic KHP cohort showed that a significant proportion (40%) of cases had higher levels of M2-like than M1-like macrophages, suggesting that pleomorphic ILC could possibly contain a higher proportion of M2-like macrophages than non-pleomorphic ILC. Further dual IHC in the KHP non-pleomorphic cases will help address this question.

In vitro studies assessing spheroid growth of ILC cell line MDA-MB-134 in the presence of M1-like and M2-like macrophages showed a significantly higher growth rate of tumour cells in the presence of M2-like macrophages further supporting the concept that M1-like macrophages have tumour-suppressive properties, whilst M2-like macrophages have pro-tumoral properties.

NanoString Whole transcriptome analysis in a subset of pleomorphic cases identified differences in gene expression between immune-hot and immune-cold tumour cells. *HOXB13* which was more highly expressed in the tumour cells in regions devoid of immune cells

compared to tumour cells closely associated with immune cells, was associated with poor outcomes in SCAN-B. *HOXB13* (Homeobox protein Hox-B13) encodes a transcription factor which belongs to the homeobox gene family which are essential for embryonic development. It acts as part of a developmental regulatory system which provides cells with specific positional identities on the anterior-posterior axis [386]. A high *HOXB13* to *IL17BR* (Interleukin 17 Receptor B) expression ratio has been associated with an increased risk of disease relapse and poor outcomes in ER+ breast cancer patients [438, 439]. Furthermore, high levels expression of HOXB13 protein expression assessed with IHC were associated with resistance to tamoxifen therapy due to the suppression of ER and activation of the mTOR signalling pathway through IL-6 [440, 441].

Differential gene expression between immune-hot and immune-cold CAFs identified two genes; *CXCL9* and *GOLT1B*, which appeared to hold prognostic significance in SCAN-B with high levels being associated with worse outcomes. *CXCL9* (C-X-C Motif Chemokine Ligand 9) encodes a cytokine that affects the growth, movement, or activation state of cells that participate in the immune or inflammatory response [386]. It is thought to be involved in immune cell migration and activation within the tumour microenvironment [442] and therefore higher expression is associated with more immune cells consistent with being more highly expressed in immune-hot ROIs. *GOLT1B* (Golgi Transport 1B) encodes a protein which is involved in positive regulation of I-kappaB kinase/NF-kappaB signalling and is located in the endoplasmic reticulum [386]. It has previously been associated with tumour progression through regulation of the immune microenvironment in breast cancer [443].

Overall the work in this chapter has characterised and provided new insights into the complex immune microenvironment in ILC at the protein and transcriptomic levels with spatial context and assessed the prognostic significance of these findings.

Chapter 6: Mapping the subclonal heterogeneity of metastatic ILC

6.1 Background

A clone is defined as an asexually reproducing population which shares the same ancestry and may be defined at different levels, for example a group of tumour cells which share the same genetics [444]. A subclone defines a cell group that has diverged away from its original clonal identity through the acquisition of new mutations [444]. Intra-tumoral heterogeneity is the presence of a diverse range of subclones within any given tumour which have a breadth of unique molecular profiles. These subclones can be genetically diverse and can have different gene expression profiles, epigenetic regulations, cell differentiation states and tumour microenvironmental requirements [444, 445]. In addition to genetic diversity, these non-genetic factors are key to producing a highly complex and dynamic tumour cell population which is continuously developing through clonal evolution. Clonal evolution refers to continuous process of clonal expansion, genetic diversification and natural selection of subclones within an adaptive ecosystem [444]. Genetic changes and non-genetic factors which confer a selective advantage to a specific subclone may thus promote its dominance within the tumour cell population. Through successive rounds of tumour cell diversification and selection, enrichment of the fittest tumour subclone variants takes place resulting in a highly dynamic and adaptable tumour cell population which has the potential to drive disease progression and fuel therapy resistance [444, 445]. When a selective pressure is placed on a tumour through a therapeutic agent, distinct subclones with molecular profiles which confer a survival advantage evade cell inhibition or death and have the potential to drive progression. These cells may either inherently possess or acquire mechanisms of survival advantage through the process of clonal evolution [444, 445].

Temporal heterogeneity refers to molecular changes and differences between subclones over time [445]. Spatial heterogeneity refers to variation in the distribution of subclones within a given space or area in terms of their location and concentration [445]. As tumour subclones evolve at a molecular level over time, their spatial locations within the immune microenvironment with respect to various immune subpopulations as well stromal components such as cancer-associated fibroblasts may be key determining their survival potential.

Large scale genome analysis of primary breast cancers demonstrates that most mutations are found in only a small proportion of tumour cell subclones [446] and not evenly distributed spatially within individual tumours [447]. Metastatic breast cancers share most of their genomic alterations with the corresponding primary tumour, yet they may also exhibit additional mutations not detected in the primary, which may be subclonal in the primary [448-453]. Understanding the molecular profiles of these subclones may be important for developing new targeted treatments in ILC. Novel new approaches such as lineage tracing of tumour cells and single cell/nuclei DNA and RNA sequencing offer new opportunities to model this.

Laboratory based models are critical in enabling a better understanding of the unique biology of ILC. However the number of well-established and useful ILC tumour models in limited given the lower incidence of ILC compared to IC-NST, lower numbers of ILC patients included in clinical trials and also the intrinsic features of ILC which means they fail to easily transplant and grow as organoids [431]. Only a few ILC cell lines have been studied extensively [431]. ILC cell lines tend to grow at a significantly slower rate than their IC-NST counterparts, and few are derived from primary tumours, with the majority being derived from sites of metastatic disease e.g. ascites or pleural effusions [431]. Moreover, from a demographic perspective most are derived from Caucasian patients. Whilst clinically the majority of ILCs are of a luminal A phenotype (ER^+, PR^+, HER2+), few of the available ILC models express the oestrogen receptor. The loss of E-cadherin is a defining feature of ILC and consistent with this, the majority of the ILC cells lines harbour a pathogenic *CDH1* mutation [431]. However whilst *TP53* mutations are rare in ILC, the commonly used ILC cells lines show mutant *TP53* [431].

ILC cell lines can also be grown in 3D culture in the form of organoids, spheroids and organotypic cultures from tumour slices or co-cultures. Growth in these formats more accurately represents the intra-tumoral heterogeneity and mechanical, biochemical and spatial features of the tumour compared to growth in 2D format [431]. ILC tumour tissue from patients can also be established as a model of organoids, known as patient derived organoids (PDOs), which better preserve the original genomic features of the primary tumour [431]. A limitation of 2D and 3D-based models is that they lack the complex tumour-stromal interactions which are a key feature of ILC.

The generation of mouse models which accurately recapitulate the unique features of ILC and its metastatic spread has also been challenging. ILCs can be modelled using xenografts which involve tumour cells growing in immunodeficient mice and these models have helped to reveal important features of ILC disease relapse and mechanisms of treatment resistance [431]. There are four main routes of injection which have been used to generate xenografts which

include systemic (intracardiac/intracarotid injections), local (e.g. renal capsule, intra-bone marrow, intracranial), beneath the skin (subcutaneous/intradermic) and orthotopic (mammary fat pad or inside the milk ducts using the mouse intraductal model). Xenograft models include cell line derived xenografts and patient-derived xenografts (PDXs) whereby fresh primary or metastatic tumour tissue is directly implanted into an immunocompromised mouse. The advantage of PDXs is that they retain the intra-tumoral heterogeneity, genetics and polyclonality of the original patient tumour. The biggest challenge in the development of ILC PDXs is the significantly lower engraftment rate of ER+ tumours such as ILC compared to triple-negative and HER2+ breast cancers [431]. Furthermore given that ILCs are generally slow-growing, ILC xenografts are equally notoriously slow to grow. Currently there are few ILC PDXs that been reported.

The modelling of ER+ breast cancers *in vivo* has presented challenges due to its specific microenvironmental requirements. It has proven challenging to accurately recapitulate the nature of human ER+ breast cancers since cell line xenografts require non-physiological hormonal supplements which can be poorly tolerated [454] and PDXs are difficult to establish. Traditionally ER+ tumour cells have been grafted into the mouse mammary fat pads. However this has been shown to induce TGFβ/SLUG signalling and results in differentiation of the cells into a basal-like as opposed to luminal phenotype [455]. The mouse intra-ductal (MIND) model involves the grafting of tumour cells directly into the mouse milk ducts and has been shown to suppress SLUG, enabling the tumour cells to grow and develop initially *in situ* within the ducts under the influence of physiological hormones, retaining a luminal phenotype thus more closely recapitulating human disease [455]. Intraductal ER+ PDXs have been shown to be re-transplantable and genomically stable [455]. The MIND model therefore offers new opportunities for the modelling of ER+ breast cancers including ILC in the context of physiological hormone levels and initial *in situ* growth of tumour cells.

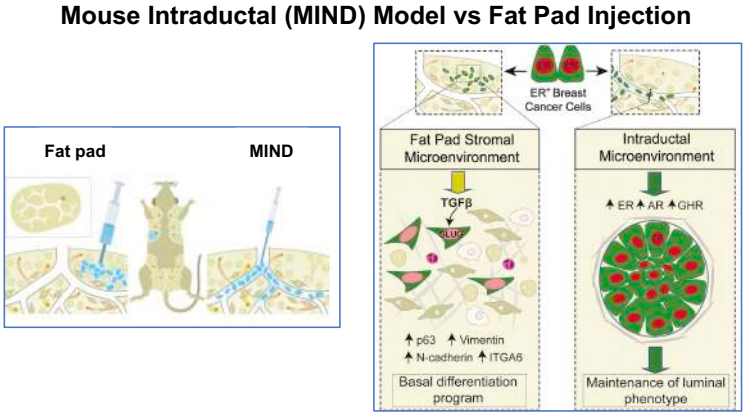


Figure 6.1: Mouse intraductal model: Schematic diagram comparing fat pad and intraductal microenvironments for the injection of ER+ breast cancer cells [455]

6.2 Chapter Aims

Overall the purpose of this chapter is to:

1) Determine if subpopulations of ILC tumour cells that are transcriptionally rewired to become metastatic pre-exist in the primary tumour.

2) Determine if these subclones hold prognostic/predictive significance in clinical cohorts of ILC.

6.3 Results

6.31 ILC MM134 brain metastatic lesions show an enrichment of cell motility and migration related pathways

To help determine if aggressive subclones of ILC tumour cells which are transcriptionally rewired to become metastatic pre-exist in the primary tumour, a single nuclei RNA sequencing experiment was performed using the ILC cell line MDA-MB-134. Tumour cells were RFP and luciferase labelled (MM134-RFP-Luc2) and 400,000 cells were injected intraductally using the MIND model into each of the 3rd and 4th mammary glands. Following a 12-month period, well-defined metastatic lesions had formed in the mice, at classical sites of ILC metastatic spread including the ovaries and brain (Figure 6.2). To study transcriptional diversity among the primary tumours and metastatic lesions and help identify gene expression pathways enriched in metastatic disease, 3 primary tumours (1 left primary and 2 right primaries), the brain and ovaries were extracted from one such mouse and single nuclei sequencing performed in the separate lesions (see methods, Figure 6.2, Table 6.1).

Sample	Initial number of cells	Human cells	After read filter	After mitochondrial filter	After doublet filter	% cells kept	Initial number of genes	Selected genes
Primary L	8315	8193	7716	7643	7317	0.88	25321	15612
Primary R1	7499	7367	7152	7077	6772	0.9	25772	16480
Primary R2	11633	11483	9955	9820	9194	0.79	27414	17249
Brain metastasis	7348	7204	6837	6178	5992	0.81	25248	16302
Ovary L	8065	1231	441	441	441	0.05	18601	17255
Ovary R	5706	793	220	220	220	0.04	17849	15846

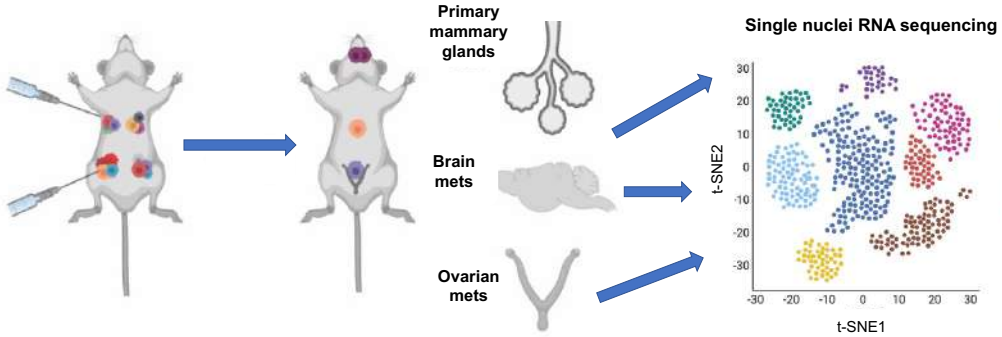
Table 6.1: Pre-processing of primary and metastatic samples: Table summarises the number of cells and genes at each stage of the filtering process during pre-processing. Primary L = left primary mammary gland, Primary R1 = right primary mammary gland 1, Primary R2 = right primary mammary gland, Ovary L = left ovary, Ovary R = right ovary. Note the number of human cells remaining post filtering in the ovarian metastasis samples were low, and thus not analysed further.

CIDER based clustering (see methods) identified 11 META-clusters (Figure 6.3, 6.4). When the proportion of contribution of the brain metastases sample to each CIDER meta-cluster was considered, meta-cluster 11 showed the highest contribution of the brain metastases (Figure 6.4). Next to gain further insights into the gene expression profiles of each individual CIDER meta-cluster, the markers of the individual meta-clusters were defined by the aggregation of markers of the constituent subclusters (see methods). Overall, within meta-cluster 11 there were 14 genes identified as the top markers of the meta-cluster, 5 of which were non-protein coding/RNA genes. Using these top markers of meta-cluster 11, Enrichr [376] was used to assess the gene expression pathways enriched in the meta-cluster using the 'Gene Ontology; biological processes' (GO-DB) database which identified an enrichment of pathways associated with cell motility and migration (Table 6.2).

Pathway	Adjusted p-value
Regulation of cellular component movement	0.0107
Cytoskeleton organisation	0.0162
Tissue development	0.0162
Cell migration	0.0162
Cell locomotion	0.0162
Regulation of intracellular signal transduction	0.0162

Table 6.2: Gene expression pathways enriched in CIDER meta-cluster 11

A



B

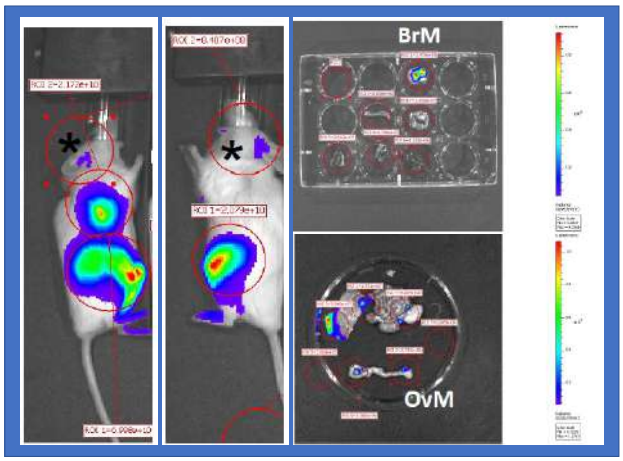


Figure 6.2: Design of MM134-RFP-Luc2 single nuclei RNA sequencing experiment: A) Schematic summarising the experiment workflow starting with intraductal injection of MM134-RFP-Luc2 cells using the MIND model followed by extraction of invasive tumour-containing primary glands and metastatic brain and ovarian lesions 12 months post initial injection. This is then followed by the generation of single cells from the tissue samples and single nuclei RNA sequencing of the pooled samples B) IVIS imaging showing positive signal in the ovaries, mammary glands and brain 12 months post-injection.

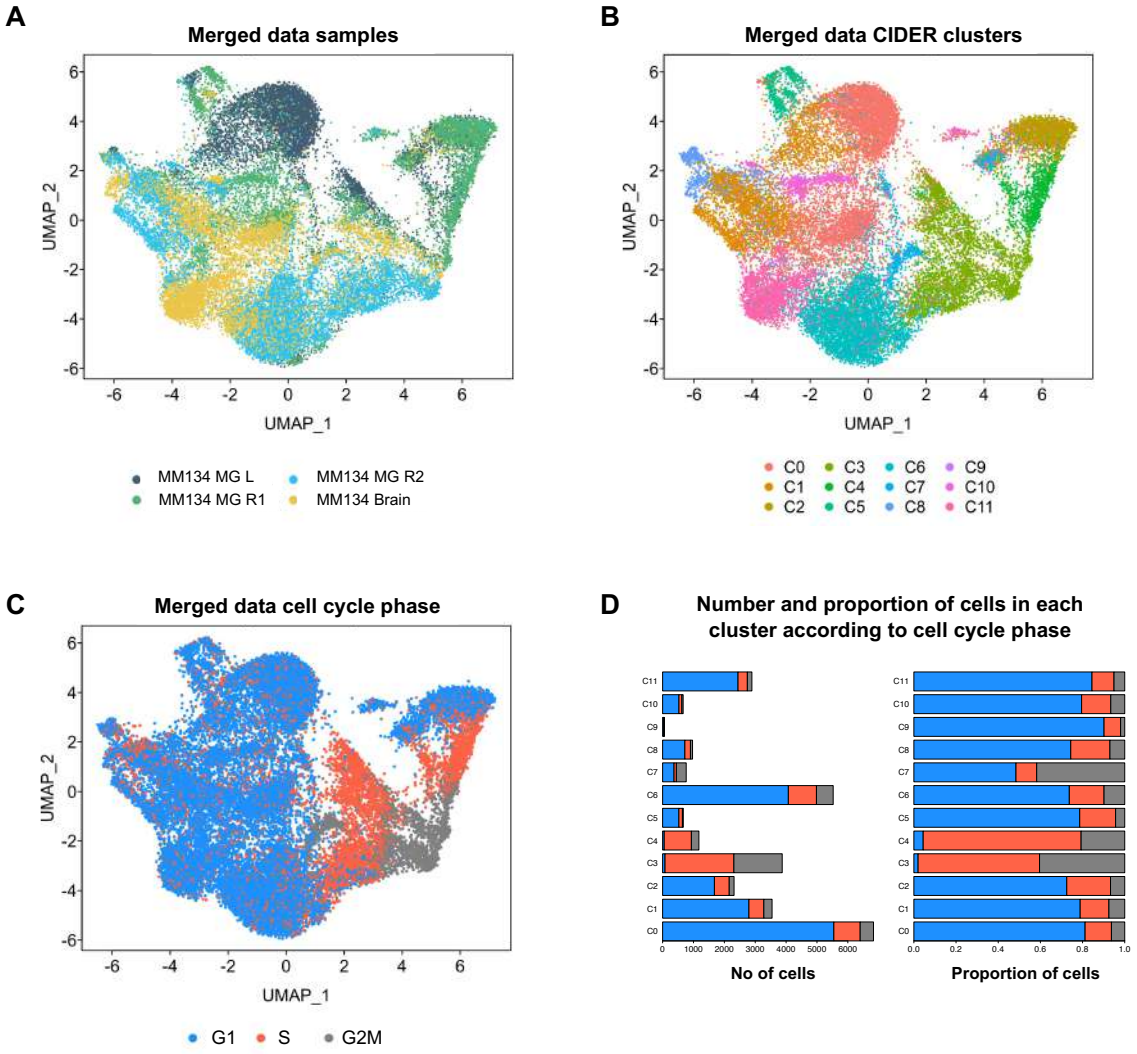


Figure 6.3: Single nuclei RNA sequencing of nuclei from 48,566 single cells: Cells were isolated from mouse mammary glands, ovaries and brain metastatic lesions following injection of MM134-RFP-Luc2 cells using the MIND model. Single nuclei uniform manifold approximation and projection (UMAP) plots for visualisation of all samples in one space (without batch effect) colour-coded by: A) individual tissue samples B) CIDER clusters C) cell cycle phase D) Bar chart showing the number of cells (left) and proportion of cells (right) falling into each cluster based upon cell cycle phase.

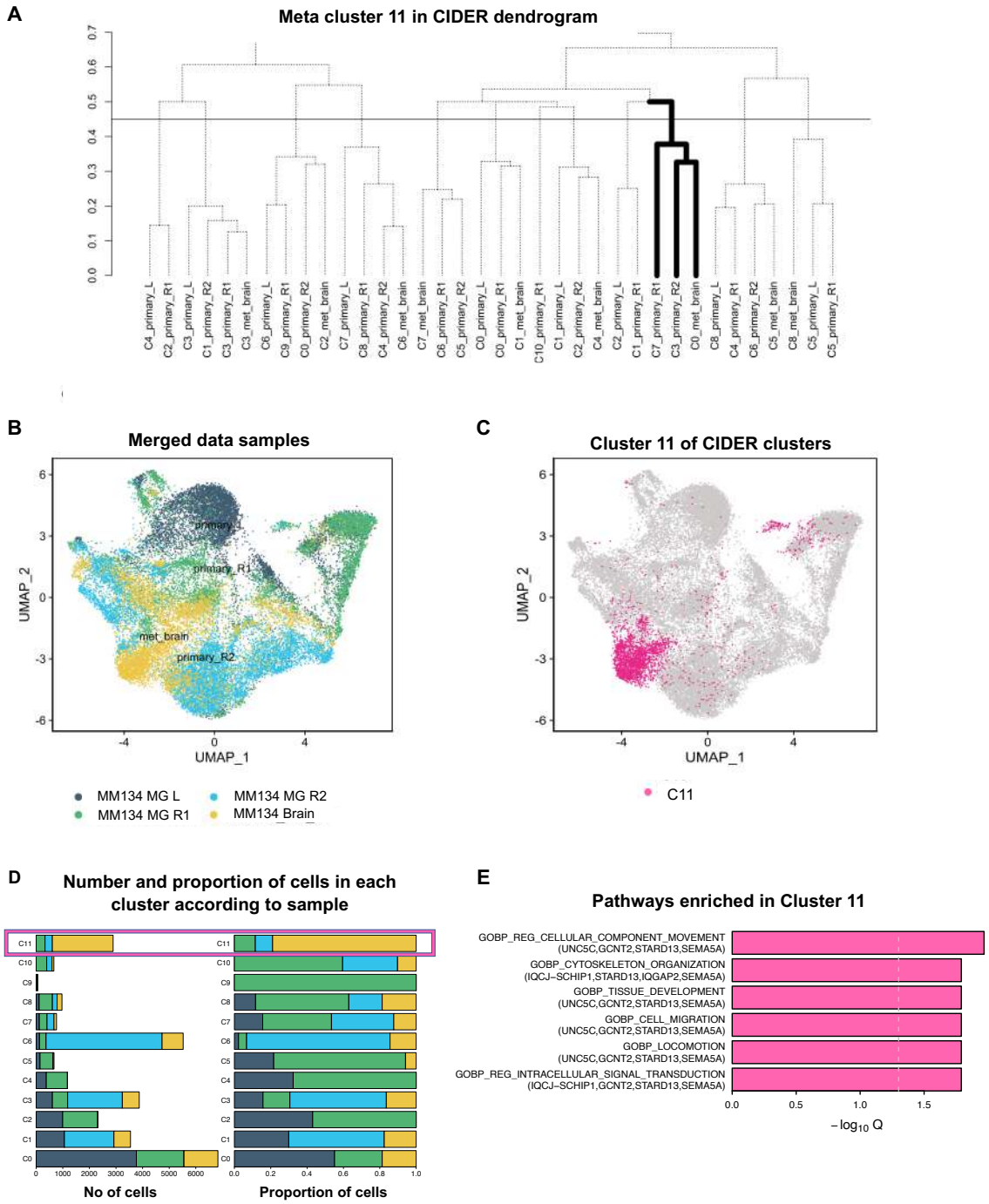


Figure 6.4: Cluster 11 shows a high contribution from metastatic brain lesions: A) Dendrogram highlighting in bold where cluster 11 is located, the composition of cluster 11, and the height of C11 in the in the dendrogram (the lower the height, the more robust the cluster is) B) UMAP (without batch effect) colour coded by individual tissue samples C) UMAP for CIDER clusters highlighting location of cells in cluster 11 only D) Bar chart showing the number of cells (left) and proportion of cells (right) falling into each cluster based upon each tissue sample with pink box indicating cluster 11 which is enriched in brain metastases E) Bar chart summarising pathways enriched in cluster 11.

6.32 *GCNT2* is associated with early relapse in ILC

Having identified the genes which were more highly expressed in meta-cluster 11 which was accounted for largely by metastatic brain tumour subclones, an assessment was next made of the prognostic significance of these genes in two well-annotated general ILC cohorts SCAN-B (n = 386) and METABRIC (n = 148) and the KHP pleomorphic ILC cohort (n = 47), using mRNA abundance data comparing differential gene expression in relapse vs no-relapse and early (< 3 years) vs late (> 6 years) relapse. This identified *GCNT2* as significantly associated with early relapse in METABRIC (LFC = 0.363, p = 0.029) and the KHP pleomorphic cohort (LFC = 3.086 p = 0.045) using statistical thresholds of $\log_2$FC (1.25) and un-adjusted p < 0.05. *IQGAP2* showed differing results in METABRIC and the KHP cohort, associated with early relapse in METABRIC (LFC = 0.505, p = 0.035) but with late relapse in the KHP pleomorphic cohort (LFC = -1.901, p = 0.031, Table 6.3).

	SCAN-B relapse vs no relapse		METABRIC relapse vs no relapse		KHP relapse vs no relapse		METABRIC early vs late relapse		SCAN-B early vs late relapse		KHP early vs late relapse	
Gene	**logFC**	**P.Value**	**logFC**	**P.Value**	**logFC**	**P.Value**	**logFC**	**P.Value**	**logFC**	**P.Value**	**logFC**	**P.Value**
UNC5C	NA	NA	-0.056	0.037	NA	NA	-0.045	0.232	NA	NA	NA	NA
IQCJ-SCHIP1	-0.179	0.007	-0.056	0.037	NA	NA	-0.045	0.232	NA	NA	NA	NA
AL139383.1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
GCNT2	NA	NA	-0.123	0.175	0.434	0.546	0.363	0.029	NA	NA	3.086	0.045
STARD13	-0.146	0.068	NA	NA	-0.002	0.998	-0.397	0.101	-0.244	0.351	-0.002	0.998
TPRG1	-0.412	0.059	NA	NA	-0.849	0.401	-0.338	0.238	-1.183	0.036	-0.849	0.401
IQGAP2	NA	NA	-0.181	0.162	-0.066	0.897	0.505	0.035	NA	NA	-1.901	0.031
BCAS1	NA	NA	0.206	0.176	-0.910	0.292	0.173	0.474	-0.361	0.335	-0.910	0.292
COLEC12	NA	NA	-0.155	0.408	-1.617	0.132	NA	NA	NA	NA	-1.617	0.132
SEMA5A	-0.262	0.005	NA	NA	-0.379	0.705	NA	NA	NA	NA	-0.379	0.705
MIR4713HG	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
LINC02613	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
LINC01948	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
LINC00882	-0.089	0.129	NA	NA	NA	NA	NA	NA	-0.133	0.403	NA	NA

Table 6.3: Assessment of the prognostic significance of 14 genes associated with metastatic lesions in SCAN-B, METABRIC and the KHP pleomorphic ILC cohort: (NA = not detected, red boxes indicate significant results).

6.33 Molecular barcoding is used to study the evolution of tumour metastases in ILC

The preliminary single nuclei sequencing data highlighted that subpopulations of cells already possess the transcriptional features that are enriched in metastasis (i.e. are 'transcriptionally primed'). However, from these profiling studies alone it not possible to determine if subclones are selected for in a stochastic manner. One way to address this question is to use lineage tracing (cellular barcoding) to assess if metastatic lesions harbour enrichment of the same cellular barcode when injected from the same starting barcoded population. If genetically defined subclones pre-exist and have a selective advantage, these should be enriched in multiple metastatic lesions. If this process is random, different barcodes should be observed in different metastatic lesions, which then maybe suggestive of epigenetic plasticity governing metastasis, whereby all cells have the potential to acquire the transcriptional alterations that enable them to spread.

To study this a barcoding experiment was designed (Figure 6.5), which involved the transduction of a starter population of MM134-RFP-Luc2 cells with Cellecta's CloneTracker™ Lentiviral Barcode Libraries [456], in order to genomically label each cell with a uniquely-identifiable short nucleotide sequence (barcode). This resulted in a starter population whereby each cell had a unique barcode. Given that the barcodes become genomically integrated, they are heritable and therefore the progeny of each starter population cell contain the same sequence enabling the tracing of the clonal expansion of each starting population cell. The barcoded cells were selectable using puromycin since the barcodes also contained a puromycin-resistance selection marker and a multiplicity of infection of 5% was used (Figure 6.4). Barcoded cells were further expanded and 400,000 cells injected using the MIND method into the 3rd and 4th right and left mammary glands of 30 mice. Two died during surgery.

To assess whether the tumour cells had successfully engrafted, one mouse (mouse 704) was sacrificed day 19 post initial injection which confirmed the presence and *in situ* growth of the MM134-RFP-Luc2 cells in the mammary ducts. Lamin AC IHC was used to confirm that cells were of human origin and immunofluorescence imaging used to confirm the presence of RFP+ cells (Figure 6.6). IVIS signal in the glands of the remaining mice was performed every 3 weeks thereafter (Figure 6.7).

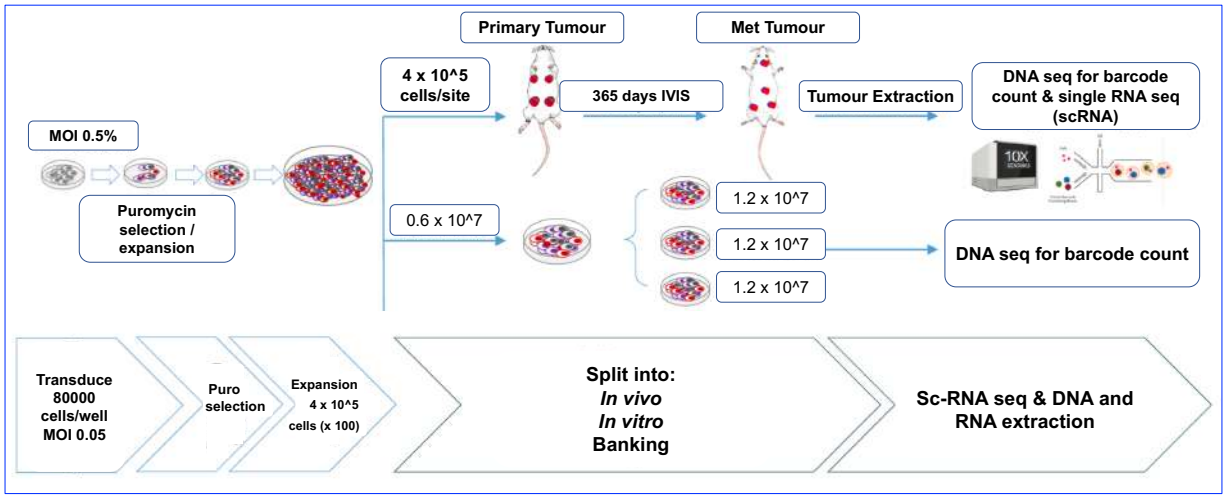


Figure 6.5: Experimental design of ILC barcoding experiment: Schematic diagram summarising the initial design of the MM134-RFP-Luc2 *in vivo* lineage tracing experiment.

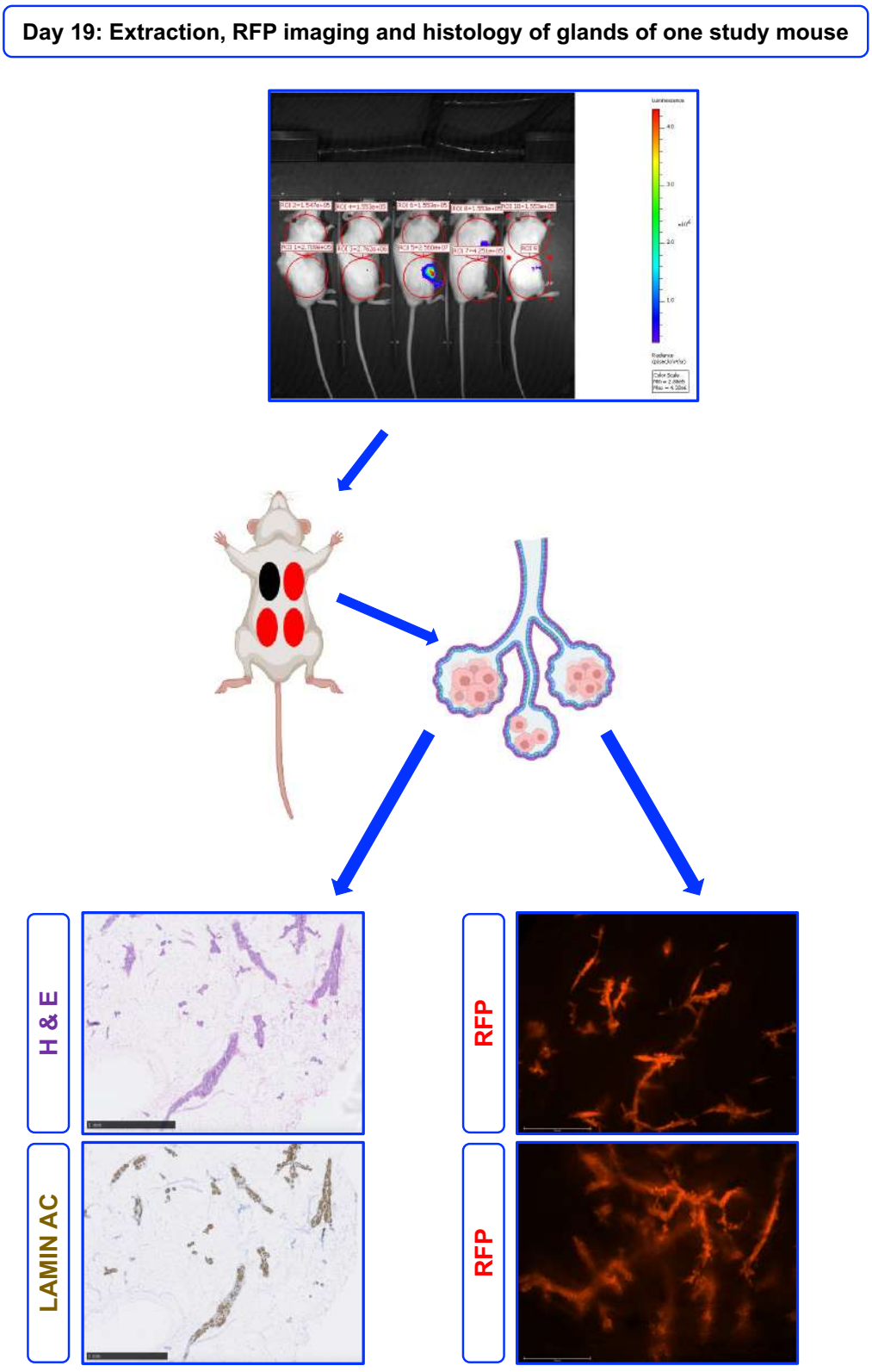


Figure 6.6: Barcoded MM134-RFP-Luc2 cell line *in vivo* growth in mouse 704 19 days following injection using the MIND model: Schematic diagram showing the selection of mouse with positive IVIS gland signal day 19 post intraductal injection, and representative H&E showing *in situ* tumour cell growth, IHC showing positive Lamin AC staining confirming growth of cells of human origin, and RFP imaging confirming MM134-RFP-Luc2 cells localized within the mammary ducts.

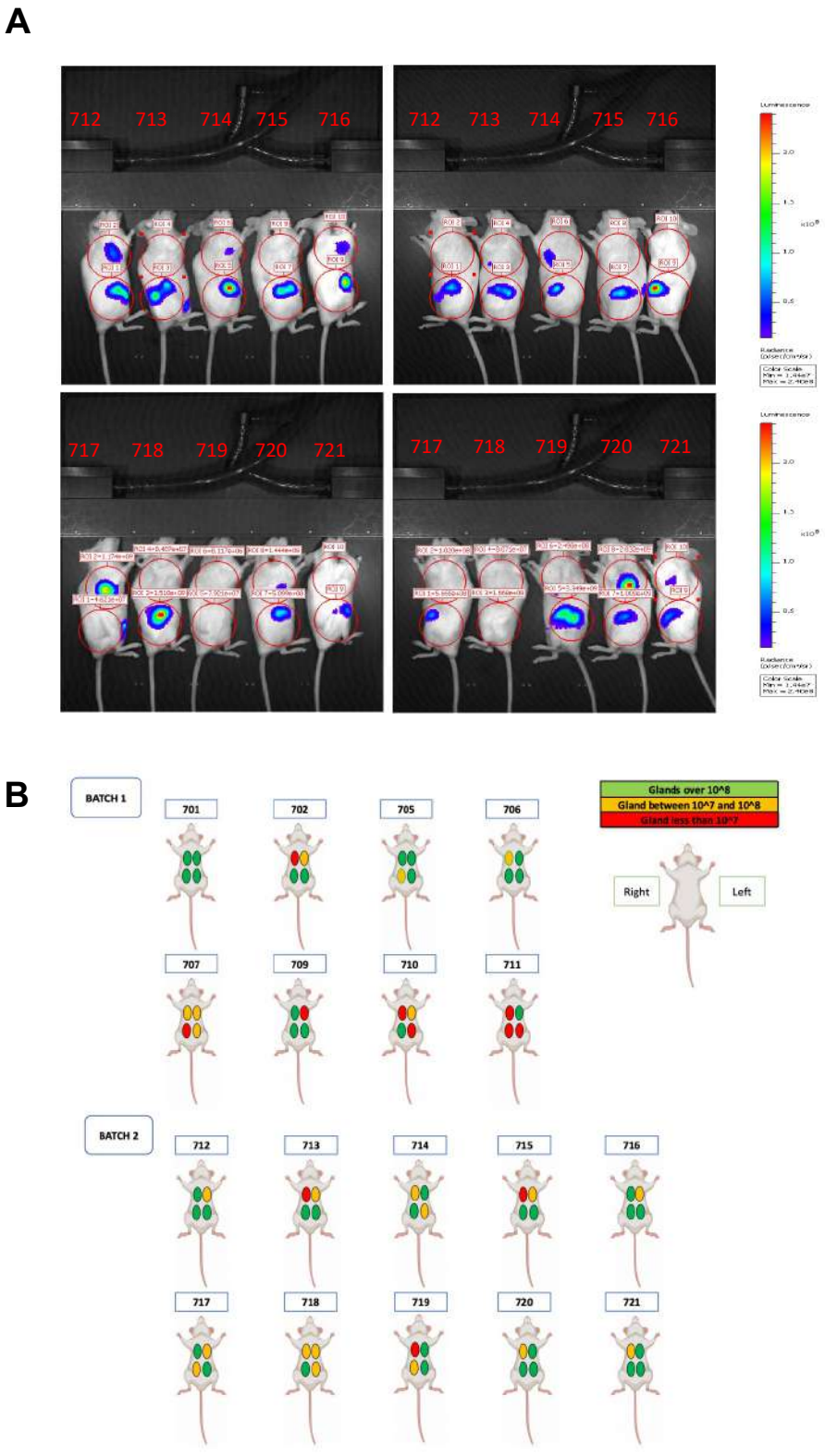


Figure 6.7: Barcoded MM134-RFP-Luc2 form primary tumours *in vivo* following intraductal injection: A) IVIS imaging of a representative mice in the cohort day 217 - 219 post injection B) Schematic diagram showing representative mice summarising the level IVIS signal within each gland.

6.34 The ILC starting population is dominated by a low number of barcodes

Following sacrifice of mouse 704 19 days post injection, and DNA extraction from IVIS signal positive sites, barcode presence and diversity was assessed in the mammary glands (3rd left, 4th left) and lymph nodes (3rd right and 4th left) in addition to the parental starting cells and a barcoded IC-NST cell line control (MCF7). This showed that the MM134-RFP-Luc2 starting population itself had a significant skew and appeared to be dominated by a small number of barcodes rather than a rich barcode diversity that the experiment was designed to contain, as evidenced in the cumulative frequency plot (Figure 6.8A). Greater diversity was observed in some other glands such as the 3rd left mammary gland and 4th right lymph node (Figure 6.8B, C).

Further DNA was extracted and sequenced from the mammary glands of 8 independent sacrificed mice with positive IVIS signal together with a repeat of the initial parental starting cells prior to injection (Figure 6.9). This showed some diversity of barcode in some glands e.g., 731 3rd right (Figure 6.9 A) and 705 4th left (Figure 6.9 E) but significant skew towards a low number of dominant barcodes in other samples (Figure 6.9 G, H) including the starting population (Figure 6.9 I).

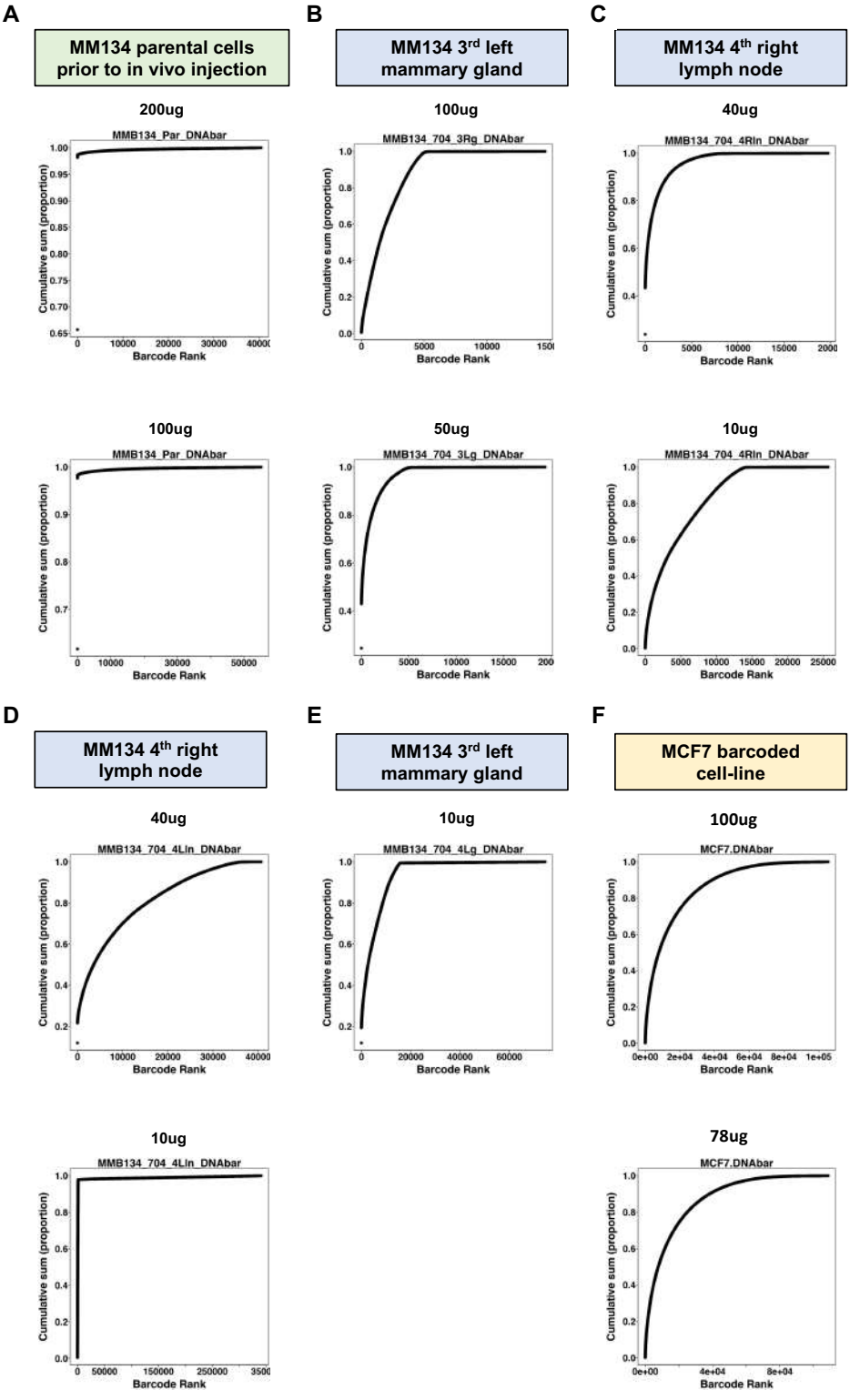


Figure 6.8: Low barcode diversity is present in the barcoded MM134-RFP-Luc2 starting population: Cumulative frequency plots showing barcode diversity in: A) barcoded MM134-RFP-Luc2 starting population prior to injection (green) B – E) mouse 704 mammary gland tumours and IVIS positive lymph nodes (blue) and F) a barcoded IC-NST control, MCF7 (yellow). This control shows a rich diversity of barcodes. Two frequency plots per sample reflect PCR amplification from the same sample on different dates and the concentration of DNA used is summarised above each graph.

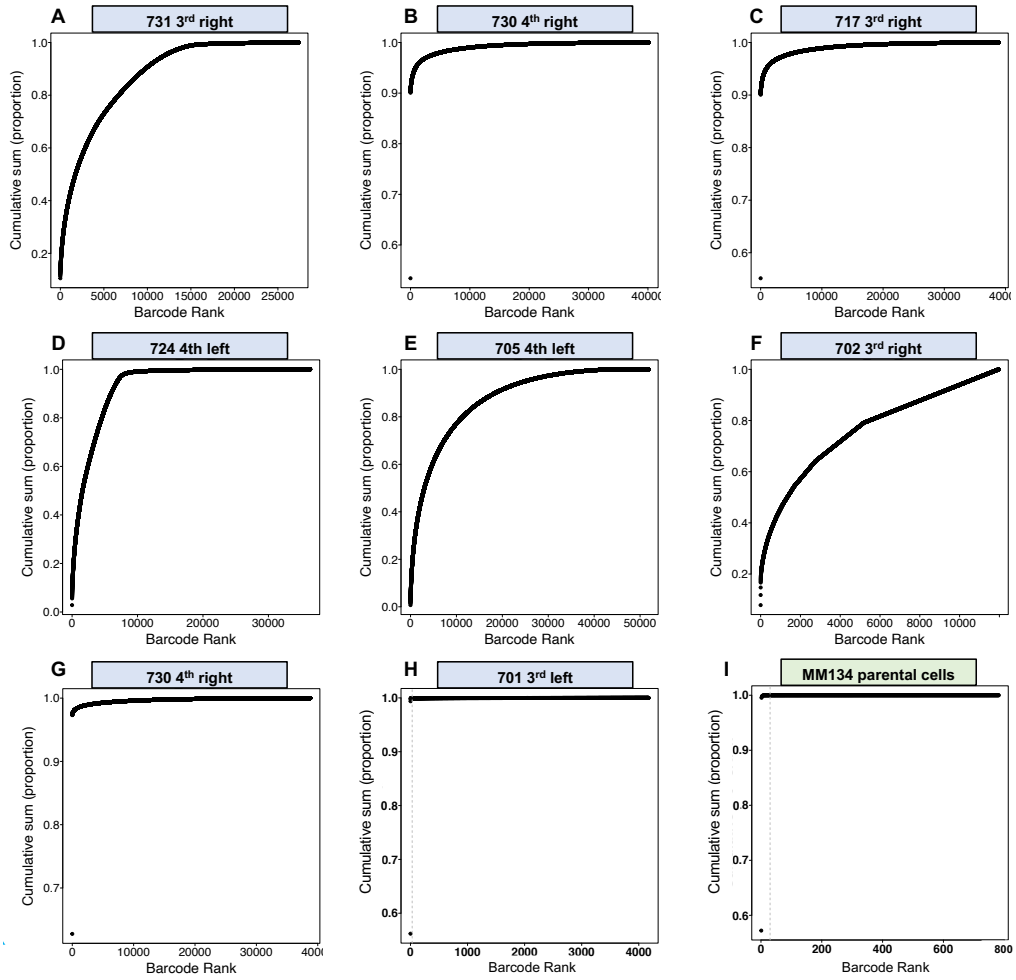


Figure 6.9: Heterogeneity in barcode diversity is observed among further primary mammary glands: Cumulative frequency plots showing barcode diversity in A) Mouse 731 3rd right gland B) Mouse 730 4th right gland C) Mouse 717 3rd right gland D) Mouse 724 4th left gland E) Mouse 705 4th left gland F) Mouse 702 3rd right G) Mouse 730 4th right gland H) Mouse 701 3rd left gland I) barcoded MM134-RFP-Luc2 starting population prior to injection (green). Note significant skew of barcodes in G) – I).

6.4 Discussion

Overall the preliminary work in this chapter has demonstrated the utility of single nuclei sequencing as a novel approach for identifying subclonal transcriptomic alterations which drive aggressive tumour behaviour in an ILC MIND *in vivo* model, which may otherwise go undetected when using traditional bulk RNA sequencing approaches. This initial experiment only used one ILC cell line and one site of metastatic spread (brain metastases). Further work will be completed going forward using additional ILC cells lines and further ILC metastatic sites to gain further insights into the drivers of metastatic disease in ILC at the single nuclei level and to assess transcriptomic differences between subclones driving relapse at different ILC metastatic sites. The findings will be compared to bulk data generated from similar ILC MIND *in vivo* models.

The experiment also identified challenges specific to the metastatic site of study, since for the ovarian metastases there was a high level of contamination of mouse stromal cells identified at the pre-processing stage with a very small proportion of human cells remaining for analysis. Fluorescence-activated cell sorting (FACS) will be a useful approach going forward for removing mouse contaminant cells, particularly in the context of ovarian lesions. A further limitation of the experiment was that by performing single nuclei as opposed to single cell sequencing, some transcripts located in the cytoplasm would have been lost, although the approach does have some advantages over single cell sequencing [457] in that it provides a less biased cell coverage and does not seem to suffer cellular isolation based transcriptional artifacts. Further it can be used with archived frozen tissue which is practically useful.

When the prognostic significance of the genes enriched in brain metastases were assessed in larger ILC cohorts including the KHP pleomorphic ILC cohort, *GCNT2* was identified as significantly associated with early relapse (< 3 years of primary diagnosis) in both METABRIC and the KHP pleomorphic ILC cohorts. *GCNT2* (Glucosaminyl (N-Acetyl) Transferase 2) encodes a branching enzyme which converts linear into branched poly-N-acetyllactosaminoglycans. It is responsible for the formation of the blood group I antigen and is associated with the development and maturation of erythrocytes [386]. It has been shown to be over-expressed among highly metastatic breast cancer cell lines and associated with a metastatic phenotype in breast tumour samples [458]. Interestingly in the KHP pleomorphic cohort the patient whose tumour had the highest level of expression of *GCNT2* relapsed early with brain metastases. Further functional *in vivo* studies have demonstrated that it enhances cell detachment and tumour cell adhesion to endothelial cells, and promotes cell migration

and invasion, as well as being involved in epithelial to mesenchymal transition (EMT) [458]. Further experiments will be useful in assessing if this gene is consistently enriched in brain metastatic lesions and whether it is expressed in other ILC metastatic sites.

Lineage tracing is an approach which has the potential to gain further insights into the metastatic process in ILC *in vivo* models, which cannot be performed in patients, as replicates can be used in the *in vivo* setting. A limitation however is that the mice used are immunodeficient and therefore this eliminates the effect of the immune system on the metastatic process and interactions between aggressive tumour subclones and various immune subpopulations cannot be studied. Whilst the barcoding element of the experiment in this chapter using the barcoding approach for lineage tracing did not work out as planned, the experiment demonstrated that the MIND approach worked effectively, and it provides a basis for future *in vivo* lineage tracing experiments in metastatic ILC.

Overall, the work in this chapter has mapped the sub-clonal heterogeneity of metastatic ILC by using novel single nuclei RNA sequencing approaches and identied that *GCNT2*, a gene significantly more highly expressed in ILC brain metastatic lesions compared to the primary tumour, is also associated with early disease relapse in independent ILC cohorts. This experiment provides a foundation for future ILC single nuclei experiments. The barcoding experiment also provides a foundation for future work. Having identified the transcriptomic profiles of aggressive ILC tumour subclones driving metastases in the *in vivo* models, their spatial locations will be further assessed in the human patient tumours from the NanoString WTA analysis, to better understand their interactions and crosstalk will other tumour, stromal and immune cell populations.

Chapter 7: Discussion

7.1 Discussion and future perspectives

Invasive Lobular Carcinoma (ILC) is the second most common histological subtype of breast cancer accounting for 10 - 15% of all cases. It is a unique disease entity with distinct histological appearances, molecular alterations and clinicopathologic features. A subgroup of ILC patients have clinically aggressive disease with metastases occurring early (< 3 years) after primary diagnosis. In addition a rare ILC subtype pleomorphic ILC, is also associated with more aggressive tumours and early relapse. Overall these patients have limited treatment options, representing a clinically unmet need. There has been a need to better understand the molecular basis and transcriptional drivers of aggressive disease biology in ILC, as well as the ILC immune landscape to help identify potential new drug targets to improve clinical outcomes for these patients.

To address this unmet need, the purpose of this PhD was to characterise the transcriptional and immune heterogeneity in clinically aggressive ILCs including pleomorphic ILC. The key aims of the project were to:

1) Identify the molecular drivers of clinically aggressive ILC
2) Evaluate the immune landscape and prognostic associations in ILC
3) Map the heterogeneity of immune infiltrate in pleomorphic ILC
4) Map the sub-clonal heterogeneity of metastatic ILC

These were addressed as follows (Figure 7.1):

Aim 1: Identify the molecular drivers of clinically aggressive ILC

Genomic alterations were characterised through targeted sequencing of a cohort of pleomorphic and non-pleomorphic ILCs with a range of different clinical outcomes. This identified differences between pleomorphic and non-pleomorphic ILCs at the genomic level and an association between *FGFR1* alterations and disease relapse in pleomorphic ILC. The transcriptomic features of aggressive ILCs were characterised in an independent pleomorphic ILC cohort (KHP cohort) whereby tissue needle macro-dissection resulted in a high tumour cell content. A prognostic gene expression signature derived from this discovery cohort validated in multiple independent matched purified and deconvolved ILC cohorts.

Transcriptomic features of clinically aggressive non-pleomorphic ILCs were also characterised.

Aim 2: Evaluate the immune landscape and prognostic associations in ILC

The immune landscape was characterised in a cohort of pleomorphic and non-pleomorphic ILCs at the histological level through the quantification of stromal TILs. This showed that generally ILCs are characterised by low stromal TILs, but a subset of cases contains higher levels. Moreover pleomorphic ILC is associated with higher levels of stromal TILs than non-pleomorphic ILC and this was after having controlled for higher rates of ER- and HER2+ disease in pleomorphic ILC. Prognostic associations were assessed through survival analysis in separate pleomorphic and non-pleomorphic cohorts which showed that the level of stromal TILs does not hold prognostic significance in ILC.

Aim 3: Map the heterogeneity of immune infiltrate in pleomorphic ILC

The heterogeneity of the immune infiltrate in pleomorphic ILC was mapped using NanoString Digital Spatial (DSP) technology which enabled the further characterisation of stromal TILs into the various immune subpopulations in a subset of histologically immunogenic pleomorphic ILCs as well as the study of their spatial locations. Validation work using IHC revealed the prognostic significance of the ratio of M2-like to M1-like macrophages within the tumour microenvironment in pleomorphic ILC where high M2/M1 ratios are associated with worse metastasis free survival. Moreover co-culture experiments assessing the rate of spheroid growth of ILC cell-line MDA-MB-134 cells in the presence of M1-like and M2-like macrophages further demonstrated that the presence of an M2-like macrophage phenotype promotes enhanced tumour cell growth. A further NanoString whole transcriptome (WTA) experiment using a subset of histologically immunogenic pleomorphic ILCs revealed differences at the transcriptomic level between i) immune-hot and immune-cold tumour cells and ii) immune-hot and immune-cold CAFs. Analysis in independent ILC cohorts identified differentially expressed genes between these groups which are associated with relapse in independent ILC cohorts (*HOXB13*, *CXCL9* and *GOLT1B*).

Aim 4: Map the sub-clonal heterogeneity of metastatic ILC

The sub-clonal heterogeneity of metastatic ILC was mapped using novel single nuclei RNA sequencing approaches. Single nuclei from matched primary and metastatic lesions from an ILC *in vivo* model using MDA-MB-134 cells was performed and the genes and pathways enriched in brain metastatic lesions identified. Their prognostic significance was assessed in independent ILC cohorts which identified that *GCNT2* was associated with early disease relapse in independent ILC cohorts. Assessment within the pleomorphic ILC KHP cohort

revealed that the patient with the highest *GCNT2* expression in their primary tumour also developed early relapse and brain metastases. This work provides a foundation for future ILC single nuclei experiments. Additionally, an ILC barcoding experiment using MDA-MB-134 cells was completed providing a foundation for future work to help determine if subpopulations of ILC tumour cells that are transcriptionally rewired to become metastatic pre-exist in the primary tumour.

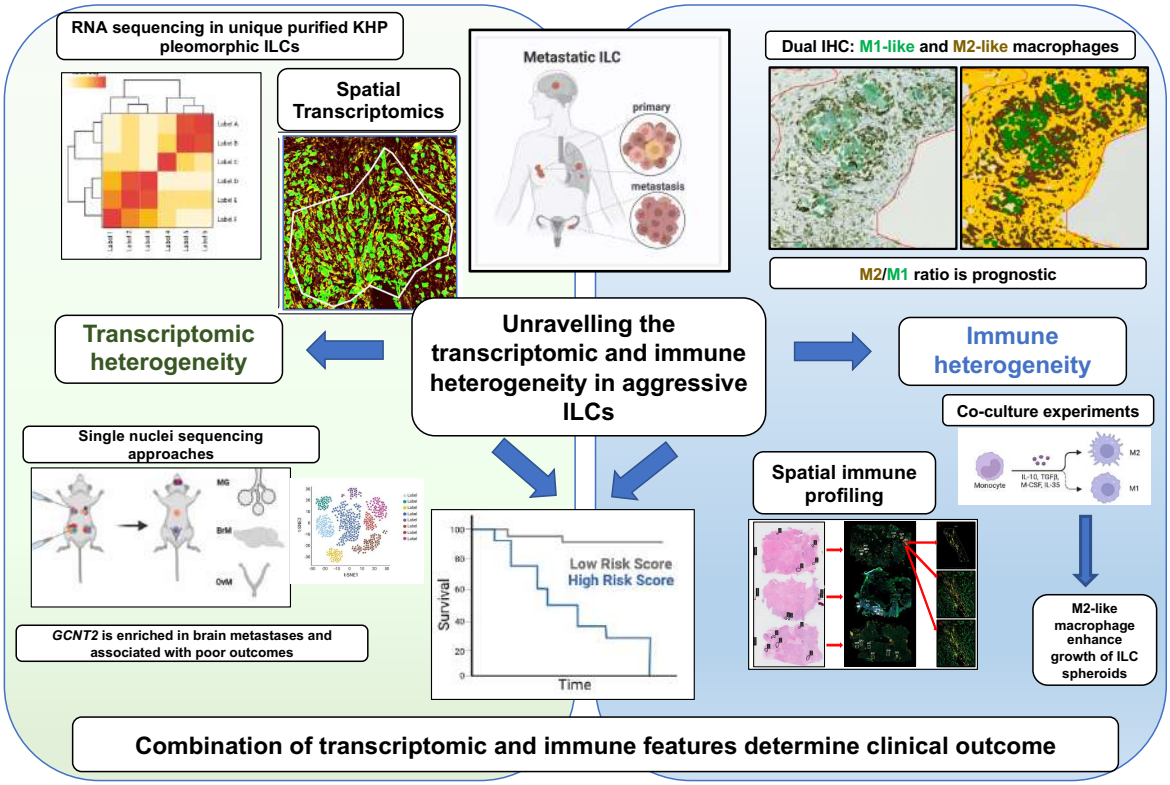


Figure 7.1: Unravelling the transcriptomic and immune heterogeneity in aggressive ILCs

7.2 Clinical Utility of findings

7.2.1 Immune heterogeneity in aggressive ILCs

Overall the work characterising the immune landscape in the KHP cohort has highlighted the relevance of macrophages in ILC and importantly their phenotypes in shaping the nature of the tumour immune microenvironment. Results showing that the ratio of M2-like to M1-like macrophages holds prognostic significance in pleomorphic ILC, suggests that rather than focusing solely on T-centric therapies, treatments targeting immunosuppressive M2-like macrophages may represent an important additional immune-based therapeutic strategy in ILC and may enhance responses to ICIs by promoting T cell infiltration into the tumour microenvironment.

Previous studies have focused on distinguishing between stromal and intra-tumoral locations of macrophages and prognostic associations. For example a recent study performed survival analysis in 30 ILC multiplex IHC cases and showed that higher tumoral but not stromal M1-like macrophages were associated with improved DFS [340]. However the work in this PhD suggests that regardless of spatial location the total relative abundance of M2-like compared to M1-like macrophages is key in shaping the tumour immune microenvironment in pleomorphic ILC and that this can tip the balance towards an immunosuppressive microenvironment favouring tumour progression and disease recurrence, or an inflammatory microenvironment favouring tumour suppression. This is useful given that it can be challenging and involves a degree of subjectivity to accurately distinguish the stromal and tumoral locations of immune cells in ILC, a tumour which is characterised by a diffusely infiltrative, single-file type growth pattern with a high stromal component. Further work, characterising M1-like and M2-like macrophages in larger pleomorphic and non-pleomorphic cohorts will provide further insights into these findings.

Moreover the work in this PhD has revealed that pleomorphic ILCs appear to be more immunogenic tumours than non-pleomorphic ILCs, characterised by higher levels of macrophages and stromal TILs as well as by an increased mutational load. Furthermore, correlation analysis demonstrated some degree of positive correlation between stromal TILs and CD68+ cells suggesting a possible degree of crosstalk between the two immune subpopulations. Indeed M1-like macrophages are known to promote T cell infiltration, and have the ability to present antigens, produce T cell stimulating cytokines (IL-12 and IL-23) and activate pro-inflammatory type-I T lymphocyte responses [333]. Overall given the apparent

enhanced immunogenicity of pleomorphic ILCs compared to non-pleomorphic cases, these patients may represent ideal candidates for treatments targeting the immune-system.

Studies suggest that the abundance and polarisation states of macrophages within the tumour microenvironment are closely associated with the clinical efficacy of ICIs [2, 3]. Moreover a range of different approaches for the targeting of tumour-associated macrophages (TAMs), (considered to represent M2-like phenotypes), are currently being studied in clinical trials. These are often given in combination with ICIs or chemotherapy and one such approach is the reprogramming of M2-like macrophages into M1-like phenotypes through various drugs including Toll-like receptor agonists to promote a more inflammatory immune microenvironment [4]. Other approaches include inhibiting the recruitment of macrophages using agents targeting the CCL2/CCR2 axis and TAM depletion using kinase inhibitors and antibodies targeting the CSF-1/CSF-1R signalling pathway [4]. In TNBC, preclinical work has shown that chemotherapy combined with macrophage Inhibition resulted in T and B lymphocyte Infiltration, and a stronger anti-tumour effect and more tumour regressions compared to chemotherapy alone [5]. The work in this project suggests that ILC patients, whose tumours contain high levels of macrophages, especially pleomorphic ILCs and patients whose tumours contain high M2/M1 ratios, should be considered for such treatments in clinical trials.

Moreover previous studies characterising stromal TILs in ILC cohorts [339, 400] concluded that high levels of stromal TILs (>5% [339] and >10% [282] in the two studies respectively), are associated with a range of poor prognostic features and worse survival outcomes in one study [339]. The results of these studies would therefore suggest that it may be clinically relevant for the diagnostic histopathologist to report the level of stromal TILs in ILC at diagnosis. However these studies did not identify pleomorphic cases. In this PhD the work has provided further clarity and demonstrated that pleomorphic ILCs contain higher stromal TILs than non-pleomorphic ILCs but that stromal TILs alone are not independent predictors of worse outcome in both pleomorphic and non-pleomorphic ILC. The work in this project therefore suggests that the diagnostic reporting of the level of stromal TILs in ILC is not prognostically relevant.

7.2.2 Prognostic gene expression signature

By characterising transcriptomic heterogeneity in the first pleomorphic only ILC RNA cohort, the project identified drivers of aggressive disease behaviour at the gene expression level in

this histologically rare and understudied yet clinically aggressive ILC subtype. Whilst previous studies using unpurified bulk RNA sequencing data from ILC patients in the RATHER and TCGA datasets (n = 106 and n = 144 respectively) [32, 277] identified distinct ILC gene expression subtypes, the RNA work in this project offers the advantage of the fact that the tumours were needle macro-dissected prior to RNA extraction and sequencing, increasing the tumour purity and tumour epithelial cell signal, whilst traditional bulk RNA sequencing data in ILC is typically characterised by a low tumour purity due to the large stromal component which is characteristic of ILC. Moreover distinct gene expression subtypes were not prognostically relevant in the RATHER cohort [277]. In the TCGA analysis, a 'reactive-like' subtype was associated with an improved prognosis compared to the 'proliferative' subtype [32], although the third 'immune' subtype showed no prognostic significance. Furthermore whilst both cohorts identified an 'immune' subtype, these two separate 'immune' subtyping approaches failed to identify the same ILC cases when they were applied to the same dataset [282].

The work in this PhD through RNA sequencing, survival analysis and modelling of the KHP pleomorphic ILC cohort, enabled the generation of a clinically relevant prognostic gene expression signature which validated as predictive of overall survival in larger independent general ILC cohorts. These datasets were purified and deconvolved prior to validation to match the discovery cohort more accurately.

In an era where molecular based prognostic gene expression-based test such as OncotypeDx® and Prosigna® are becoming mainstream in many centres to guide clinical decision making, there has been a need for prognostic gene expression signatures specific for ILC, which is molecularly and clinically distinct from IC-NST for which most of the existing tests are designed. Recently LobSig, a 194-gene signature has been the first gene signature created for specifically prognosticating ILC patients [286]. However the work in this PhD has generated the first lobular signature based on signal derived predominantly from tumour epithelial cells. In future work it will be useful to compare the KHP signature to other breast cancer prognostic signatures such a LobSig and to study the underlying biology of the genes comprising the KHP signature and how they differ from the other signatures which are based on bulk tissue samples. Moreover further work will help decipher if this could be used in combination with an immune-based assessment of the tumour immune microenvironment, quantifying M2-like and M1-like macrophage phenotypes, which could be used to accurately predict outcome in ILC and identify patients who may require more intensive treatment or closer follow-up.

7.3 New insights into the transcriptomic and immune landscape of aggressive ILCs

One of the key opportunities of this PhD project was the chance to use novel technologies to gain new insights into the transcriptomic and immune heterogeneity of clinically aggressive ILCs. NanoString DSP enabled the characterisation and spatial assessment of immune subpopulations at the protein level, whilst NanoString WTA technology enabled new insights into the transcriptomic features of immune cells, tumour cells and CAFs at the gene expression level.

Traditionally there has been a focus on the prognostic relevance of TILs in breast cancer, and the NanoString DSP immuno-oncology panel enabled the characterisation of a broader and diverse range of different immune subpopulations aside from lymphocytes alone. The findings failed to show any prognostic significance of lymphocyte subpopulations instead highlighting the relevance of macrophages within the context of pleomorphic ILC turning the focus onto this immune population. Interestingly the predominance of macrophages in ILC was recently highlighted in a study which assessed a cohort of 87 ILCs and 94 treatment-naive IC-NSTs which showed that macrophages rather than T cells, represent the main immune cell population in ILC accounting for 82% and 50% of the tumour and stromal compartments respectively, compared to 49% and 35% in IC-NST [340]. The technology also enabled spatial analysis which showed that the spatial aggregation of immune cells does not appear to hold prognostic significance in pleomorphic ILC.

In the NanoString WTA experiment, in addition to immune cells, tumour cells and CAFs were also characterised at the transcriptomic level in pleomorphic ILC. Differential gene expression between immune-hot and immune-cold tumour cells identified that *HOXB13* was more highly expressed in immune-cold tumour cells and this gene was associated with poor prognosis in SCAN-B. The comparison between immune-hot and immune-cold CAFs identified two genes; *CXCL9* and *GOLT1B*, which appeared to hold prognostic significance in SCAN-B with high levels being associated with worse outcomes. *CXCL9* has previously been shown to act directly on tumour cells which express the CXCR3 receptor, and this interaction promotes tumour cell migration as well as epithelial mesenchymal transition, promoting metastases [442], consistent with the poor outcomes observed in SCAN-B. In other solid tumour types, e.g. pancreatic cancer, *CXCL9* has been shown to be expressed modestly by CAFs and has also been associated with increased numbers of T cells and immune rich areas [459].

GOLT1B has previously been associated with tumour progression through regulation of the immune microenvironment in breast cancer [443]. A recent study showed that increased *GOLT1B* expression was associated with the infiltration of a range of different immune cell populations including macrophages, dendritic cells and T cell subpopulations as well as immune checkpoint proteins [443]. Its expression was negatively correlated with overall survival in patients with increased regulatory T cells and decreased type 2 helper T cells [443]. The results from the NanoString WTA experiment highlight that CAFs also express *GOLT1B* in pleomorphic ILC and its expression appears to be associated with an enhanced immune infiltrate. Overall these findings provide a foundation for future studies interrogating the transcriptional landscape of CAFs, tumour and immune cells in clinically aggressive ILCs.

Important limitations of both the NanoString DSP and the NanoString WTA experiments include firstly the small numbers of patients included in both experiments meant that there was reduced statistical power to detect differences between the various comparison groups and the low number of patients included is reflective of the high cost of such technologies. Furthermore a limitation of both platforms is that they are unable to offer single cell resolution including co-expression of specific proteins or spatial information at a single cell level, limiting the understanding of cell-to-cell interactions. In terms of practical issues, poor tissue adherence of the breast tumours to the NanoString slides due the inherent nature of breast tissue (with a high fat component) meant that in both experiments a couple of tumours detached from the slides during the initial stages prior to ROI selection. This meant that repeat sections had to be taken, and the original design of the experiment which allocated specific combinations of tumours to each slide, was disrupted. In terms of data analysis there is also a lack of standardisation on the approaches for normalisation of the data and data analysis.

7.4 Conclusion

In conclusion this PhD has characterised the transcriptomic and immune heterogeneity in clinically aggressive ILCs. RNA sequencing in a pleomorphic only ILC cohort identified drivers of clinically aggressive disease in pleomorphic ILC at the gene expression level and enabled the generation of a prognostic gene expression which validated as a predictor of overall survival in larger general ILC cohorts. Characterisation of the immune landscape at the histological level identified that ILCs are generally immune cold tumours yet a subgroup shows higher stromal TILs. Moreover pleomorphic ILCs are characterised by higher stromal TILs than non-pleomorphic ILCs, yet stromal TILs do not hold prognostic significance in both pleomorphic and non-pleomorphic ILC.

Further characterisation and spatial assessment of immune subpopulations in pleomorphic ILC using NanoString DSP showed an association between increased levels of macrophages and early relapse. Although this did not validate in larger ILC cohorts, dual IHC identified that the relative proportions of M2-like to M1-like macrophages holds prognostic significance in pleomorphic ILC with high M2/M1 ratios predictive of worse metastasis free survival. Preliminary co-culture experiments further support the concept that M2-like macrophages have pro-tumour effects on ILC cells promoting faster growth of ILC spheroids compared to M1-like macrophages. These experiments provide a foundation for future co-culture experiments to better understand the cross-talk and interplay between these two cell types. NanoString WTA analysis identified differences at the transcriptomic level between immune hot and immune cold tumour cells and immune hot and immune cold CAFs and differentially expressed genes associated with prognosis in independent ILC cohorts. Finally, a further preliminary experiment using novel single nuclei sequencing approaches identified genes enriched in ILC brain metastatic lesions providing a foundation for future work to understand transcriptomic drivers of disease relapse at a subclonal level.

Supplementary Tables

CASE ID	Pleomoprhic (1) / non pleomorphic (0)	Histology	TILs score (%)	ER (0/1)	PR (0/1)	HER2 (0/1)	lymph nodes (n)	size (mm)	grade	Age at diagnosis (years)	Mets / no mets (1/0)	Time between diagnosis and mets (years)
137	0	classic	4	1	1	0	24	100	2	69.2	1	15.88
144	0	classic	2	1	0	0	0	60	2	43.5	1	1.98
245	0	classic	1	1	1	0	0	150	2	79.5	1	0.00
672	0	classic	2.5	1	1	0	0	20	2	42.8	1	8.40
684	0	solid	3	1	1	0	18	25	3	44.2	1	0.22
758	0	classic	7.5	1	1	0	2	24	2	56.1	1	15.05
759	0	classic	1	1	1	0	10	30	2	66.0	1	8.40
843	0	classic	5	1	1	0	2	28	3	41.7	1	6.87
1046	0	classic	4.5	1	0	0	23	18	2	47.5	1	12.74
1155	0	classic	7.5	1	1	0	2	50	2	61.5	1	2.21
1190	0	classic	2	1	1	0	5	47	2	52.3	1	9.32
1355G	1	pleomorphic + apocrine	7	1	0	0	2	22	2	62.5	1	12.92
1529	0	classic	3	1	1	0	6	28	2	38.8	1	5.31
1605	0	classic	2.5	1	0	0	0	25	2	67.8	1	4.69
1867	0	classic	3	1	1	0	2	40	2	60.1	1	3.46
2229	0	classic	3	1	1	0	1	22	2	56.3	1	0.81
2272	0	classic	2	1	1	0	6	50	2	69.3	1	5.20
2422	0	classic	3	1	1	0	5	30	2	72.8	1	6.48
3117	1	pleomorphic	5.5	1	1	0	1	17	2	78.0	1	3.03
3160	0	classic	2	1	1	0	0	50	2	61.7	1	0.00
3589	1	pleomorphic	1	1	1	0	12	120	2	39.2	1	5.73
4035	0	classic	5	1	1	0	0	50	2	59.2	1	12.20
4860	0	classic	2	1	0	0	0	30	2	75.2	1	2.60
5817	0	classic	17	1	1	0	0	30	2	54.5	1	0.00
6389	0	classic	13.5	1	0	0	1	30	2	78.9	1	6.81
6700	0	classic	5	1	0	0	0	30	2	47.3	1	3.57
6813	0	classic	1.5	1	1	0	5	70	2	51.9	1	4.44
7414	0	classic	2.5	1	0	0	0	50	2	65.6	1	12.35
7490	0	classic	2	1	0	0	0	44	2	52.2	1	2.17
7570	0	classic	0.5	1	1	0	0	40	2	58.9	1	4.13
7756	1	pleomorphic	0.5	1	0	0	0	25	2	59.0	1	8.39
8207	0	classic	3.5	1	1	0	0	20	2	50.1	1	11.70
8268	0	classic	2	1	1	0	1	25	2	61.6	1	8.24
8301	0	mixed - classic and solid	7.5	1	1	0	0	35	2	33.6	1	0.00
8327	0	classic	4.5	1	1	0	5	30	2	56.8	1	3.20
8353	0	classic	2	1	1	0	0	60	2	49.6	1	3.67
8620	0	classic	1	1	1	0	1	18	NA	54.0	1	6.96
8627	0	mixed solid and alveolar apocrine cells	12	1	0	0	8	25	2	69.1	1	1.59
8699	1	pleomorphic	4.5	1	1	0	0	30	3	40.8	1	1.36
8780	0	classic	2	1	0	0	0	20	2	57.1	1	0.00
8785	0	classic	1	1	1	0	6	25	2	58.1	1	0.00
9005	0	classic	3.5	1	1	0	0	40	2	43.2	1	8.09
9068	1	pleomorphic	42.5	1	0	0	1	40	3	45.5	1	1.13
9072	0	classic and alveolar - mixed	4	1	1	0	5	40	2	44.9	1	6.07
9129	0	classic	1	1	1	0	2	60	2	42.5	1	12.07
9165	0	alveolar not pleomorphic	0	1	0	0	0	20	2	58.9	1	1.85
9252	0	classic and alveolar and solid - mixed	3	1	1	0	6	30	2	48.7	1	3.15
9314	0	classic and alveolar and solid - mixed	1	1	1	0	1	80	2	23.1	1	2.95
9326	0	classic	1	1	0	0	0	20	2	55.1	1	3.78
9350	1	pleomorphic	1.5	1	1	0	0	70	2	59.1	1	1.94
9353	0	classic	1.5	1	1	0	5	35	2	69.5	1	4.57
9423	0	classic	1.5	1	1	0	0	60	2	52.1	1	14.50
9536	0	classic	3.8	1	1	0	0	30	2	54.6	1	3.43
9573	0	classic	2	1	1	0	0	25	2	52.3	1	7.73
9894	1	pleomorphic and solid	2	1	1	0	0	28	2	61.7	1	0.00
10293	0	classic	5	1	0	0	10	150	2	74.7	1	2.20
17011789	1	pleomorphic	10	1	1	0	0	40	3	47.9	0	NA
17013387	1	pleomorphic	7	1	0	NA	1	60	3	67.3	1	5.32
17013509	1	pleomorphic and solid areas	2.5	1	1	0	1	40	3	63.0	1	9.20
17013592	1	pleomorphic	5	1	1	0	6	40	3	57.2	1	1.57
17013766	1	pleomorphic	1.5	1	0	0	13	40	NA	49.1	1	1.38
17014363	1	pleomorphic	7	1	0	1	22	40	NA	53.9	1	0.52
17015028	1	pleomorphic	8	1	1	0	0	60	3	79.7	0	NA
17018625	1	mixed classic and pleo	7	1	1	0	8	50	2	47.3	0	NA
17020462	1	pleomorphic	2	0	0	1	2	40	2	79.8	1	1.62
17020498	1	pleomorphic	2	1	1	0	3	40	2	73.6	0	NA
17021357	1	pleomorphic	5.5	0	0	0	NA	30	NA	53.5	0	NA
17021740	0	classic	2	1	1	0	1	60	1	83.9	0	NA
17022733	1	pleomorphic	10	1	1	1	1	20	NA	36.2	0	NA
17022921	0	solid and aleolar and classic mixed non pleo	0	1	0	0	4	60	NA	53.6	1	1.49
17022935	0	alveolar classic	2	1	1	0	0	20	3	79.2	0	NA
17023306	1	pleomorphic and pleomorphic LCIS	8.5	0	0	1	33	30	3	58.1	1	1.00
17023387	1	pleomorphic	3	1	1	0	1	20	2	69.2	0	NA
17023670	0	solid and classic not pleomorphic	1	1	1	0	0	80	3	72.4	1	0.21
17025943	1	pleomorphic + lymphovascular invasion	3	1	1	0	0	20	2	65.3	0	NA
17025981	1	pleomorphic	5	1	1	1	2	50	3	88.1	1	1.07
17037705	0	classic	1	1	1	0	0	NA	2	60.0	0	NA
17037739	1	pleomorphic	4	1	1	1	0	40	3	77.9	0	NA

CASE ID	Pleomoprhic (1) / non pleomorphic (0)	Histology	TILs score (%)	ER (0/1)	PR (0/1)	HER2 (0/1)	lymph nodes (n)	size (mm)	grade	Age at diagnosis (years)	Mets / no mets (1/0)	Time between diagnosis and mets (years)
17047324	1	pleomorphic	16	1	1	0	33	40	2	66.6	1	0.00
17051250	1	pleomorphic	1	1	1	0	0	40	3	80.5	0	NA
17051586	1	pleomorphic	7	0	0	0	0	10	2	66.1	1	6.21
17051630	1	predominantly classic and minor pleo (<10%)	1	1	1	0	0	60	2	73.7	0	NA
17054621	1	pleomorphic	4	1	1	0	1	32	NA	51.7	1	21.36
17055086	1	pleomorphic	2.5	1	1	0	5	40	3	42.7	0	NA
17055238	0	classic	7.5	1	1	0	0	40	2	80.5	0	NA
17055645	1	pleomorphic	6	0	0	0	0	10	2	43.4	1	18.81
17056290	1	pleomorphic	4	1	1	0	0	24	3	69.4	1	2.22
17056531	1	pleomorphic	2	1	1	0	0	20	NA	51.7	0	NA
17056904	1	mixed alveolar pleomorphic and classic	4	1	1	0	0	30	3	92.2	0	NA
17057552	0	classic	1	1	1	0	4	60	2	49.3	0	NA
17057784	1	pleomorphic	4.5	0	0	0	11	40	2	60.8	0	NA
17057789	1	pleomorphic	1	1	0	0	0	20	2	89.1	0	NA
17058082	1	pleomorphic	2	1	1	1	0	25	3	53.8	0	NA
17058145	1	pleomorphic	4	1	1	0	3	60	3	59.6	1	11.50
17058332	1	pleomorphic	4	1	1	NA	1	60	2	40.4	1	4.07
17058347	1	pleomorphic	1.5	0	0	0	8	5	3	46.8	1	3.30
17024213	0	classic	5	1	1	0	17	50	2	73.9	0	NA
17024226 / 2284	0	classic	1.5	1	1	0	NA	21	2	71.2	1	3.61
20160 / 17029272	0	alveolar non pleomorphic (some pleomorphic LCIS)	1	1	1	0	1	25	2	58.7	1	4.93
2602 / 17024882	1	mixed classic and ploe	1	1	1	0	8	90	2	55.1	1	0.65
2851 / 17057134	0	classic	10	1	1	0	4	37	2	56.7	1	14.98
30151 / 17047580	1	pleomorphic	2.5	1	1	0	35	60	2	43.2	1	4.30
4125 / 17004447	0	classic	1	1	1	0	1	30	2	45.8	1	10.74
4148 / 17020161	0	classic	1	1	1	0	14	36	2	47.5	1	7.56
8840 / 17016111	0	classic	0	1	1	0	26	95	2	47.0	1	3.24
9102 / 17014681	0	classic	3	1	0	0	33	30	2	54.4	1	1.15
TG00506	1	pleomorphic	2.5	1	NA	NA	4	90	2	46.3	1	8.81
TG01139	1	pleomorphic	5.5	1	1	0	3	40	2	49.4	0	NA
TG01183	0	classic	1.5	1	NA	0	5	25	2	56.8	0	NA
TG01189	1	pleomorphic	3	1	NA	0	1	30	2	49.0	0	NA
TG01725	1	pleomorphic	35	NA	NA	1	1	11	3	42.7	0	NA
TG01789	0	classic	1	1	NA	0	2	21	2	54.5	0	NA
TG01941	1	pleomorphic	4.5	1	NA	0	0	50	3	52.0	0	NA
TG01958	1	pleomorphic	10	0	NA	0	5	2	2	34.3	0	NA
TG01964	1	pleomorphic	3.5	1	NA	0	2	11	2	43.8	0	NA
TG02006	1	pleomorphic	1.5	NA	NA	0	0	24	2	49.2	1	2.01
TG02072	1	pleomorphic	2	NA	NA	NA	4	60	2	44.5	0	NA
TG02166	1	pleomorphic	5	1	1	0	0	40	3	56.1	1	NA
TG02210	1	pleomorphic	6.5	1	1	0	0	13	3	57.5	0	NA
TG02222	1	pleomorphic	5	NA	NA	NA	1	12	2	54.4	0	NA
17014653	0	classic	1	1	1	0	1	50	2	69.6	0	NA
17014838	0	solid not pleomorphic	1	1	0	0	0	40	NA	79.1	1	0.00
17016969	0	classic	1	1	1	0	23	40	NA	72.8	0	NA
17017163	0	classic	2	1	0	0	20	40	NA	73.0	0	NA
17017388	0	classic	3	1	1	NA	0	20	NA	45.5	0	NA
17018357	0	classic	8	1	0	1	1	20	2	57.3	0	NA
17018737	0	mixed classic and alveolar	3	1	1	0	2	60	2	40.8	0	NA
17018751	0	clasisic	1	1	1	NA	0	20	1	67.5	0	NA
17019723	0	intracytoplasmic lumina classic	2	1	1	0	3	20	NA	46.9	0	NA
17019917	0	classic	2	1	NA	1	0	20	2	81.7	0	NA
17021537	0	classic - histocytoid	2	1	1	NA	4	50	NA	52.8	0	NA
17021562	0	classic	3	1	0	1	0	20	NA	60.7	0	NA
17022081	0	classic	6	1	1	NA	0	50	2	54.9	0	NA
17022814	0	classic	3	1	1	0	0	20	NA	54.0	0	NA
17046914	0	classic	2	1	1	0	1	20	NA	56.4	0	NA
17055597	0	classic	1	1	1	0	0	20	2	69.6	0	NA
17004445	0	classic	2	1	0	NA	0	40	NA	71.6	0	NA
17006157	0	classic	5	1	0	NA	0	20	NA	79.3	0	NA
17006800	0	classic	1	1	0	NA	0	20	NA	61.7	0	NA
17007789	0	classic	2	1	0	NA	0	10	NA	63.3	0	NA
17008110	0	classic	0	1	0	NA	2	30	NA	59.6	0	NA
17015966	0	classic	1	1	0	0	7	30	NA	79.6	0	NA
17019918	0	classic	0	1	0	NA	21	60	1	71.6	0	NA
17020262	0	classic	1	1	1	NA	0	20	NA	75.6	0	NA
17020524	0	classic	2	1	1	0	7	60	NA	62.7	0	NA
17023000	0	classic	3	1	NA	NA	0	20	NA	64.8	0	NA
17025683	0	classic	2	1	1	0	7	60	2	73.3	0	NA
17037453	0	classic	0	1	1	0	2	40	2	71.3	0	NA
17037629	0	classic	4	1	1	0	0	30	2	75.6	0	NA
17042106	0	classic	1	1	0	NA	0	40	NA	66.6	0	NA
17043414	0	classic	1	1	0	0	0	30	2	72.1	0	NA
17047322	0	classic	0	1	1	0	0	60	2	52.6	0	NA
17020041	1	pleomorphic + pleo LCIS	1	1	0	0	2	10	NA	52.7	0	NA
17022260	0	solid	3	1	0	0	0	30	NA	88.0	0	NA
17004052	1	pleomorphic	4	1	1	0	0	20	NA	80.8	0	NA
17006677	1	pleomorphic + mitoses	2	1	0	NA	0	30	NA	71.1	0	NA
17007630	1	pleomorphic with solid areas	2	1	0	NA	0	30	NA	63.1	0	NA
17041799	1	pleomorphic with solid areas	9	1	1	NA	0	30	NA	72.7	0	NA
17022542	1	pleomorphic	2	1	1	NA	3	80	NA	68.6	0	NA
17004562	1	pleomorphic	3	1	0	NA	0	40	NA	56.5	0	NA
17011975	1	pleomorphic	35	1	0	NA	0	30	2	63.1	0	NA
17042643	0	classic	4	1	1	NA	0	20	NA	63.7	0	NA
17049882	0	classic and solid (not pleo)	3	1	1	0	0	40	2	80.6	0	NA

Supplementary Table 1: Clinical detail of cohort

ABL1	BCOR	CDKN2C	FADD	HRAS
ACVR1	BCORL1	CEBPA	FAM46C	IDH1
AKT1	BIRC3	CHEK1	FANCI	IDH2
AKT2	BRAF	CHEK2	FANCL	IGF1R
AKT3	BRCA1	CIC	FAT1	IRF4
ALK	BRCA2	CKS1B	FBXW7	IRS2
AMER1	BRIP1	CREBBP	FGF10	JAK2
ANTRX2	BTG1	CRLF1	FGFR1	KRAS
APC	BTK	CTNNB1	FGFR2	KBTBD4
AR	CALR	CXCR4	FGFR3	KDM6A
ARAF	CASP8	DAXX	FGFR4	KDR
ARID1A	CBL	DDR2	FH	KEAP1
ARID1B	CCND1	DDX3X	FLT3	KIT
ARID2	CCND2	DICER1	FOXL2	KLF2
ASXL1	CCNE1	DROSHA	FOXO1	KMT2A
ATM	CCNE2	ELP1	GATA1	KMT2C
ATR	CD79B	EGFR	GATA3	KMT2D
ATRX	CDH1	EMSY	GNA11	LIN28B
AURKA	CDK12	EP300	GNAQ	LZTR1
AXIN1	CDK2	EPHB2	GNAS	MAP2K1
AXIN2	CDK4	ERBB2	GPR161	MAP2K2
B2M	CDK6	ERBB3	H3F3A	MAP2K4
BAP1	CDKN1A	ESR1	HIST1H3B	MAP3K1
BARD1	CDKN1B	ETV6	HIST1H3C	MAPK1
BBC3	CDKN2A	EZH2	HIST2H3A	MCL1
BCL2	CDKN2B	F2R	HIST2H3C	MDM2
MDM4	OTX1	RAD50	SMARCA4	TCF3
MEN1	PALB2	RAD51B	SMARCB1	TERT
MET	PAX5	RAD51C	SMARCE1	TET2
MLH1	PBRM1	RAD51D	SMO	TFE3
MN1	PAX5	RAD54L	RUNX1	TP53
MPL	PBRM1	RAF1	SDHA	TP63
MRE11A	PDCD1LG2	RB1	SDHB	TSC1
MSH2	PDGFRA	RBM10	SDHC	TSC2
MSH6	PHOX2B	RET	SDHD	U2AF1
MTOR	PIK3CA	RhoA	SETBP2	VHL
MYC	PIK3CD	RICTOR	SETD2	WT1
MYCL	PIK3R1	RIT1	SF3B1	YAP1
MYCN	PIN1	RNF43	SH2B3	DPYD
MYD88	PMS1	ROS1	SMAD2	EBF1
NF1	PMS2	RUNX1	SMAD3	IL3RA
NF2	POLD1	SDHA	SMAD4	KIAA1549
NFE2	POLE	SDHB	SMARCA4	MYOD1
NFE2L2	POT1	SDHC	SMARCB1	TG
NOTCH1	PPM1D	SDHD	SMARCE1	YES1
NOTCH2	PPP2R2A	SETBP2	SMO	YWHAE
NOTCH3	PRKAR1A	SETD2	SOX2	H3F3B
NPM1	PTCH1	SF3B1	SRSF2	ID3
NRAS	PTCH2	SH2B3	STAG2	
NTRK1	PTEN	SMAD2	STK11	
NTRK2	PTPN11	SMAD3	SUFU	
NTRK3	RAD21	SMAD4	TCEB1	

Supplementary Table 2: RMH200Solid targeted gene panel

Immune Cell Profiling Panel (Core)	IO Drug Target Panel	Immune Activation Status Panel	Immune Cell Typing Panel	Pan-Tumor Panel	MAPK Signaling Panel (Early Access)	PI3K/AKT Signaling Panel (Early Access)
Beta-2-microglobulin	4-1BB CD137	CD25	CD14	Bcl-2	BRAF	p-AKT (S473)
CD3	ARG1	CD27	CD34	EpCAM	p-CRAF (S338)	p-GSK3 (S9)
CD4	B7-H3	CD40	CD45RO	ER alpha	EGFR	p-GSK3αβ (S21/S9)
CD8	GITR	CD44	CD66b	Her2/ErbB2	p-JNK (T183/Y185)	INPP4B
CD11c	IDO1	CD80	CD163	MART1	p-MEK1 (S217/S221)	MET
CD20	LAG3	CD127	FAPalpha	NY-ESO-1	p-P38 (T180/Y182)	Pan-AKT
CD45	OX40L	ICOS	FOXP3*	PR	p44/42 MAPK ERK1/2	PLCG1
CD56	STING	PD-L2		PTEN†	p-p44/42 MAPK ERK1/2 (T202/Y204)	p-PRAS40 (T246)
CD68	Tim-3			S100B	p-P90RSK (T359/S363)	p-Tuberin (T1462)
CTLA4	VISTA				Pan-RAS	
Fibronectin						
GZMB†						
HLA-DR						
Ki-67						
Pan-cytokeratin						
PD-1						
PD-L1						
SMA						
Housekeeping genes/negative controls						
Ms IgG1						
Ms IgG2a						
Rb IgG						
GAPDH						
Histone H3						
S6						

Supplementary Table 3: NanoString GeoMx Immuno-oncology panel

Protein	Coefficient relapse vs no relapse	p-value	q-value	Coefficient early vs late relapse	p-value	q-value
S100B	-0.983	0.013	0.983	-0.465	0.360	0.804
CD11c	-0.657	0.053	0.983	-0.257	0.348	0.804
PanCk	0.782	0.076	0.983	-0.061	0.938	0.972
EpCAM	0.750	0.118	0.983	0.665	0.365	0.804
Histone H3	-0.224	0.121	0.983	-0.157	0.316	0.804
Her2	1.377	0.168	0.983	3.038	0.078	0.804
FOXP3	-0.386	0.169	0.983	-0.301	0.565	0.870
CD8	-0.529	0.170	0.983	-0.595	0.188	0.804
Phospho-GSK3B (S9)	0.444	0.176	0.983	-0.607	0.171	0.804
PD-L1	-0.464	0.187	0.983	-0.279	0.565	0.870
Phospho-p38 MAPK (T180/Y182)	-0.250	0.206	0.983	-0.322	0.184	0.804
PR	-0.419	0.228	0.983	0.094	0.856	0.966
CD66b	-0.456	0.237	0.983	0.208	0.667	0.903
CD3	-0.430	0.259	0.983	-0.257	0.606	0.903
GITR	-0.295	0.274	0.983	-0.133	0.714	0.922
pan-RAS	0.212	0.303	0.983	-0.017	0.959	0.972
Bcl-2	0.308	0.304	0.983	-0.157	0.767	0.935
Phospho-PRAS40 (T246)	0.547	0.304	0.983	0.712	0.423	0.856
OX40L	-0.299	0.326	0.983	0.219	0.674	0.903
MET	-0.267	0.327	0.983	-0.762	0.130	0.804
Phospho-JNK (T183/Y185)	-0.632	0.342	0.983	-1.160	0.262	0.804
FAP-alpha	0.246	0.367	0.983	0.460	0.342	0.804
CD45	-0.257	0.382	0.983	-0.175	0.491	0.857
Phospho-MEK1 (S217/S221)	0.255	0.427	0.983	0.845	0.176	0.804
IDO1	-0.414	0.431	0.983	-1.308	0.114	0.804
Phospho-Tuberin (T1462)	0.195	0.432	0.983	-0.452	0.077	0.804
Phospho-AKT1 (S473)	0.239	0.434	0.983	-0.019	0.957	0.972
CD20	-0.554	0.436	0.983	0.314	0.634	0.903
STING	-0.312	0.483	0.983	-0.806	0.260	0.804
CTLA4	0.293	0.486	0.983	-0.646	0.309	0.804
Beta-2-microglobulin	-0.201	0.488	0.983	-0.165	0.762	0.935
SMA	0.190	0.505	0.983	-0.659	0.094	0.804
Fibronectin	0.209	0.505	0.983	0.047	0.916	0.968
MART1	-0.135	0.508	0.983	0.212	0.454	0.856
CD27	-0.222	0.525	0.983	-0.270	0.531	0.870
CD45RO	-0.226	0.532	0.983	0.296	0.568	0.870
INPP4B	0.258	0.546	0.983	-1.068	0.093	0.804
Ki-67	-0.246	0.555	0.983	-0.443	0.418	0.856
CD80	-0.151	0.562	0.983	0.671	0.163	0.804
ER-alpha	0.244	0.568	0.983	-0.858	0.334	0.804
CD127	-0.174	0.570	0.983	-0.895	0.109	0.804
Phospho-GSK3A (S21)/Phospho-GSK3B (S9)	0.099	0.597	0.983	-0.311	0.174	0.804
NY-ESO-1	0.080	0.631	0.983	-0.056	0.773	0.935
CD4	-0.171	0.645	0.983	0.062	0.907	0.968
ICOS	-0.156	0.671	0.983	0.191	0.664	0.903
Phospho-p90 RSK (T359/S363)	-0.105	0.685	0.983	-0.704	0.085	0.804
PTEN	-0.090	0.695	0.983	-0.406	0.275	0.804
CD25	-0.072	0.712	0.983	-0.210	0.421	0.856
EGFR	-0.094	0.725	0.983	-0.036	0.876	0.966
CD68	0.172	0.731	0.983	1.608	0.055	0.804
CD40	0.119	0.738	0.983	0.126	0.672	0.903
GAPDH	0.101	0.756	0.983	-0.207	0.637	0.903
CD163	0.151	0.767	0.983	0.842	0.321	0.804
Phospho-c-RAF (S338)	0.042	0.788	0.983	-0.184	0.448	0.856
Phospho-p44/42 MAPK ERK1/2 (T202/Y204)	0.103	0.792	0.983	0.540	0.330	0.804
VISTA	0.099	0.795	0.983	-0.020	0.974	0.974
p44/42 MAPK ERK1/2	-0.043	0.809	0.983	-0.207	0.477	0.856
GZMB	-0.064	0.809	0.983	-0.221	0.559	0.870
S6	-0.063	0.811	0.983	-0.383	0.470	0.856
PLCG1	0.067	0.816	0.983	-0.874	0.050	0.804
Pan-AKT	-0.068	0.822	0.983	-0.689	0.214	0.804
LAG3	0.043	0.825	0.983	-0.051	0.871	0.966
PD-L2	-0.055	0.826	0.983	0.250	0.480	0.856
CD44	0.066	0.860	0.990	0.310	0.294	0.804
CD14	0.048	0.864	0.990	0.530	0.241	0.804
CD34	-0.089	0.879	0.990	-1.064	0.294	0.804
Phospho-AKT (phospho T308)	0.094	0.884	0.990	-0.095	0.910	0.968
BRAF	-0.016	0.930	0.993	-0.368	0.125	0.804
4-1BB	-0.010	0.948	0.993	-0.032	0.847	0.966
CD56	-0.008	0.950	0.993	-0.107	0.532	0.870
ARG1	0.018	0.954	0.993	0.090	0.860	0.966
PD-1	0.006	0.966	0.993	0.217	0.336	0.804
HLA-DR	-0.008	0.970	0.993	-0.133	0.703	0.922
Tim-3	-0.006	0.980	0.993	0.149	0.725	0.922
B7-H3	-0.004	0.993	0.993	0.134	0.826	0.966

Supplementary Table 4: NanoString DSP differential protein expression data

	Case ID	M2/M1 ratio
Relapse (n = 10)	17013592	4.34
	1355G	2.69
	17014363	3.36
	17013387	3.69
	17023306	3.08
	17051586	2.52
	17047324	1.39
	17025981	2.68
	17055645	2.25
	TG02166	3.22
No relapse (n = 9)	17011789	3.22
	17018625	2.80
	17015028	0.78
	17021357	2.87
	17057784	1.51
	17055238	3.15
	TG01139	3.37
	TG01725	3.50
	TG01958	2.32

	Case ID	M2/M1 ratio
Early relapse (n = 5)	17013592	4.34
	17014363	3.36
	17023306	3.08
	17047324	1.39
	17025981	2.68
Late relapse (n = 4)	1355G	2.69
	17013387	3.69
	17051586	2.52
	17055645	2.25

Supplementary Table 5: NanoString DSP M2/M1 ratios per patient

134 Overall Survival genes				
MPHOSPH8	PPARD	KHDRBS1	CDR2L	SLC39A10
KMT2D	ZNF281	GOLGA1	MFSD11	SOAT1
RUNX2	SLC7A2	ZNF493	XPR1	USP3
PDPR	MTMR14	CYLD	FBLIM1	AP2A2
BAG5	FAM83H	SMAD5	TCP11L2	CAMSAP2
TRABD	ZMYM3	MKNK1	VCPIP1	EXOSC6
STEAP2	ICA1	EIF4E	IRAK1	MROH1
RABGAP1L	PARN	COL6A2	ZDHHC21	ACAD10
RNF4	MPLKIP	MTMR1	KLC4	SLC35A3
NXPE3	ASCC2	AKNA	WRNIP1	CREBRF
ATE1	TMEM132A	FTO	CD109	UBE2B
PJA2	TTC39B	RFX1	SCAPER	RXRA
UQCRB	ADCK2	RAB30	CYB5R3	DNAJC3
SDCBP	CTSO	DNAJC13	ACTR1A	STXBP4
TLK2	MAPK9	RBM5	DAPK2	
BOD1L1	RAP2B	DTX2	PTPN1	
DHX57	TTC7A	IGF1R	ENG	
DOP1A	LETM1	NEK11	GNPNAT1	
ZNF606	RBMS1	PARP9	SNX25	
CRYBG3	KMT2E	LITAF	ARL1	
RPL13	IGFBP5	MRPS14	ARID2	
ALDH2	UBXN6	TTC21B	BRI3BP	
ERP44	ADCY6	MAP3K20	TUFT1	
ARMC8	TGFBR2	PPP2R3A	TRIM65	
APBB2	DNAH5	IL17RD	KDM4C	
GALNT7	GFUS	SSH1	MED15	
TAMM41	CYBRD1	YJU2B	RBM15B	
PHF2	C11orf54	ZC3H13	ANKRD26	
BTN2A1	CDK16	NUDT3	KIAA0513	
YPEL2	NIPA2	CHMP1B	UBP1	

Supplementary Table 6: 134 significant overall survival genes from KHP pleomorphic RNA cohort

356 Metastasis free survival genes							
SLC9A1	ATE1	PPARD	LETM1	G2E3	HSD17B4	TMEM25	STXBP4
TM2D1	PJA2	PARP6	RBMS1	DNAJC13	CD109	AHI1	COX7C
REPS2	PPP1R21	SELENOT	SAV1	SCRIB	AHDC1	PPM1K	BDP1
TMTC3	PTPDC1	TLE5	PALD1	UHRF2	STK19	UBP1	PRR14
MEIS2	ZNF33A	SLCO3A1	MAF	PDE12	AKAP17A	ZDHHC17	CCDC186
TCF12	ZDHHC18	SLC7A2	KMT2E	RBM5	UFM1	FBXL4	NCKAP1L
MDFIC	NAGLU	RPS27	ATF2	DTX2	RNF166	SLC39A10	TMEM167A
PURA	ATM	ZNF460	CYP1B1	SPART	PTPN1	CUL3	CCNL1
MAPKAP1	PLEKHG5	IKBKB	CMTM4	ODR4	CREM	SOAT1	AKAP10
MOSMO	HECTD2	EPHB4	CALCOCO1	PPOX	NHLRC2	PCMT1	PRRC2C
MAPK1IP1L	DDX1	BAG3	NXN	IGF1R	SYTL4	ZNF445	RNF146
MPHOSPH8	ZNF638	RPGR	ADCY6	RHOT2	SEPTIN9	SMARCAD1	PDP1
KMT2D	USP25	TGFBR1	CEPT1	ZNF146	FLRT2	EXOSC6	BPNT2
ACAP2	ACVR2B	HNRNPU	MFSD4B	PARP9	ENG	EP400	YPEL2
CLUAP1	VAV3	ZRANB1	ARHGAP35	CCNG1	ACP6	BCL3	CACNB3
GRINA	LASP1	MTX3	INTS1	SAMD9L	ANAPC5	SEL1L	SPEF2
RUNX2	TBC1D8	MPLKIP	PRPF3	S100PBP	SNX18	ZBTB41	DNASE1
NUP43	LRP5	KDELR2	GFUS	LITAF	SMIM8	EPC1	TDP1
RAB27A	SDCBP	SORBS1	CYBRD1	DDIT4	SNX25	COL9A2	TMEM250
XRN1	DAZAP2	RANBP2	ORC4	RBM26	PTEN	BCL9	PRPF4B
PDPR	ABCC10	C1QTNF3	C2CD3	MRPS14	RAB11FIP2	RTTN	AKNA
TSPAN4	CORO1C	FMNL2	TBRG1	TTC21B	ARRDC1	ZNF875	PHLPP1
CSNK2A1	GNA12	VPS36	SH3TC1	TET2	ARL1	GANC	ANKRD40
AZIN2	BOD1L1	GTF3C5	WDR36	COQ7	RELT	KLHL9	KLHDC4
ARPC1B	DHX57	DACT1	NIPA2	MAP3K20	SLC25A46	SLC20A2	WRNIP1
PPP2R5C	COG3	DCAF5	PIP5K1C	ITSN2	CSE1L	CERS4	SLC25A22
SCAMP4	IKZF1	SRP54	AKT3	C2orf69	TTC39A	ORC2	SLC30A1
MRI1	DOP1A	FAM126B	SULF1	IL17RD	ARID2	TSN	CMTR2
TBC1D15	ZNF235	CWF19L2	BTN3A2	SSH1	CCAR2	INTS8	HMG20B
SLMAP	EI24	SMARCD2	IFI44	BRF1	BCL2L2	SLC35A3	ANKRD26
CRNKL1	NIPBL	PLEKHG2	PDCD4	PREX1	SPATA20	PER1	TTL
TTC14	ZNF280D	TMEM132A	MTAP	EPG5	HSPG2	PPFIBP2	CLUH
PMP22	ZEB2	TNFAIP1	GPX4	YJU2B	SENP6	RALGAPA1	RESF1
TRABD	PIAS2	ZEB1	SEC22B	ZC3H13	COBLL1	UBE2B	ZNF75A
LAMTOR1	PHYKPL	CHST12	MBD5	NUDT3	CTDSP1	PHF23	MPRIP
LZTR1	ZNF37A	MTURN	NIBAN2	SPRYD4	TUFT1	EMSY	RAB5C
NAV2	NBEAL1	PLEKHH2	SUN2	DDX46	ACSS1	RXRA	MRRF
LLGL2	ZFHX4	PSMB7	ZNF91	CDK19	NEMF	DNAJC3	RPS27L
BRAT1	CSNK1G1	MAN1B1	N4BP2	PLEKHA1	RAP1B	ANKRD12	ZDHHC21
RABGAP1L	ARMC8	AP1G1	MKNK1	DOT1L	FASN	TTC19	MED15
CTDSPL2	HINFP	CTSO	TPGS2	MFSD11	KDM4C	RAB22A	ZNF551
SLC35A4	SOCS5	ZNF428	ARFGAP1	XPR1	STRBP	VPS26B	
CEP170	NUB1	ZNF121	ENOSF1	FBLIM1	MLH3	NR2C2	
FAM91A1	RBBP6	LETMD1	FHIP2B	SACS	USP8	NOTCH4	
BAZ2B	GMIP	KIFAP3	COL6A2	STK11	EEA1	BRAF	

Supplementary Table 6: 356 significant metastasis free survival genes from KHP pleomorphic RNA cohort

References

1. McCart Reed, A.E., et al., *Invasive lobular carcinoma of the breast: morphology, biomarkers and 'omics.* Breast Cancer Res, 2015. **17**: p. 12.
2. Li, C.I. and J.R. Daling, *Changes in breast cancer incidence rates in the United States by histologic subtype and race/ethnicity, 1995 to 2004.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(12): p. 2773-80.
3. Li, C.I., et al., *Trends in incidence rates of invasive lobular and ductal breast carcinoma.* JAMA, 2003. **289**(11): p. 1421-4.
4. Pestalozzi, B.C., et al., *Distinct clinical and prognostic features of infiltrating lobular carcinoma of the breast: combined results of 15 International Breast Cancer Study Group clinical trials.* J Clin Oncol, 2008. **26**(18): p. 3006-14.
5. Luveta, J., et al., *Invasive Lobular Breast Cancer as a Distinct Disease: Implications for Therapeutic Strategy.* Oncol Ther, 2020. **8**(1): p. 1-11.
6. Chen, Z., et al., *Invasive lobular carcinoma of the breast: A special histological type compared with invasive ductal carcinoma.* PLoS One, 2017. **12**(9): p. e0182397.
7. Li, C.I., et al., *Hormone replacement therapy in relation to risk of lobular and ductal breast carcinoma in middle-aged women.* Cancer, 2000. **88**(11): p. 2570-7.
8. Manjer, J., et al., *Increased incidence of small and well-differentiated breast tumours in post-menopausal women following hormone-replacement therapy.* Int J Cancer, 2001. **92**(6): p. 919-22.
9. Chen, C.L., et al., *Hormone replacement therapy in relation to breast cancer.* JAMA, 2002. **287**(6): p. 734-41.
10. Daling, J.R., et al., *Relation of regimens of combined hormone replacement therapy to lobular, ductal, and other histologic types of breast carcinoma.* Cancer, 2002. **95**(12): p. 2455-64.
11. Newcomb, P.A., et al., *Postmenopausal estrogen and progestin use in relation to breast cancer risk.* Cancer Epidemiol Biomarkers Prev, 2002. **11**(7): p. 593-600.
12. Newcomer, L.M., et al., *Postmenopausal hormone therapy and risk of breast cancer by histologic type (United States).* Cancer Causes Control, 2003. **14**(3): p. 225-33.
13. Ursin, G., et al., *Does menopausal hormone replacement therapy interact with known factors to increase risk of breast cancer?* J Clin Oncol, 2002. **20**(3): p. 699-706.
14. Li, C.I., et al., *Relationship between established breast cancer risk factors and risk of seven different histologic types of invasive breast cancer.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(5): p. 946-54.
15. Lee, S., et al., *Postmenopausal hormone therapy and breast cancer risk: the Multiethnic Cohort.* Int J Cancer, 2006. **118**(5): p. 1285-91.
16. Tjonneland, A., et al., *Hormone replacement therapy in relation to breast carcinoma incidence rate ratios: a prospective Danish cohort study.* Cancer, 2004. **100**(11): p. 2328-37.
17. Garcia-Closas, M., et al., *Established breast cancer risk factors by clinically important tumour characteristics.* Br J Cancer, 2006. **95**(1): p. 123-9.
18. Li, C.I., et al., *Relationship between menopausal hormone therapy and risk of ductal, lobular, and ductal-lobular breast carcinomas.* Cancer Epidemiol Biomarkers Prev, 2008. **17**(1): p. 43-50.
19. Reeves, G.K., et al., *Hormonal therapy for menopause and breast-cancer risk by histological type: a cohort study and meta-analysis.* Lancet Oncol, 2006. **7**(11): p. 910-8.
20. Rosenberg, L.U., et al., *Menopausal hormone therapy and other breast cancer risk factors in relation to the risk of different histological subtypes of breast cancer: a case-control study.* Breast Cancer Res, 2006. **8**(1): p. R11.

21. Brinton, L.A., et al., *Menopausal hormone therapy and breast cancer risk in the NIH-AARP Diet and Health Study Cohort.* Cancer Epidemiol Biomarkers Prev, 2008. **17**(11): p. 3150-60.
22. Fournier, A., et al., *Use of different postmenopausal hormone therapies and risk of histology- and hormone receptor-defined invasive breast cancer.* J Clin Oncol, 2008. **26**(8): p. 1260-8.
23. Flesch-Janys, D., et al., *Risk of different histological types of postmenopausal breast cancer by type and regimen of menopausal hormone therapy.* Int J Cancer, 2008. **123**(4): p. 933-41.
24. Slanger, T.E., et al., *Menopausal hormone therapy and risk of clinical breast cancer subtypes.* Cancer Epidemiol Biomarkers Prev, 2009. **18**(4): p. 1188-96.
25. Calle, E.E., et al., *Postmenopausal hormone use and breast cancer associations differ by hormone regimen and histologic subtype.* Cancer, 2009. **115**(5): p. 936-45.
26. Kotsopoulos, J., et al., *Risk factors for ductal and lobular breast cancer: results from the nurses' health study.* Breast Cancer Res, 2010. **12**(6): p. R106.
27. Phipps, A.I., et al., *Risk factors for ductal, lobular, and mixed ductal-lobular breast cancer in a screening population.* Cancer Epidemiol Biomarkers Prev, 2010. **19**(6): p. 1643-54.
28. Work, M.E., et al., *Risk factors for uncommon histologic subtypes of breast cancer using centralized pathology review in the Breast Cancer Family Registry.* Breast Cancer Res Treat, 2012. **134**(3): p. 1209-20.
29. Nyante, S.J., et al., *Risk factors for specific histopathological types of postmenopausal breast cancer in the NIH-AARP Diet and Health Study.* Am J Epidemiol, 2013. **178**(3): p. 359-71.
30. Li, C.I., et al., *Use of menopausal hormone therapy and risk of ductal and lobular breast cancer among women 55-74 years of age.* Breast Cancer Res Treat, 2014. **145**(2): p. 481-9.
31. *National Institute for Health and Care Excellence, UK. Early and locally advanced breast cancer: diagnosis and management. 2018. Clinical Guideline NG101*. 2018 [cited 2021 04.07.21]; Available from: https://www.nice.org.uk/guidance/ng101.
32. Ciriello, G., et al., *Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer.* Cell, 2015. **163**(2): p. 506-19.
33. Metzger Filho, O., et al., *Relative Effectiveness of Letrozole Compared With Tamoxifen for Patients With Lobular Carcinoma in the BIG 1-98 Trial.* J Clin Oncol, 2015. **33**(25): p. 2772-9.
34. Cristofanilli, M., et al., *Invasive lobular carcinoma classic type: response to primary chemotherapy and survival outcomes.* J Clin Oncol, 2005. **23**(1): p. 41-8.
35. UK, C.R. *Breast cancer incidence (invasive) statistics*. 2023 [cited 2023 17th April 2023]; Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/incidence-invasive.
36. McCart Reed, A.E., et al., *Invasive lobular carcinoma of the breast: the increasing importance of this special subtype.* Breast Cancer Res, 2021. **23**(1): p. 6.
37. Farhat, G.N., et al., *Changes in invasive breast cancer and ductal carcinoma in situ rates in relation to the decline in hormone therapy use.* J Clin Oncol, 2010. **28**(35): p. 5140-6.
38. Yang, L.Y., L.P. Yang, and B. Zhu, *Clinicopathological characteristics and survival outcomes of invasive lobular carcinoma in different races.* Oncotarget, 2017. **8**(43): p. 74287-74298.
39. Dossus, L. and P.R. Benusiglio, *Lobular breast cancer: incidence and genetic and non-genetic risk factors.* Breast Cancer Res, 2015. **17**: p. 37.
40. Collaborative Group on Hormonal Factors in Breast, C., *Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies.* Lancet Oncol, 2012. **13**(11): p. 1141-51.

41. Collaborative Group on Hormonal Factors in Breast, C., *Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies.* Lancet, 1996. **347**(9017): p. 1713-27.
42. Iversen, L., et al., *Lifetime cancer risk and combined oral contraceptives: the Royal College of General Practitioners' Oral Contraception Study.* Am J Obstet Gynecol, 2017. **216**(6): p. 580 e1-580 e9.
43. Organisation, W.H., *Alcohol is one of the biggest risk factors for breast cancer.* 2021.
44. *Food, nutrition, physical activity, and the prevention of cancer: a global perspective. Washington DC: AICR; 2007.* 2007.
45. Travis, R.C. and T.J. Key, *Oestrogen exposure and breast cancer risk.* Breast Cancer Res, 2003. **5**(5): p. 239-47.
46. Middleton, L.P., et al., *Multicentric mammary carcinoma: evidence of monoclonal proliferation.* Cancer, 2002. **94**(7): p. 1910-6.
47. Eliyatkin, N., et al., *Properties of Synchronous Versus Metachronous Bilateral Breast Carcinoma with Long Time Follow Up.* Asian Pac J Cancer Prev, 2015. **16**(12): p. 4921-6.
48. Duraker, N., et al., *A Comparison of the Clinicopathological Features, Metastasis Sites and Survival Outcomes of Invasive Lobular, Invasive Ductal and Mixed Invasive Ductal and Lobular Breast Carcinoma.* Eur J Breast Health, 2020. **16**(1): p. 22-31.
49. Arpino, G., et al., *Infiltrating lobular carcinoma of the breast: tumor characteristics and clinical outcome.* Breast Cancer Res, 2004. **6**(3): p. R149-56.
50. Lopez, J.K. and L.W. Bassett, *Invasive lobular carcinoma of the breast: spectrum of mammographic, US, and MR imaging findings.* Radiographics, 2009. **29**(1): p. 165-76.
51. Parvaiz, M.A., et al., *Breast MRI in Invasive Lobular Carcinoma: A Useful Investigation in Surgical Planning?* Breast J, 2016. **22**(2): p. 143-50.
52. Wasif, N., et al., *Invasive lobular vs. ductal breast cancer: a stage-matched comparison of outcomes.* Ann Surg Oncol, 2010. **17**(7): p. 1862-9.
53. Mathew, A., et al., *Distinct Pattern of Metastases in Patients with Invasive Lobular Carcinoma of the Breast.* Geburtshilfe Frauenheilkd, 2017. **77**(6): p. 660-666.
54. Wong, Y.M., et al., *Infiltrative pattern of metastatic invasive lobular breast carcinoma in the abdomen: a pictorial review.* Insights Imaging, 2021. **12**(1): p. 181.
55. Lamovec, J. and M. Bracko, *Metastatic pattern of infiltrating lobular carcinoma of the breast: an autopsy study.* J Surg Oncol, 1991. **48**(1): p. 28-33.
56. Harris, M., et al., *A comparison of the metastatic pattern of infiltrating lobular carcinoma and infiltrating duct carcinoma of the breast.* Br J Cancer, 1984. **50**(1): p. 23-30.
57. DiPiro, P.J., et al., *Lobular breast cancer: patterns of intraabdominal metastatic spread on imaging and prognostic significance.* Abdom Radiol (NY), 2019. **44**(1): p. 362-369.
58. Inoue, M., et al., *Specific sites of metastases in invasive lobular carcinoma: a retrospective cohort study of metastatic breast cancer.* Breast Cancer, 2017. **24**(5): p. 667-672.
59. Yousef, G.M., et al., *Invasive lobular carcinoma of the breast presenting as retroperitoneal fibrosis: a case report.* J Med Case Rep, 2010. **4**: p. 175.
60. Winston, C.B., et al., *Metastatic lobular carcinoma of the breast: patterns of spread in the chest, abdomen, and pelvis on CT.* AJR Am J Roentgenol, 2000. **175**(3): p. 795-800.
61. Brown, D.L., et al., *Primary versus secondary ovarian malignancy: imaging findings of adnexal masses in the Radiology Diagnostic Oncology Group Study.* Radiology, 2001. **219**(1): p. 213-8.
62. Koyama, T., et al., *Secondary ovarian tumors: spectrum of CT and MR features with pathologic correlation.* Abdom Imaging, 2007. **32**(6): p. 784-95.

63. Mavaddat, N., et al., *Pathology of breast and ovarian cancers among BRCA1 and BRCA2 mutation carriers: results from the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA).* Cancer Epidemiol Biomarkers Prev, 2012. **21**(1): p. 134-47.
64. Barroso-Sousa, R. and O. Metzger-Filho, *Differences between invasive lobular and invasive ductal carcinoma of the breast: results and therapeutic implications.* Ther Adv Med Oncol, 2016. **8**(4): p. 261-6.
65. Montagner, M. and E. Sahai, *In vitro Models of Breast Cancer Metastatic Dormancy.* Front Cell Dev Biol, 2020. **8**: p. 37.
66. Du, T., et al., *Invasive lobular and ductal breast carcinoma differ in immune response, protein translation efficiency and metabolism.* Sci Rep, 2018. **8**(1): p. 7205.
67. Narbe, U., et al., *The Distribution of Circulating Tumor Cells Is Different in Metastatic Lobular Compared to Ductal Carcinoma of the Breast-Long-Term Prognostic Significance.* Cells, 2020. **9**(7).
68. Bergeron, A., et al., *Triple-negative breast lobular carcinoma: a luminal androgen receptor carcinoma with specific ESRRA mutations.* Mod Pathol, 2021. **34**(7): p. 1282-1296.
69. Sahin, S., et al., *Invasive Pleomorphic Lobular Histology Is an Adverse Prognostic Factor on Survival in Patients with Breast Cancer.* Am Surg, 2017. **83**(4): p. 359-364.
70. Eusebi, V., F. Magalhaes, and J.G. Azzopardi, *Pleomorphic lobular carcinoma of the breast: an aggressive tumor showing apocrine differentiation.* Hum Pathol, 1992. **23**(6): p. 655-62.
71. Bentz, J.S., N. Yassa, and F. Clayton, *Pleomorphic lobular carcinoma of the breast: clinicopathologic features of 12 cases.* Mod Pathol, 1998. **11**(9): p. 814-22.
72. Haque, W., et al., *Outcomes of pleomorphic lobular carcinoma versus invasive lobular carcinoma.* Breast, 2019. **43**: p. 67-73.
73. Pathologists, T.R.C.o., *Pathology reporting of breast disease in surgical excision specimens incorporating the dataset for histological reporting of breast cancer.* 2016.
74. Wen, H.Y. and E. Brogi, *Lobular Carcinoma In Situ.* Surg Pathol Clin, 2018. **11**(1): p. 123-145.
75. Pieri, A., J. Harvey, and N. Bundred, *Pleomorphic lobular carcinoma in situ of the breast: Can the evidence guide practice?* World J Clin Oncol, 2014. **5**(3): p. 546-53.
76. Bagaria, S.P., et al., *The florid subtype of lobular carcinoma in situ: marker or precursor for invasive lobular carcinoma?* Ann Surg Oncol, 2011. **18**(7): p. 1845-51.
77. Al-Baimani, K., et al., *Invasive Pleomorphic Lobular Carcinoma of the Breast: Pathologic, Clinical, and Therapeutic Considerations.* Clin Breast Cancer, 2015. **15**(6): p. 421-5.
78. Lakhani S, E.I., Schnitt SJ, et al., *WHO Classification of Tumours 5th Edition: Breast Tumours. Lyon: International Agency for Research on Cancer(IARC)*. 2018.
79. Porter, A.J., et al., *Mammographic and ultrasound features of invasive lobular carcinoma of the breast.* J Med Imaging Radiat Oncol, 2014. **58**(1): p. 1-10.
80. Iorfida, M., et al., *Invasive lobular breast cancer: subtypes and outcome.* Breast Cancer Res Treat, 2012. **133**(2): p. 713-23.
81. Wheeler, D.T., et al., *Tubulolobular carcinoma of the breast: an analysis of 27 cases of a tumor with a hybrid morphology and immunoprofile.* Am J Surg Pathol, 2004. **28**(12): p. 1587-93.
82. Vlug, E., et al., *Lobular breast cancer: pathology, biology, and options for clinical intervention.* Arch Immunol Ther Exp (Warsz), 2014. **62**(1): p. 7-21.
83. Grabenstetter, A., et al., *E-cadherin immunohistochemical expression in invasive lobular carcinoma of the breast: correlation with morphology and CDH1 somatic alterations.* Hum Pathol, 2020. **102**: p. 44-53.
84. Pai, K., P. Baliga, and B.L. Shrestha, *E-cadherin expression: a diagnostic utility for differentiating breast carcinomas with ductal and lobular morphologies.* J Clin Diagn Res, 2013. **7**(5): p. 840-4.

85. Sarrio, D., et al., *Epigenetic and genetic alterations of APC and CDH1 genes in lobular breast cancer: relationships with abnormal E-cadherin and catenin expression and microsatellite instability.* Int J Cancer, 2003. **106**(2): p. 208-15.
86. Rakha, E.A., et al., *Clinical and biological significance of E-cadherin protein expression in invasive lobular carcinoma of the breast.* Am J Surg Pathol, 2010. **34**(10): p. 1472-9.
87. Vermeulen, J.F., et al., *Nuclear Kaiso expression is associated with high grade and triple-negative invasive breast cancer.* PLoS One, 2012. **7**(5): p. e37864.
88. Da Silva, L., et al., *Aberrant expression of E-cadherin in lobular carcinomas of the breast.* Am J Surg Pathol, 2008. **32**(5): p. 773-83.
89. Dabbs, D.J., R. Bhargava, and M. Chivukula, *Lobular versus ductal breast neoplasms: the diagnostic utility of p120 catenin.* Am J Surg Pathol, 2007. **31**(3): p. 427-37.
90. Goldhirsch, A., et al., *Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011.* Ann Oncol, 2011. **22**(8): p. 1736-47.
91. Rakha, E.A., J.S. Reis-Filho, and I.O. Ellis, *Basal-like breast cancer: a critical review.* J Clin Oncol, 2008. **26**(15): p. 2568-81.
92. Rakha, E.A., et al., *Invasive lobular carcinoma of the breast: response to hormonal therapy and outcomes.* Eur J Cancer, 2008. **44**(1): p. 73-83.
93. Scabia, V., et al., *Estrogen receptor positive breast cancers have patient specific hormone sensitivities and rely on progesterone receptor.* Nat Commun, 2022. **13**(1): p. 3127.
94. DeMayo, F.J., et al., *Mechanisms of action of estrogen and progesterone.* Ann N Y Acad Sci, 2002. **955**: p. 48-59; discussion 86-8, 396-406.
95. Slamon, D.J., et al., *Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene.* Science, 1987. **235**(4785): p. 177-82.
96. Slamon, D.J., et al., *Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer.* Science, 1989. **244**(4905): p. 707-12.
97. Schlam, I. and S.M. Swain, *HER2-positive breast cancer and tyrosine kinase inhibitors: the time is now.* NPJ Breast Cancer, 2021. **7**(1): p. 56.
98. van Maaren, M.C., et al., *Ten-year recurrence rates for breast cancer subtypes in the Netherlands: A large population-based study.* Int J Cancer, 2019. **144**(2): p. 263-272.
99. Marczyk, V.R., et al., *Overall Survival for HER2-Positive Breast Cancer Patients in the HER2-Targeted Era: Evidence From a Population-Based Study.* Clin Breast Cancer, 2022. **22**(5): p. 418-423.
100. Riedlinger, G.M., et al., *Targetable alterations in invasive pleomorphic lobular carcinoma of the breast.* Breast Cancer Res, 2021. **23**(1): p. 7.
101. Rosa-Rosa, J.M., et al., *High Frequency of ERBB2 Activating Mutations in Invasive Lobular Breast Carcinoma with Pleomorphic Features.* Cancers (Basel), 2019. **11**(1).
102. Lien, H.C., et al., *Frequent alterations of HER2 through mutation, amplification, or overexpression in pleomorphic lobular carcinoma of the breast.* Breast Cancer Res Treat, 2015. **150**(2): p. 447-55.
103. Zhu, S., et al., *IRS2 mutations linked to invasion in pleomorphic invasive lobular carcinoma.* JCI Insight, 2018. **3**(8).
104. Yarden, Y. and M.X. Sliwkowski, *Untangling the ErbB signalling network.* Nat Rev Mol Cell Biol, 2001. **2**(2): p. 127-37.
105. Spector, N.L. and K.L. Blackwell, *Understanding the mechanisms behind trastuzumab therapy for human epidermal growth factor receptor 2-positive breast cancer.* J Clin Oncol, 2009. **27**(34): p. 5838-47.
106. Mitri, Z., T. Constantine, and R. O'Regan, *The HER2 Receptor in Breast Cancer: Pathophysiology, Clinical Use, and New Advances in Therapy.* Chemother Res Pract, 2012. **2012**: p. 743193.
107. Bose, R., et al., *Activating HER2 mutations in HER2 gene amplification negative breast cancer.* Cancer Discov, 2013. **3**(2): p. 224-37.

108. Connell, C.M. and G.J. Doherty, *Activating HER2 mutations as emerging targets in multiple solid cancers.* ESMO Open, 2017. **2**(5): p. e000279.
109. Gupta, G.K., et al., *Perspectives on Triple-Negative Breast Cancer: Current Treatment Strategies, Unmet Needs, and Potential Targets for Future Therapies.* Cancers (Basel), 2020. **12**(9).
110. Thike, A.A., et al., *Triple-negative breast cancer: clinicopathological characteristics and relationship with basal-like breast cancer.* Mod Pathol, 2010. **23**(1): p. 123-33.
111. Dent, R., et al., *Triple-negative breast cancer: clinical features and patterns of recurrence.* Clin Cancer Res, 2007. **13**(15 Pt 1): p. 4429-34.
112. Lehmann, B.D., et al., *Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies.* J Clin Invest, 2011. **121**(7): p. 2750-67.
113. Jiang, Y.Z., et al., *Genomic and Transcriptomic Landscape of Triple-Negative Breast Cancers: Subtypes and Treatment Strategies.* Cancer Cell, 2019. **35**(3): p. 428-440 e5.
114. Burstein, M.D., et al., *Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer.* Clin Cancer Res, 2015. **21**(7): p. 1688-98.
115. Plasilova, M.L., et al., *Features of triple-negative breast cancer: Analysis of 38,813 cases from the national cancer database.* Medicine (Baltimore), 2016. **95**(35): p. e4614.
116. Zhao, S., et al., *Clinicopathologic features and prognoses of different histologic types of triple-negative breast cancer: A large population-based analysis.* Eur J Surg Oncol, 2018. **44**(4): p. 420-428.
117. Thompson, K.J., et al., *Luminal androgen receptor breast cancer subtype and investigation of the microenvironment and neoadjuvant chemotherapy response.* NAR Cancer, 2022. **4**(2): p. zcac018.
118. Butler, D. and M. Rosa, *Pleomorphic lobular carcinoma of the breast: a morphologically and clinically distinct variant of lobular carcinoma.* Arch Pathol Lab Med, 2013. **137**(11): p. 1688-92.
119. Jacobs, M., F. Fan, and O. Tawfik, *Clinicopathologic and biomarker analysis of invasive pleomorphic lobular carcinoma as compared with invasive classic lobular carcinoma: an experience in our institution and review of the literature.* Ann Diagn Pathol, 2012. **16**(3): p. 185-9.
120. Jung, H.N., et al., *Are the imaging features of the pleomorphic variant of invasive lobular carcinoma different from classic ILC of the breast?* Breast, 2013. **22**(3): p. 324-9.
121. DiCostanzo, D., et al., *Prognosis in infiltrating lobular carcinoma. An analysis of "classical" and variant tumors.* Am J Surg Pathol, 1990. **14**(1): p. 12-23.
122. Cleton-Jansen, A.M., *E-cadherin and loss of heterozygosity at chromosome 16 in breast carcinogenesis: different genetic pathways in ductal and lobular breast cancer?* Breast Cancer Res, 2002. **4**(1): p. 5-8.
123. Reis-Filho, J.S., et al., *Pleomorphic lobular carcinoma of the breast: role of comprehensive molecular pathology in characterization of an entity.* J Pathol, 2005. **207**(1): p. 1-13.
124. Simpson, P.T., et al., *Molecular profiling pleomorphic lobular carcinomas of the breast: evidence for a common molecular genetic pathway with classic lobular carcinomas.* J Pathol, 2008. **215**(3): p. 231-44.
125. Buchanan, C.L., et al., *Is pleomorphic lobular carcinoma really a distinct clinical entity?* J Surg Oncol, 2008. **98**(5): p. 314-7.
126. Narendra, S., et al., *Clinical outcome in pleomorphic lobular carcinoma: a case-control study with comparison to classic invasive lobular carcinoma.* Ann Diagn Pathol, 2015. **19**(2): p. 64-9.
127. Liu, Y.L., et al., *Invasive Lobular Breast Carcinoma: Pleomorphic Versus Classical Subtype, Associations and Prognosis.* Clin Breast Cancer, 2018. **18**(2): p. 114-120.

128. Rakha, E.A., et al., *Pleomorphic lobular carcinoma of the breast: is it a prognostically significant pathological subtype independent of histological grade?* Mod Pathol, 2013. **26**(4): p. 496-501.
129. Mukhtar, R.A., et al., *Mitotic score and pleomorphic histology in invasive lobular carcinoma of the breast: impact on disease-free survival.* Breast Cancer Res Treat, 2020. **181**(1): p. 23-29.
130. Rakha, E.A., et al., *Histologic grading is an independent prognostic factor in invasive lobular carcinoma of the breast.* Breast Cancer Res Treat, 2008. **111**(1): p. 121-7.
131. Talman, M.L., M.B. Jensen, and F. Rank, *Invasive lobular breast cancer. Prognostic significance of histological malignancy grading.* Acta Oncol, 2007. **46**(6): p. 803-9.
132. Fodor, J., et al., *Comparison of mastectomy with breast-conserving surgery in invasive lobular carcinoma: 15-Year results.* Rep Pract Oncol Radiother, 2011. **16**(6): p. 227-31.
133. Abel, M.K., et al., *Breast conservation therapy versus mastectomy in the surgical management of invasive lobular carcinoma measuring 4 cm or greater.* Am J Surg, 2021. **221**(1): p. 32-36.
134. Piper, M.L., et al., *Success rates of re-excision after positive margins for invasive lobular carcinoma of the breast.* NPJ Breast Cancer, 2019. **5**: p. 29.
135. Jackson, S.P. and J. Bartek, *The DNA-damage response in human biology and disease.* Nature, 2009. **461**(7267): p. 1071-8.
136. Borrego-Soto, G., R. Ortiz-Lopez, and A. Rojas-Martinez, *Ionizing radiation-induced DNA injury and damage detection in patients with breast cancer.* Genet Mol Biol, 2015. **38**(4): p. 420-32.
137. Jutzy, J.M.S., et al., *The Evolution of Radiation Therapy in Metastatic Breast Cancer: From Local Therapy to Systemic Agent.* Int J Breast Cancer, 2018. **2018**: p. 4786819.
138. Vo, T.N., et al., *Outcomes of breast-conservation therapy for invasive lobular carcinoma are equivalent to those for invasive ductal carcinoma.* Am J Surg, 2006. **192**(4): p. 552-5.
139. Stecklein, S.R., X. Shen, and M.P. Mitchell, *Post-Mastectomy Radiation Therapy for Invasive Lobular Carcinoma: A Comparative Utilization and Outcomes Study.* Clin Breast Cancer, 2016. **16**(4): p. 319-26.
140. Cardoso, F., et al., *Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.* Ann Oncol, 2019. **30**(10): p. 1674.
141. Mills, M.N., et al., *Increased Risk for Ipsilateral Breast Tumor Recurrence in Invasive Lobular Carcinoma after Accelerated Partial Breast Irradiation Brachytherapy.* Oncologist, 2021. **26**(11): p. e1931-e1938.
142. Murray Brunt, A., et al., *Hypofractionated breast radiotherapy for 1 week versus 3 weeks (FAST-Forward): 5-year efficacy and late normal tissue effects results from a multicentre, non-inferiority, randomised, phase 3 trial.* Lancet, 2020. **395**(10237): p. 1613-1626.
143. Haviland, J.S., et al., *The UK Standardisation of Breast Radiotherapy (START) trials of radiotherapy hypofractionation for treatment of early breast cancer: 10-year follow-up results of two randomised controlled trials.* Lancet Oncol, 2013. **14**(11): p. 1086-1094.
144. Amjad, M.T., A. Chidharla, and A. Kasi, *Cancer Chemotherapy*, in *StatPearls*. 2023: Treasure Island (FL).
145. Alba, E., et al., *High Proliferation Predicts Pathological Complete Response to Neoadjuvant Chemotherapy in Early Breast Cancer.* Oncologist, 2016. **21**(2): p. 150-5.
146. Telli, M.L., et al., *Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer.* Clin Cancer Res, 2016. **22**(15): p. 3764-73.
147. Ni Chonghaile, T., et al., *Pretreatment mitochondrial priming correlates with clinical response to cytotoxic chemotherapy.* Science, 2011. **334**(6059): p. 1129-33.

148. Sinicrope, F.A., *DNA mismatch repair and adjuvant chemotherapy in sporadic colon cancer.* Nat Rev Clin Oncol, 2010. **7**(3): p. 174-7.
149. Bracci, L., et al., *Immune-based mechanisms of cytotoxic chemotherapy: implications for the design of novel and rationale-based combined treatments against cancer.* Cell Death Differ, 2014. **21**(1): p. 15-25.
150. Park, Y.H., et al., *Chemotherapy induces dynamic immune responses in breast cancers that impact treatment outcome.* Nat Commun, 2020. **11**(1): p. 6175.
151. Charfare, H., S. Limongelli, and A.D. Purushotham, *Neoadjuvant chemotherapy in breast cancer.* Br J Surg, 2005. **92**(1): p. 14-23.
152. Delpech, Y., et al., *Clinical benefit from neoadjuvant chemotherapy in oestrogen receptor-positive invasive ductal and lobular carcinomas.* Br J Cancer, 2013. **108**(2): p. 285-91.
153. Mathieu, M.C., et al., *The poor responsiveness of infiltrating lobular breast carcinomas to neoadjuvant chemotherapy can be explained by their biological profile.* Eur J Cancer, 2004. **40**(3): p. 342-51.
154. Lips, E.H., et al., *Lobular histology and response to neoadjuvant chemotherapy in invasive breast cancer.* Breast Cancer Res Treat, 2012. **136**(1): p. 35-43.
155. Tubiana-Hulin, M., et al., *Response to neoadjuvant chemotherapy in lobular and ductal breast carcinomas: a retrospective study on 860 patients from one institution.* Ann Oncol, 2006. **17**(8): p. 1228-33.
156. Tamirisa, N., et al., *The impact of chemotherapy sequence on survival in node-positive invasive lobular carcinoma.* J Surg Oncol, 2019. **120**(2): p. 132-141.
157. Abotaleb, M., et al., *Chemotherapeutic agents for the treatment of metastatic breast cancer: An update.* Biomed Pharmacother, 2018. **101**: p. 458-477.
158. Truin, W., et al., *Effect of adjuvant chemotherapy in postmenopausal patients with invasive ductal versus lobular breast cancer.* Ann Oncol, 2012. **23**(11): p. 2859-2865.
159. Nurgalieva, Z., C.C. Liu, and X.L. Du, *Chemotherapy use and risk of bone marrow suppression in a large population-based cohort of older women with breast and ovarian cancer.* Med Oncol, 2011. **28**(3): p. 716-25.
160. Mayer, E.L., *Early and late long-term effects of adjuvant chemotherapy.* Am Soc Clin Oncol Educ Book, 2013: p. 9-14.
161. Barrios, C.H., et al., *What is the role of chemotherapy in estrogen receptor-positive, advanced breast cancer?* Ann Oncol, 2009. **20**(7): p. 1157-62.
162. Lips, E.H., et al., *Neoadjuvant chemotherapy in ER+ HER2- breast cancer: response prediction based on immunohistochemical and molecular characteristics.* Breast Cancer Res Treat, 2012. **131**(3): p. 827-36.
163. Harb, W.A., *Management of patients with hormone receptor-positive breast cancer with visceral disease: challenges and treatment options.* Cancer Manag Res, 2015. **7**: p. 37-46.
164. Patel, R., et al., *An emerging generation of endocrine therapies in breast cancer: a clinical perspective.* NPJ Breast Cancer, 2023. **9**(1): p. 20.
165. Dutertre, M. and C.L. Smith, *Molecular mechanisms of selective estrogen receptor modulator (SERM) action.* J Pharmacol Exp Ther, 2000. **295**(2): p. 431-7.
166. Artero, A., J.J. Tarin, and A. Cano, *The adverse effects of estrogen and selective estrogen receptor modulators on hemostasis and thrombosis.* Semin Thromb Hemost, 2012. **38**(8): p. 797-807.
167. Wardell, S.E., J.R. Marks, and D.P. McDonnell, *The turnover of estrogen receptor alpha by the selective estrogen receptor degrader (SERD) fulvestrant is a saturable process that is not required for antagonist efficacy.* Biochem Pharmacol, 2011. **82**(2): p. 122-30.
168. Osborne, C.K., A. Wakeling, and R.I. Nicholson, *Fulvestrant: an oestrogen receptor antagonist with a novel mechanism of action.* Br J Cancer, 2004. **90 Suppl 1**(Suppl 1): p. S2-6.
169. Lloyd, M.R., et al., *Next-generation selective estrogen receptor degraders and other novel endocrine therapies for management of metastatic hormone receptor-positive*

breast cancer: current and emerging role. Ther Adv Med Oncol, 2022. **14**: p. 17588359221113694.
170. *Phase 3 Trial of Elacestrant vs. Standard of Care for the Treatment of Patients With ER+/HER2- Advanced Breast Cancer*.
171. Brueggemeier, R.W., *Aromatase inhibitors--mechanisms of steroidal inhibitors.* Breast Cancer Res Treat, 1994. **30**(1): p. 31-42.
172. Howell, A., et al., *Results of the ATAC (Arimidex, Tamoxifen, Alone or in Combination) trial after completion of 5 years' adjuvant treatment for breast cancer.* Lancet, 2005. **365**(9453): p. 60-2.
173. Razavi, P., et al., *The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers.* Cancer Cell, 2018. **34**(3): p. 427-438 e6.
174. Figg, W.D., 2nd, K. Cook, and R. Clarke, *Aromatase inhibitor plus ovarian suppression as adjuvant therapy in premenopausal women with breast cancer.* Cancer Biol Ther, 2014. **15**(12): p. 1586-7.
175. Dixon, J.M., et al., *Invasive lobular carcinoma: response to neoadjuvant letrozole therapy.* Breast Cancer Res Treat, 2011. **130**(3): p. 871-7.
176. *Fulvestrant and/or Anastrozole in Treating Postmenopausal Patients With Stage II-III Breast Cancer Undergoing Surgery*.
177. Early Breast Cancer Trialists' Collaborative, G., et al., *Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials.* Lancet, 2011. **378**(9793): p. 771-84.
178. Fisher, B., et al., *A randomized clinical trial evaluating tamoxifen in the treatment of patients with node-negative breast cancer who have estrogen-receptor-positive tumors.* N Engl J Med, 1989. **320**(8): p. 479-84.
179. Krauss, K. and E. Stickeler, *Endocrine Therapy in Early Breast Cancer.* Breast Care (Basel), 2020. **15**(4): p. 337-346.
180. Francis, P.A., et al., *Tailoring Adjuvant Endocrine Therapy for Premenopausal Breast Cancer.* N Engl J Med, 2018. **379**(2): p. 122-137.
181. David Cameron; Jack Cuzick; Meredith Regan; Mitch Dowsett; Ivana Sestak; Jonas Bergh; Sandra M Swain; John Bartlett; Early Breast Cancer Trialists' Collaborative Group, R.K.H.S.O.O.M.D.D.H.P.J.B.R.G.R.P.R.B.E.S.R.B.D.R., *Effectiveness of aromatase inhibitors versus tamoxifen in lobular compared to ductal carcinoma: Individual patient data meta-analysis of 9328 women with central histopathology, and 7654 women with e-Cadherin status.* Cancer Res, 2022.
182. *Suppression of Ovarian Function With Either Tamoxifen or Exemestane Compared With Tamoxifen Alone in Treating Premenopausal Women With Hormone-Responsive Breast Cancer*.
183. *Triptorelin With Either Exemestane or Tamoxifen in Treating Premenopausal Women With Hormone-Responsive Breast Cancer*.
184. Baudino, T.A., *Targeted Cancer Therapy: The Next Generation of Cancer Treatment.* Curr Drug Discov Technol, 2015. **12**(1): p. 3-20.
185. Watt, A.C. and S. Goel, *Cellular mechanisms underlying response and resistance to CDK4/6 inhibitors in the treatment of hormone receptor-positive breast cancer.* Breast Cancer Res, 2022. **24**(1): p. 17.
186. Scott, S.C., S.S. Lee, and J. Abraham, *Mechanisms of therapeutic CDK4/6 inhibition in breast cancer.* Semin Oncol, 2017. **44**(6): p. 385-394.
187. Crozier, L., et al., *CDK4/6 inhibitors induce replication stress to cause long-term cell cycle withdrawal.* EMBO J, 2022. **41**(6): p. e108599.
188. Finn, R.S., et al., *Palbociclib and Letrozole in Advanced Breast Cancer.* N Engl J Med, 2016. **375**(20): p. 1925-1936.
189. Goetz, M.P., et al., *MONARCH 3: Abemaciclib As Initial Therapy for Advanced Breast Cancer.* J Clin Oncol, 2017. **35**(32): p. 3638-3646.
190. Rugo, H.S., et al., *Palbociclib plus letrozole as first-line therapy in estrogen receptor-positive/human epidermal growth factor receptor 2-negative advanced breast cancer with extended follow-up.* Breast Cancer Res Treat, 2019. **174**(3): p. 719-729.

191. Cristofanilli, M., et al., *Fulvestrant plus palbociclib versus fulvestrant plus placebo for treatment of hormone-receptor-positive, HER2-negative metastatic breast cancer that progressed on previous endocrine therapy (PALOMA-3): final analysis of the multicentre, double-blind, phase 3 randomised controlled trial.* Lancet Oncol, 2016. **17**(4): p. 425-439.
192. Turner, N.C., et al., *Overall Survival with Palbociclib and Fulvestrant in Advanced Breast Cancer.* N Engl J Med, 2018. **379**(20): p. 1926-1936.
193. Sledge, G.W., Jr., et al., *The Effect of Abemaciclib Plus Fulvestrant on Overall Survival in Hormone Receptor-Positive, ERBB2-Negative Breast Cancer That Progressed on Endocrine Therapy-MONARCH 2: A Randomized Clinical Trial.* JAMA Oncol, 2020. **6**(1): p. 116-124.
194. Dickler, M.N., et al., *MONARCH 1, A Phase II Study of Abemaciclib, a CDK4 and CDK6 Inhibitor, as a Single Agent, in Patients with Refractory HR(+)/HER2(-) Metastatic Breast Cancer.* Clin Cancer Res, 2017. **23**(17): p. 5218-5224.
195. Slamon, D.J., et al., *Overall Survival with Ribociclib plus Fulvestrant in Advanced Breast Cancer.* N Engl J Med, 2020. **382**(6): p. 514-524.
196. Im, S.A., et al., *Overall Survival with Ribociclib plus Endocrine Therapy in Breast Cancer.* N Engl J Med, 2019. **381**(4): p. 307-316.
197. Tripathy, D., et al., *Ribociclib plus endocrine therapy for premenopausal women with hormone-receptor-positive, advanced breast cancer (MONALEESA-7): a randomised phase 3 trial.* Lancet Oncol, 2018. **19**(7): p. 904-915.
198. Neven, P., et al., *Abemaciclib plus fulvestrant in hormone receptor-positive, human epidermal growth factor receptor 2-negative advanced breast cancer in premenopausal women: subgroup analysis from the MONARCH 2 trial.* Breast Cancer Res, 2021. **23**(1): p. 87.
199. Di Cosimo, S., et al., *Palbociclib with Fulvestrant or Letrozole in Endocrine-Sensitive Patients with HR-Positive/HER2-Negative Advanced Breast Cancer: A Detailed Safety Analysis of the Randomized PARSIFAL Trial.* Oncologist, 2023. **28**(1): p. 23-32.
200. Tolaney, S.M., et al., *Abemaciclib plus trastuzumab with or without fulvestrant versus trastuzumab plus standard-of-care chemotherapy in women with hormone receptor-positive, HER2-positive advanced breast cancer (monarcHER): a randomised, open-label, phase 2 trial.* Lancet Oncol, 2020. **21**(6): p. 763-775.
201. O'Sullivan, C.C., V.J. Suman, and M.P. Goetz, *The emerging role of CDK4/6i in HER2-positive breast cancer.* Ther Adv Med Oncol, 2019. **11**: p. 1758835919887665.
202. Johnston, S.R.D. and V. Andre, *Abemaciclib plus endocrine therapy for hormone receptor-positive, HER2-negative, node-positive, high-risk, early breast cancer - Authors' reply.* Lancet Oncol, 2023. **24**(3): p. e104.
203. Loibl, S., et al., *Palbociclib for Residual High-Risk Invasive HR-Positive and HER2-Negative Early Breast Cancer-The Penelope-B Trial.* J Clin Oncol, 2021. **39**(14): p. 1518-1530.
204. Gnant, M., et al., *Adjuvant Palbociclib for Early Breast Cancer: The PALLAS Trial Results (ABCSG-42/AFT-05/BIG-14-03).* J Clin Oncol, 2022. **40**(3): p. 282-293.
205. Gao, J.J., et al., *CDK4/6 inhibitor treatment for patients with hormone receptor-positive, HER2-negative, advanced or metastatic breast cancer: a US Food and Drug Administration pooled analysis.* Lancet Oncol, 2020. **21**(2): p. 250-260.
206. Gao, J.J., et al., *Overall survival in patients with hormone receptor-positive, HER2-negative, advanced or metastatic breast cancer treated with a cyclin-dependent kinase 4/6 inhibitor plus fulvestrant: a US Food and Drug Administration pooled analysis.* Lancet Oncol, 2021. **22**(11): p. 1573-1581.
207. Wilson, F.R., et al., *Herceptin(R) (trastuzumab) in HER2-positive early breast cancer: protocol for a systematic review and cumulative network meta-analysis.* Syst Rev, 2017. **6**(1): p. 196.

208. Valabrega, G., F. Montemurro, and M. Aglietta, *Trastuzumab: mechanism of action, resistance and future perspectives in HER2-overexpressing breast cancer.* Ann Oncol, 2007. **18**(6): p. 977-84.
209. Collins, D.M., et al., *Trastuzumab induces antibody-dependent cell-mediated cytotoxicity (ADCC) in HER-2-non-amplified breast cancer cell lines.* Ann Oncol, 2012. **23**(7): p. 1788-95.
210. Ferraro, E., J.Z. Drago, and S. Modi, *Implementing antibody-drug conjugates (ADCs) in HER2-positive breast cancer: state of the art and future directions.* Breast Cancer Res, 2021. **23**(1): p. 84.
211. Battisti, N.M.L., et al., *Safety and efficacy of T-DM1 in patients with advanced HER2-positive breast cancer The Royal Marsden experience.* Cancer Treat Res Commun, 2020. **24**: p. 100188.
212. Modi, S., et al., *Trastuzumab Deruxtecan in Previously Treated HER2-Low Advanced Breast Cancer.* N Engl J Med, 2022. **387**(1): p. 9-20.
213. Schroeder, R.L., C.L. Stevens, and J. Sridhar, *Small molecule tyrosine kinase inhibitors of ErbB2/HER2/Neu in the treatment of aggressive breast cancer.* Molecules, 2014. **19**(9): p. 15196-212.
214. Chan, A., et al., *Final Efficacy Results of Neratinib in HER2-positive Hormone Receptor-positive Early-stage Breast Cancer From the Phase III ExteNET Trial.* Clin Breast Cancer, 2021. **21**(1): p. 80-91 e7.
215. Cameron, D., et al., *A phase III randomized comparison of lapatinib plus capecitabine versus capecitabine alone in women with advanced breast cancer that has progressed on trastuzumab: updated efficacy and biomarker analyses.* Breast Cancer Res Treat, 2008. **112**(3): p. 533-43.
216. Guarneri, V., et al., *Trastuzumab-lapatinib as neoadjuvant therapy for HER2-positive early breast cancer: Survival analyses of the CHER-Lob trial.* Eur J Cancer, 2021. **153**: p. 133-141.
217. Murthy, R.K., et al., *Tucatinib, Trastuzumab, and Capecitabine for HER2-Positive Metastatic Breast Cancer.* N Engl J Med, 2020. **382**(7): p. 597-609.
218. Ma, F., et al., *Pyrotinib or Lapatinib Combined With Capecitabine in HER2-Positive Metastatic Breast Cancer With Prior Taxanes, Anthracyclines, and/or Trastuzumab: A Randomized, Phase II Study.* J Clin Oncol, 2019. **37**(29): p. 2610-2619.
219. Wu, J., et al., *Neoadjuvant pyrotinib, trastuzumab, and docetaxel for HER2-positive breast cancer (PHEDRA): a double-blind, randomized phase 3 trial.* BMC Med, 2022. **20**(1): p. 498.
220. Saura, C., et al., *Neratinib Plus Capecitabine Versus Lapatinib Plus Capecitabine in HER2-Positive Metastatic Breast Cancer Previously Treated With >/= 2 HER2-Directed Regimens: Phase III NALA Trial.* J Clin Oncol, 2020. **38**(27): p. 3138-3149.
221. Pohlmann, P.R., I.A. Mayer, and R. Mernaugh, *Resistance to Trastuzumab in Breast Cancer.* Clin Cancer Res, 2009. **15**(24): p. 7479-7491.
222. Goutsouliak, K., et al., *Towards personalized treatment for early stage HER2-positive breast cancer.* Nat Rev Clin Oncol, 2020. **17**(4): p. 233-250.
223. Metzger-Filho, O., et al., *Magnitude of trastuzumab benefit in patients with HER2-positive, invasive lobular breast carcinoma: results from the HERA trial.* J Clin Oncol, 2013. **31**(16): p. 1954-60.
224. Piccart-Gebhart, M.J., et al., *Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer.* N Engl J Med, 2005. **353**(16): p. 1659-72.
225. Cameron, D., et al., *11 years' follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive early breast cancer: final analysis of the HERceptin Adjuvant (HERA) trial.* Lancet, 2017. **389**(10075): p. 1195-1205.
226. Miricescu, D., et al., *PI3K/AKT/mTOR Signaling Pathway in Breast Cancer: From Molecular Landscape to Clinical Aspects.* Int J Mol Sci, 2020. **22**(1).
227. Ortega, M.A., et al., *Signal Transduction Pathways in Breast Cancer: The Important Role of PI3K/Akt/mTOR.* J Oncol, 2020. **2020**: p. 9258396.

228. Paplomata, E. and R. O'Regan, *The PI3K/AKT/mTOR pathway in breast cancer: targets, trials and biomarkers.* Ther Adv Med Oncol, 2014. **6**(4): p. 154-66.
229. Mollon, L.E., et al., *A Systematic Literature Review of the Prognostic and Predictive Value of PIK3CA Mutations in HR(+)/HER2(-) Metastatic Breast Cancer.* Clin Breast Cancer, 2020. **20**(3): p. e232-e243.
230. Andre, F., et al., *Alpelisib for PIK3CA-Mutated, Hormone Receptor-Positive Advanced Breast Cancer.* N Engl J Med, 2019. **380**(20): p. 1929-1940.
231. Howell, S.J., et al., *Fulvestrant plus capivasertib versus placebo after relapse or progression on an aromatase inhibitor in metastatic, oestrogen receptor-positive, HER2-negative breast cancer (FAKTION): overall survival, updated progression-free survival, and expanded biomarker analysis from a randomised, phase 2 trial.* Lancet Oncol, 2022. **23**(7): p. 851-864.
232. Teo, K., et al., *E-cadherin loss induces targetable autocrine activation of growth factor signalling in lobular breast cancer.* Sci Rep, 2018. **8**(1): p. 15454.
233. Noguchi, S., et al., *Efficacy of everolimus with exemestane versus exemestane alone in Asian patients with HER2-negative, hormone-receptor-positive breast cancer in BOLERO-2.* Breast Cancer, 2014. **21**(6): p. 703-14.
234. Gabriel N. Hortobagyi , S.N., Patrick Neven , Puttisak Puttawibul , Daniel Heng , Thomas BrechenmacherFrancois Philippe Ringeisen , Stephen Saletan , Thomas Denis Bachelot, *Everolimus plus exemestane in patients with advanced invasive lobular carcinoma: Efficacy and safety results from BOLERO-2.*

. Journal of Clinical Oncology 32, no. 26_suppl (September 10, 2014) 152-152., 2014.
235. Cavallo, F., et al., *2011: the immune hallmarks of cancer.* Cancer Immunol Immunother, 2011. **60**(3): p. 319-26.
236. Dillman, R.O., *Cancer immunotherapy.* Cancer Biother Radiopharm, 2011. **26**(1): p. 1-64.
237. Yadav, D., et al., *Cancer immunotherapy by immune checkpoint blockade and its advanced application using bio-nanomaterials.* Semin Cancer Biol, 2022. **86**(Pt 2): p. 909-922.
238. Iranzo, P., et al., *Overview of Checkpoint Inhibitors Mechanism of Action: Role of Immune-Related Adverse Events and Their Treatment on Progression of Underlying Cancer.* Front Med (Lausanne), 2022. **9**: p. 875974.
239. Liu, Y., et al., *Regulation of arginase I activity and expression by both PD-1 and CTLA-4 on the myeloid-derived suppressor cells.* Cancer Immunol Immunother, 2009. **58**(5): p. 687-97.
240. Chikuma, S., et al., *PD-1-mediated suppression of IL-2 production induces CD8+ T cell anergy in vivo.* J Immunol, 2009. **182**(11): p. 6682-9.
241. Youngblood, B., et al., *Chronic virus infection enforces demethylation of the locus that encodes PD-1 in antigen-specific CD8(+) T cells.* Immunity, 2011. **35**(3): p. 400-12.
242. Yokosuka, T., et al., *Programmed cell death 1 forms negative costimulatory microclusters that directly inhibit T cell receptor signaling by recruiting phosphatase SHP2.* J Exp Med, 2012. **209**(6): p. 1201-17.
243. Hui, E., et al., *T cell costimulatory receptor CD28 is a primary target for PD-1-mediated inhibition.* Science, 2017. **355**(6332): p. 1428-1433.
244. Adams, S., et al., *Pembrolizumab monotherapy for previously untreated, PD-L1-positive, metastatic triple-negative breast cancer: cohort B of the phase II KEYNOTE-086 study.* Ann Oncol, 2019. **30**(3): p. 405-411.
245. Adams, S., et al., *Pembrolizumab monotherapy for previously treated metastatic triple-negative breast cancer: cohort A of the phase II KEYNOTE-086 study.* Ann Oncol, 2019. **30**(3): p. 397-404.
246. Nanda, R., et al., *Pembrolizumab in Patients With Advanced Triple-Negative Breast Cancer: Phase Ib KEYNOTE-012 Study.* J Clin Oncol, 2016. **34**(21): p. 2460-7.

247. Cortes, J., et al., *Pembrolizumab plus Chemotherapy in Advanced Triple-Negative Breast Cancer.* N Engl J Med, 2022. **387**(3): p. 217-226.
248. Jacob, S.L., L.A. Huppert, and H.S. Rugo, *Role of Immunotherapy in Breast Cancer.* JCO Oncol Pract, 2023: p. OP2200483.
249. Schmid, P., et al., *Pembrolizumab for Early Triple-Negative Breast Cancer.* N Engl J Med, 2020. **382**(9): p. 810-821.
250. Loi, S., et al., *Pembrolizumab plus trastuzumab in trastuzumab-resistant, advanced, HER2-positive breast cancer (PANACEA): a single-arm, multicentre, phase 1b-2 trial.* Lancet Oncol, 2019. **20**(3): p. 371-382.
251. Ahn, H.K., et al., *Response Rate and Safety of a Neoadjuvant Pertuzumab, Atezolizumab, Docetaxel, and Trastuzumab Regimen for Patients With ERBB2-Positive Stage II/III Breast Cancer: The Neo-PATH Phase 2 Nonrandomized Clinical Trial.* JAMA Oncol, 2022. **8**(9): p. 1271-1277.
252. Emens, L.A., et al., *Trastuzumab emtansine plus atezolizumab versus trastuzumab emtansine plus placebo in previously treated, HER2-positive advanced breast cancer (KATE2): a phase 2, multicentre, randomised, double-blind trial.* Lancet Oncol, 2020. **21**(10): p. 1283-1295.
253. Huober, J., et al., *Atezolizumab With Neoadjuvant Anti-Human Epidermal Growth Factor Receptor 2 Therapy and Chemotherapy in Human Epidermal Growth Factor Receptor 2-Positive Early Breast Cancer: Primary Results of the Randomized Phase III IMpassion050 Trial.* J Clin Oncol, 2022. **40**(25): p. 2946-2956.
254. *A Study Evaluating the Efficacy and Safety of Adjuvant Atezolizumab or Placebo and Trastuzumab Emtansine for Participants With HER2-Positive Breast Cancer at High Risk of Recurrence Following Preoperative Therapy*.
255. *Neoadjuvant Treatment of HER2 Positive Early High-risk and Locally Advanced Breast Cancer*.
256. Rugo, H.S., et al., *Safety and Antitumor Activity of Pembrolizumab in Patients with Estrogen Receptor-Positive/Human Epidermal Growth Factor Receptor 2-Negative Advanced Breast Cancer.* Clin Cancer Res, 2018. **24**(12): p. 2804-2811.
257. *Study of Nivolumab Versus Placebo in Combination With Neoadjuvant Chemotherapy and Adjuvant Endocrine Therapy in Participants With High-risk, Estrogen Receptor-Positive (ER+), Human Epidermal Growth Factor Receptor 2-Negative (HER2-) Primary Breast Cancer*.
258. *Study of Pembrolizumab (MK-3475) Plus Chemotherapy Versus Placebo Plus Chemotherapy for HR+/HER2- Locally Recurrent Inoperable or Metastatic Breast Cancer (MK-3475-B49/KEYNOTE-B49)*.
259. *Study of Pembrolizumab (MK-3475) Versus Placebo in Combination With Neoadjuvant Chemotherapy & Adjuvant Endocrine Therapy in the Treatment of Early-Stage Estrogen Receptor-Positive, Human Epidermal Growth Factor Receptor 2-Negative (ER+/HER2-) Breast Cancer (MK-3475-756/KEYNOTE-756)*.
260. *AssessinG Efficacy of Carboplatin and ATezOlizumab in Metastatic Lobular Breast Cancer*. https://ClinicalTrials.gov/show/NCT03147040.
261. Mittendorf, E.A., et al., *Neoadjuvant atezolizumab in combination with sequential nab-paclitaxel and anthracycline-based chemotherapy versus placebo and chemotherapy in patients with early-stage triple-negative breast cancer (IMpassion031): a randomised, double-blind, phase 3 trial.* Lancet, 2020. **396**(10257): p. 1090-1100.
262. Schmid, P., et al., *Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer.* N Engl J Med, 2018. **379**(22): p. 2108-2121.
263. Presti, D., et al., *Tumor infiltrating lymphocytes (TILs) as a predictive biomarker of response to checkpoint blockers in solid tumors: A systematic review.* Crit Rev Oncol Hematol, 2022. **177**: p. 103773.
264. McNamara, M.G., et al., *Impact of high tumor mutational burden in solid tumors and challenges for biomarker application.* Cancer Treat Rev, 2020. **89**: p. 102084.

265. Marabelle, A., et al., *Efficacy of Pembrolizumab in Patients With Noncolorectal High Microsatellite Instability/Mismatch Repair-Deficient Cancer: Results From the Phase II KEYNOTE-158 Study.* J Clin Oncol, 2020. **38**(1): p. 1-10.
266. Nan, Z., et al., *The Predictive Efficacy of Tumor Mutation Burden (TMB) on Nonsmall Cell Lung Cancer Treated by Immune Checkpoint Inhibitors: A Systematic Review and Meta-Analysis.* Biomed Res Int, 2021. **2021**: p. 1780860.
267. Ning, B., et al., *The Predictive Value of Tumor Mutation Burden on Clinical Efficacy of Immune Checkpoint Inhibitors in Melanoma: A Systematic Review and Meta-Analysis.* Front Pharmacol, 2022. **13**: p. 748674.
268. Iwase, T., et al., *A Novel Immunomodulatory 27-Gene Signature to Predict Response to Neoadjuvant Immunochemotherapy for Primary Triple-Negative Breast Cancer.* Cancers (Basel), 2021. **13**(19).
269. Bajrami, I., et al., *E-Cadherin/ROS1 Inhibitor Synthetic Lethality in Breast Cancer.* Cancer Discov, 2018. **8**(4): p. 498-515.
270. *Crizotinib in Lobular Breast, Diffuse Gastric and Triple Negative Lobular Breast Cancer or CDH1-mutated Solid Tumours*. https://ClinicalTrials.gov/show/NCT03620643.
271. *Neoadjuvant Study of Targeting ROS1 in Combination With Endocrine Therapy in Invasive Lobular Carcinoma of the Breast (ROSALINE)*. https://ClinicalTrials.gov/show/NCT04551495.
272. *Endocrine Response in Women With Invasive Lobular Breast Cancer*. https://ClinicalTrials.gov/show/NCT02206984.
273. *Palbociclib and Endocrine Therapy for LObular Breast Cancer Preoperative Study (PELOPS)*. https://ClinicalTrials.gov/show/NCT02764541.
274. *Basket Study of Neratinib in Participants With Solid Tumors Harboring Somatic HER2 or EGFR Exon 18 Mutations*.
275. Fernández-Martos;, H.W.V.B.C.S.M.O.K.J.H.W.F.-C.B.A.M.B.M.E.B.A.C.C., *Neratinib + trastuzumab + fulvestrant for HER2-mutant, hormone receptor-positive, metastatic breast cancer: Updated results from the phase 2 SUMMIT 'basket' trial* Cancer Research 2020.
276. Desmedt, C., et al., *Genomic Characterization of Primary Invasive Lobular Breast Cancer.* J Clin Oncol, 2016. **34**(16): p. 1872-81.
277. Michaut, M., et al., *Integration of genomic, transcriptomic and proteomic data identifies two biologically distinct subtypes of invasive lobular breast cancer.* Sci Rep, 2016. **6**: p. 18517.
278. Dabbs, D.J., et al., *The spectrum of morphomolecular abnormalities of the E-cadherin/catenin complex in pleomorphic lobular carcinoma of the breast.* Appl Immunohistochem Mol Morphol, 2007. **15**(3): p. 260-6.
279. Droufakou, S., et al., *Multiple ways of silencing E-cadherin gene expression in lobular carcinoma of the breast.* Int J Cancer, 2001. **92**(3): p. 404-8.
280. Alexander, J., et al., *Assessment of the Molecular Heterogeneity of E-Cadherin Expression in Invasive Lobular Breast Cancer.* Cancers (Basel), 2022. **14**(2).
281. Hansford, S., et al., *Hereditary Diffuse Gastric Cancer Syndrome: CDH1 Mutations and Beyond.* JAMA Oncol, 2015. **1**(1): p. 23-32.
282. Desmedt, C., et al., *Immune Infiltration in Invasive Lobular Breast Cancer.* J Natl Cancer Inst, 2018. **110**(7): p. 768-776.
283. Hurtado, A., et al., *FOXA1 is a key determinant of estrogen receptor function and endocrine response.* Nat Genet, 2011. **43**(1): p. 27-33.
284. Theodorou, V., et al., *GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility.* Genome Res, 2013. **23**(1): p. 12-22.
285. Ping, Z., et al., *ERBB2 mutation is associated with a worse prognosis in patients with CDH1 altered invasive lobular cancer of the breast.* Oncotarget, 2016. **7**(49): p. 80655-80663.
286. McCart Reed, A.E., et al., *LobSig is a multigene predictor of outcome in invasive lobular carcinoma.* NPJ Breast Cancer, 2019. **5**: p. 18.

287. Lee, J.Y., et al., *Lobular Carcinomas In Situ Display Intralesion Genetic Heterogeneity and Clonal Evolution in the Progression to Invasive Lobular Carcinoma.* Clin Cancer Res, 2019. **25**(2): p. 674-686.
288. Harrison, B.T., et al., *Genomic profiling of pleomorphic and florid lobular carcinoma in situ reveals highly recurrent ERBB2 and ERRB3 alterations.* Mod Pathol, 2020. **33**(7): p. 1287-1297.
289. Shamir, E.R., Y.Y. Chen, and G. Krings, *Genetic analysis of pleomorphic and florid lobular carcinoma in situ variants: frequent ERBB2/ERBB3 alterations and clonal relationship to classic lobular carcinoma in situ and invasive lobular carcinoma.* Mod Pathol, 2020. **33**(6): p. 1078-1091.
290. Richard, F., et al., *Characterization of Stromal Tumor-infiltrating Lymphocytes and Genomic Alterations in Metastatic Lobular Breast Cancer.* Clin Cancer Res, 2020. **26**(23): p. 6254-6265.
291. Pareja, F., et al., *The genomic landscape of metastatic histologic special types of invasive breast cancer.* NPJ Breast Cancer, 2020. **6**: p. 53.
292. Sokol, E.S., et al., *Loss of function of NF1 is a mechanism of acquired resistance to endocrine therapy in lobular breast cancer.* Ann Oncol, 2019. **30**(1): p. 115-123.
293. Ercan, C., et al., *p53 mutations in classic and pleomorphic invasive lobular carcinoma of the breast.* Cell Oncol (Dordr), 2012. **35**(2): p. 111-8.
294. Deniziaut, G., et al., *ERBB2 mutations associated with solid variant of high-grade invasive lobular breast carcinomas.* Oncotarget, 2016. **7**(45): p. 73337-73346.
295. Desmedt, C., et al., *ESR1 mutations in metastatic lobular breast cancer patients.* NPJ Breast Cancer, 2019. **5**: p. 9.
296. Cao, L., et al., *Frequent amplifications of ESR1, ERBB2 and MDM4 in primary invasive lobular breast carcinoma.* Cancer Lett, 2019. **461**: p. 21-30.
297. Basudan, A., et al., *Frequent ESR1 and CDK Pathway Copy-Number Alterations in Metastatic Breast Cancer.* Mol Cancer Res, 2019. **17**(2): p. 457-468.
298. Rakha, E.A., et al., *Nottingham Prognostic Index Plus (NPI+): a modern clinical decision making tool in breast cancer.* Br J Cancer, 2014. **110**(7): p. 1688-97.
299. Metzger-Filho, O., et al., *Genomic grade adds prognostic value in invasive lobular carcinoma.* Ann Oncol, 2013. **24**(2): p. 377-384.
300. Beumer, I.J., et al., *Prognostic Value of MammaPrint((R)) in Invasive Lobular Breast Cancer.* Biomark Insights, 2016. **11**: p. 139-146.
301. Laenkholm, A.V., et al., *Population-based Study of Prosigna-PAM50 and Outcome Among Postmenopausal Women With Estrogen Receptor-positive and HER2-negative Operable Invasive Lobular or Ductal Breast Cancer.* Clin Breast Cancer, 2020. **20**(4): p. e423-e432.
302. Filipits, M., et al., *A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors.* Clin Cancer Res, 2011. **17**(18): p. 6012-20.
303. Sestak, I., et al., *Prognostic Value of EndoPredict in Women with Hormone Receptor-Positive, HER2-Negative Invasive Lobular Breast Cancer.* Clin Cancer Res, 2020. **26**(17): p. 4682-4687.
304. Bomeisl, P.E., et al., *Comparison of Oncotype DX Recurrence Score by Histologic Types of Breast Carcinoma.* Arch Pathol Lab Med, 2015. **139**(12): p. 1546-9.
305. Felts, J.L., et al., *An Analysis of Oncotype DX Recurrence Scores and Clinicopathologic Characteristics in Invasive Lobular Breast Cancer.* Breast J, 2017. **23**(6): p. 677-686.
306. Wilson, P.C., et al., *Breast cancer histopathology is predictive of low-risk Oncotype Dx recurrence score.* Breast J, 2018. **24**(6): p. 976-980.
307. Anderson, N.M. and M.C. Simon, *The tumor microenvironment.* Curr Biol, 2020. **30**(16): p. R921-R925.
308. Salgado, R., 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): p. 259-71.

309. Dvorak, H.F., *Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing.* N Engl J Med, 1986. **315**(26): p. 1650-9.
310. Hu, D., et al., *Cancer-associated fibroblasts in breast cancer: Challenges and opportunities.* Cancer Commun (Lond), 2022. **42**(5): p. 401-434.
311. Nakagawa, S., et al., *Tumor microenvironment in invasive lobular carcinoma: possible therapeutic targets.* Breast Cancer Res Treat, 2016. **155**(1): p. 65-75.
312. Gomez-Cuadrado, L., et al., *Characterisation of the Stromal Microenvironment in Lobular Breast Cancer.* Cancers (Basel), 2022. **14**(4).
313. Dushyanthen, S., et al., *Relevance of tumor-infiltrating lymphocytes in breast cancer.* BMC Med, 2015. **13**: p. 202.
314. Emens, L.A., *Breast Cancer Immunotherapy: Facts and Hopes.* Clin Cancer Res, 2018. **24**(3): p. 511-520.
315. Alexander, E.L. and B. Wetzel, *Human lymphocytes: similarity of B and T cell surface morphology.* Science, 1975. **188**(4189): p. 732-4.
316. Sauls, R.S., C. McCausland, and B.N. Taylor, *Histology, T-Cell Lymphocyte*, in *StatPearls*. 2023: Treasure Island (FL).
317. Horii, M. and T. Matsushita, *Regulatory B cells and T cell Regulation in Cancer.* J Mol Biol, 2021. **433**(1): p. 166685.
318. Fabbri, M., C. Smart, and R. Pardi, *T lymphocytes.* Int J Biochem Cell Biol, 2003. **35**(7): p. 1004-8.
319. Raskov, H., et al., *Cytotoxic CD8(+) T cells in cancer and cancer immunotherapy.* Br J Cancer, 2021. **124**(2): p. 359-367.
320. Andersen, M.H., et al., *Cytotoxic T cells.* J Invest Dermatol, 2006. **126**(1): p. 32-41.
321. Samji, T. and K.M. Khanna, *Understanding memory CD8(+) T cells.* Immunol Lett, 2017. **185**: p. 32-39.
322. Tay, R.E., E.K. Richardson, and H.C. Toh, *Revisiting the role of CD4(+) T cells in cancer immunotherapy-new insights into old paradigms.* Cancer Gene Ther, 2021. **28**(1-2): p. 5-17.
323. Speiser, D.E., et al., *CD4(+) T cells in cancer.* Nat Cancer, 2023. **4**(3): p. 317-329.
324. Whiteside, T.L., *The role of regulatory T cells in cancer immunology.* Immunotargets Ther, 2015. **4**: p. 159-71.
325. Jiang, D., et al., *Clinicopathological and prognostic significance of FOXP3+ tumor infiltrating lymphocytes in patients with breast cancer: a meta-analysis.* BMC Cancer, 2015. **15**: p. 727.
326. Wang, Y., et al., *Regulatory T cells are an important prognostic factor in breast cancer: a systematic review and meta-analysis.* Neoplasma, 2016. **63**(5): p. 789-98.
327. Zhou, Y., et al., *Prognostic value of tumor-infiltrating Foxp3+ regulatory T cells in patients with breast cancer: a meta-analysis.* J Cancer, 2017. **8**(19): p. 4098-4105.
328. Liu, S., et al., *Prognostic significance of FOXP3+ tumor-infiltrating lymphocytes in breast cancer depends on estrogen receptor and human epidermal growth factor receptor-2 expression status and concurrent cytotoxic T-cell infiltration.* Breast Cancer Res, 2014. **16**(5): p. 432.
329. Mahmoud, S.M., et al., *An evaluation of the clinical significance of FOXP3+ infiltrating cells in human breast cancer.* Breast Cancer Res Treat, 2011. **127**(1): p. 99-108.
330. Dijkgraaf, E.M., et al., *Chemotherapy alters monocyte differentiation to favor generation of cancer-supporting M2 macrophages in the tumor microenvironment.* Cancer Res, 2013. **73**(8): p. 2480-92.
331. Fujiwara, N. and K. Kobayashi, *Macrophages in inflammation.* Curr Drug Targets Inflamm Allergy, 2005. **4**(3): p. 281-6.
332. Boutilier, A.J. and S.F. Elsawa, *Macrophage Polarization States in the Tumor Microenvironment.* Int J Mol Sci, 2021. **22**(13).
333. Allavena, P., et al., *The inflammatory micro-environment in tumor progression: the role of tumor-associated macrophages.* Crit Rev Oncol Hematol, 2008. **66**(1): p. 1-9.

334. Mahmoud, S.M., et al., *Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer.* J Clin Oncol, 2011. **29**(15): p. 1949-55.
335. Gao, G., et al., *Prognostic value of tumor-infiltrating lymphocytes in patients with triple-negative breast cancer: a systematic review and meta-analysis.* BMC Cancer, 2020. **20**(1): p. 179.
336. Asano, Y., et al., *Prediction of Treatment Response to Neoadjuvant Chemotherapy in Breast Cancer by Subtype Using Tumor-infiltrating Lymphocytes.* Anticancer Res, 2018. **38**(4): p. 2311-2321.
337. Denkert, C., et al., *Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer.* J Clin Oncol, 2010. **28**(1): p. 105-13.
338. 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, 2018. **19**(1): p. 40-50.
339. Tille, J.C., et al., *Tumor-infiltrating lymphocytes are associated with poor prognosis in invasive lobular breast carcinoma.* Mod Pathol, 2020. **33**(11): p. 2198-2207.
340. Onkar, S., et al., *Immune landscape in invasive ductal and lobular breast cancer reveals a divergent macrophage-driven microenvironment.* Nat Cancer, 2023. **4**(4): p. 516-534.
341. Bankhead, P., et al., *QuPath: Open source software for digital pathology image analysis.* Sci Rep, 2017. **7**(1): p. 16878.
342. Uwe Schmidt, M.W., Coleman Broaddus, and Gene Myers, *Cell Detection with Star-convex Polygons.* 2018.
343. Chen, X., et al., *Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications.* Bioinformatics, 2016. **32**(8): p. 1220-2.
344. Ye, K., et al., *Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads.* Bioinformatics, 2009. **25**(21): p. 2865-71.
345. Cibulskis, K., et al., *Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples.* Nat Biotechnol, 2013. **31**(3): p. 213-9.
346. Fernandez, E.M., et al., *Cancer-Specific Thresholds Adjust for Whole Exome Sequencing-based Tumor Mutational Burden Distribution.* JCO Precis Oncol, 2019. **3**.
347. Wingett, S.W. and S. Andrews, *FastQ Screen: A tool for multi-genome mapping and quality control.* F1000Res, 2018. **7**: p. 1338.
348. Ewels, P., et al., *MultiQC: summarize analysis results for multiple tools and samples in a single report.* Bioinformatics, 2016. **32**(19): p. 3047-8.
349. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.
350. Aken, B.L., et al., *The Ensembl gene annotation system.* Database (Oxford), 2016. **2016**.
351. Wongvibulsin, S., K.C. Wu, and S.L. Zeger, *Clinical risk prediction with random forests for survival, longitudinal, and multivariate (RF-SLAM) data analysis.* BMC Med Res Methodol, 2019. **20**(1): p. 1.
352. Saal, L.H., et al., *The Sweden Cancerome Analysis Network - Breast (SCAN-B) Initiative: a large-scale multicenter infrastructure towards implementation of breast cancer genomic analyses in the clinical routine.* Genome Med, 2015. **7**(1): p. 20.
353. Curtis, C., et al., *The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups.* Nature, 2012. **486**(7403): p. 346-52.
354. Anghel, C.V., et al., *ISOpureR: an R implementation of a computational purification algorithm of mixed tumour profiles.* BMC Bioinformatics, 2015. **16**: p. 156.
355. Jimenez-Sanchez, A., O. Cast, and M.L. Miller, *Comprehensive Benchmarking and Integration of Tumor Microenvironment Cell Estimation Methods.* Cancer Res, 2019. **79**(24): p. 6238-6246.

356. Sflomos, G., et al., *Intraductal xenografts show lobular carcinoma cells rely on their own extracellular matrix and LOXL1.* EMBO Mol Med, 2021. **13**(3): p. e13180.
357. Baxter, E.W., et al., *Standardized protocols for differentiation of THP-1 cells to macrophages with distinct M(IFNgamma+LPS), M(IL-4) and M(IL-10) phenotypes.* J Immunol Methods, 2020. **478**: p. 112721.
358. Zheng, G.X., et al., *Massively parallel digital transcriptional profiling of single cells.* Nat Commun, 2017. **8**: p. 14049.
359. Stuart, T., et al., *Comprehensive Integration of Single-Cell Data.* Cell, 2019. **177**(7): p. 1888-1902 e21.
360. Hafemeister, C. and R. Satija, *Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression.* Genome Biol, 2019. **20**(1): p. 296.
361. Traag, V.A., *Faster unfolding of communities: speeding up the Louvain algorithm.* Phys Rev E Stat Nonlin Soft Matter Phys, 2015. **92**(3): p. 032801.
362. Hu, Z., A.A. Ahmed, and C. Yau, *CIDER: an interpretable meta-clustering framework for single-cell RNA-seq data integration and evaluation.* Genome Biol, 2021. **22**(1): p. 337.
363. Nguyen, H.C.T., et al., *Benchmarking integration of single-cell differential expression.* Nat Commun, 2023. **14**(1): p. 1570.
364. Marot, G., et al., *Moderated effect size and P-value combinations for microarray meta-analyses.* Bioinformatics, 2009. **25**(20): p. 2692-9.
365. Tizhoosh, H.R., et al., *Searching Images for Consensus: Can AI Remove Observer Variability in Pathology?* Am J Pathol, 2021. **191**(10): p. 1702-1708.
366. Weydert, J.A., B.R. De Young, and M.B. Cohen, *A preliminary diagnosis service provides prospective blinded dual-review of all general surgical pathology cases in an academic practice.* Am J Surg Pathol, 2005. **29**(6): p. 801-5.
367. Weir, M.M., E. Jan, and T.J. Colgan, *Interinstitutional pathology consultations. A reassessment.* Am J Clin Pathol, 2003. **120**(3): p. 405-12.
368. Renshaw, A.A. and E.W. Gould, *Comparison of disagreement and amendment rates by tissue type and diagnosis: identifying cases for directed blinded review.* Am J Clin Pathol, 2006. **126**(5): p. 736-9.
369. Tomaszewski, J.E. and V.A. LiVolsi, *Mandatory second opinion of pathologic slides: is it necessary?* Cancer, 1999. **86**(11): p. 2198-200.
370. Schaumann, N., et al., *Lobular neoplasia and invasive lobular breast cancer: Inter-observer agreement for histological grading and subclassification.* Pathol Res Pract, 2019. **215**(11): p. 152611.
371. Hermsen, M., et al., *Deep Learning-Based Histopathologic Assessment of Kidney Tissue.* J Am Soc Nephrol, 2019. **30**(10): p. 1968-1979.
372. Ehteshami Bejnordi, B., et al., *Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer.* JAMA, 2017. **318**(22): p. 2199-2210.
373. Litjens, G., et al., *A survey on deep learning in medical image analysis.* Med Image Anal, 2017. **42**: p. 60-88.
374. Rutgers, J.J., et al., *Interobserver variability between experienced and inexperienced observers in the histopathological analysis of Wilms tumors: a pilot study for future algorithmic approach.* Diagn Pathol, 2021. **16**(1): p. 77.
375. Campanella, G., et al., *Clinical-grade computational pathology using weakly supervised deep learning on whole slide images.* Nat Med, 2019. **25**(8): p. 1301-1309.
376. Chen, E.Y., et al., *Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool.* BMC Bioinformatics, 2013. **14**: p. 128.
377. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.

378. Morris, L.G., et al., *Recurrent somatic mutation of FAT1 in multiple human cancers leads to aberrant Wnt activation.* Nat Genet, 2013. **45**(3): p. 253-61.
379. Li, Z., et al., *Loss of the FAT1 Tumor Suppressor Promotes Resistance to CDK4/6 Inhibitors via the Hippo Pathway.* Cancer Cell, 2018. **34**(6): p. 893-905 e8.
380. Liang, S., et al., *CDK12: A Potent Target and Biomarker for Human Cancer Therapy.* Cells, 2020. **9**(6).
381. Choi, H.J., et al., *CDK12 drives breast tumor initiation and trastuzumab resistance via WNT and IRS1-ErbB-PI3K signaling.* EMBO Rep, 2019. **20**(10): p. e48058.
382. Naidoo, K., et al., *Evaluation of CDK12 Protein Expression as a Potential Novel Biomarker for DNA Damage Response-Targeted Therapies in Breast Cancer.* Mol Cancer Ther, 2018. **17**(1): p. 306-315.
383. Quereda, V., et al., *Therapeutic Targeting of CDK12/CDK13 in Triple-Negative Breast Cancer.* Cancer Cell, 2019. **36**(5): p. 545-558 e7.
384. Arriola, E., et al., *Genomic analysis of the HER2/TOP2A amplicon in breast cancer and breast cancer cell lines.* Lab Invest, 2008. **88**(5): p. 491-503.
385. Sircoulomb, F., et al., *Genome profiling of ERBB2-amplified breast cancers.* BMC Cancer, 2010. **10**: p. 539.
386. [cited 2023 01.09.2023]; Available from: https://www.genecards.org/cgi-bin/carddisp.pl?gene=GCNT2.
387. Elbauomy Elsheikh, S., et al., *FGFR1 amplification in breast carcinomas: a chromogenic in situ hybridisation analysis.* Breast Cancer Res, 2007. **9**(2): p. R23.
388. !!! INVALID CITATION !!! {}.
389. Drago, J.Z., et al., *FGFR1 Amplification Mediates Endocrine Resistance but Retains TORC Sensitivity in Metastatic Hormone Receptor-Positive (HR(+)) Breast Cancer.* Clin Cancer Res, 2019. **25**(21): p. 6443-6451.
390. Coombes, R.C., et al., *Results of the phase IIa RADICAL trial of the FGFR inhibitor AZD4547 in endocrine resistant breast cancer.* Nat Commun, 2022. **13**(1): p. 3246.
391. Zheng, S., et al., *Overexpression of CBX2 in breast cancer promotes tumor progression through the PI3K/AKT signaling pathway.* Am J Transl Res, 2019. **11**(3): p. 1668-1682.
392. Ahn, B.Y., et al., *Genetic screen identifies insulin-like growth factor binding protein 5 as a modulator of tamoxifen resistance in breast cancer.* Cancer Res, 2010. **70**(8): p. 3013-9.
393. Lim, S.M., M.H. Hong, and H.R. Kim, *Immunotherapy for Non-small Cell Lung Cancer: Current Landscape and Future Perspectives.* Immune Netw, 2020. **20**(1): p. e10.
394. Carreau, N. and A. Pavlick, *Revolutionizing treatment of advanced melanoma with immunotherapy.* Surg Oncol, 2022. **42**: p. 101180.
395. Fasano, M., et al., *Immunotherapy for head and neck cancer: Present and future.* Crit Rev Oncol Hematol, 2022. **174**: p. 103679.
396. Stanton, S.E., S. Adams, and M.L. Disis, *Variation in the Incidence and Magnitude of Tumor-Infiltrating Lymphocytes in Breast Cancer Subtypes: A Systematic Review.* JAMA Oncol, 2016. **2**(10): p. 1354-1360.
397. Sikandar, B., et al., *Increased Tumour Infiltration of CD4+ and CD8+ T-Lymphocytes in Patients with Triple Negative Breast Cancer Suggests Susceptibility to Immune Therapy.* Asian Pac J Cancer Prev, 2017. **18**(7): p. 1827-1832.
398. Narang, P., et al., *The neoepitope landscape of breast cancer: implications for immunotherapy.* BMC Cancer, 2019. **19**(1): p. 200.
399. Safonov, A., et al., *Immune Gene Expression Is Associated with Genomic Aberrations in Breast Cancer.* Cancer Res, 2017. **77**(12): p. 3317-3324.
400. Alexandrov, L.B., et al., *Signatures of mutational processes in human cancer.* Nature, 2013. **500**(7463): p. 415-21.
401. Chalmers, Z.R., et al., *Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden.* Genome Med, 2017. **9**(1): p. 34.

402. Zehir, A., et al., *Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients.* Nat Med, 2017. **23**(6): p. 703-713.
403. Barroso-Sousa, R., et al., *Prevalence and mutational determinants of high tumor mutation burden in breast cancer.* Ann Oncol, 2020. **31**(3): p. 387-394.
404. Haricharan, S., et al., *Somatic mutation load of estrogen receptor-positive breast tumors predicts overall survival: an analysis of genome sequence data.* Breast Cancer Res Treat, 2014. **146**(1): p. 211-20.
405. Shah, S.P., et al., *The clonal and mutational evolution spectrum of primary triple-negative breast cancers.* Nature, 2012. **486**(7403): p. 395-9.
406. 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, 2014. **25**(8): p. 1544-50.
407. De Angelis, C., et al., *Evaluation of the Predictive Role of Tumor Immune Infiltrate in Patients with HER2-Positive Breast Cancer Treated with Neoadjuvant Anti-HER2 Therapy without Chemotherapy.* Clin Cancer Res, 2020. **26**(3): p. 738-745.
408. Ingold Heppner, B., et al., *Tumor-Infiltrating Lymphocytes: A Predictive and Prognostic Biomarker in Neoadjuvant-Treated HER2-Positive Breast Cancer.* Clin Cancer Res, 2016. **22**(23): p. 5747-5754.
409. Brown, S.D., et al., *Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival.* Genome Res, 2014. **24**(5): p. 743-50.
410. Ward, J.P., M.M. Gubin, and R.D. Schreiber, *The Role of Neoantigens in Naturally Occurring and Therapeutically Induced Immune Responses to Cancer.* Adv Immunol, 2016. **130**: p. 25-74.
411. 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, 2013. **31**(7): p. 860-7.
412. Liu, Z., et al., *TP53 Mutations Promote Immunogenic Activity in Breast Cancer.* J Oncol, 2019. **2019**: p. 5952836.
413. Mei, P., et al., *High tumor mutation burden is associated with DNA damage repair gene mutation in breast carcinomas.* Diagn Pathol, 2020. **15**(1): p. 50.
414. Chen, D.S. and I. Mellman, *Oncology meets immunology: the cancer-immunity cycle.* Immunity, 2013. **39**(1): p. 1-10.
415. Gajewski, T.F., H. Schreiber, and Y.X. Fu, *Innate and adaptive immune cells in the tumor microenvironment.* Nat Immunol, 2013. **14**(10): p. 1014-22.
416. Tsujikawa, T., et al., *Prognostic significance of spatial immune profiles in human solid cancers.* Cancer Sci, 2020. **111**(10): p. 3426-3434.
417. Arakelyan, A., et al., *Histocultures (tissue explants) in human retrovirology.* Methods Mol Biol, 2014. **1087**: p. 233-48.
418. Ganguly, D., et al., *The role of dendritic cells in autoimmunity.* Nat Rev Immunol, 2013. **13**(8): p. 566-77.
419. Chen, Q., et al., *CD3(+)CD20(+) T cells and their roles in human diseases.* Hum Immunol, 2019. **80**(3): p. 191-194.
420. Zhu, Y., et al., *STING: a master regulator in the cancer-immunity cycle.* Mol Cancer, 2019. **18**(1): p. 152.
421. Wolf, Y., A.C. Anderson, and V.K. Kuchroo, *TIM3 comes of age as an inhibitory receptor.* Nat Rev Immunol, 2020. **20**(3): p. 173-185.
422. Chistiakov, D.A., et al., *CD68/macrosialin: not just a histochemical marker.* Lab Invest, 2017. **97**(1): p. 4-13.
423. Grant, E.J., et al., *The role of CD27 in anti-viral T-cell immunity.* Curr Opin Virol, 2017. **22**: p. 77-88.
424. Bernier, G.M., *beta 2-Microglobulin: structure, function and significance.* Vox Sang, 1980. **38**(6): p. 323-7.
425. Liao, Y., et al., *Expression pattern of CD11c on lung immune cells after disseminated murine cytomegalovirus infection.* Virol J, 2017. **14**(1): p. 132.

426. Vudattu, N.K., et al., *Reduced numbers of IL-7 receptor (CD127) expressing immune cells and IL-7-signaling defects in peripheral blood from patients with breast cancer.* Int J Cancer, 2007. **121**(7): p. 1512-9.
427. Zanoni, I. and F. Granucci, *Role of CD14 in host protection against infections and in metabolism regulation.* Front Cell Infect Microbiol, 2013. **3**: p. 32.
428. Jordan, A.R., et al., *The Role of CD44 in Disease Pathophysiology and Targeted Treatment.* Front Immunol, 2015. **6**: p. 182.
429. Fortis, S.P., et al., *Differential intratumoral distributions of CD8 and CD163 immune cells as prognostic biomarkers in breast cancer.* J Immunother Cancer, 2017. **5**: p. 39.
430. Barmpoutis, P., et al., *Tertiary lymphoid structures (TLS) identification and density assessment on H&E-stained digital slides of lung cancer.* PLoS One, 2021. **16**(9): p. e0256907.
431. Sflomos, G., et al., *Atlas of Lobular Breast Cancer Models: Challenges and Strategic Directions.* Cancers (Basel), 2021. **13**(21).
432. *GeoMx® Human Whole Transcriptome Atlas*. [cited 2023; Available from: https://nanostring.com/products/geomx-digital-spatial-profiler/geomx-rna-assays/geomx-whole-transcriptome-atlas/.
433. Holness, C.L. and D.L. Simmons, *Molecular cloning of CD68, a human macrophage marker related to lysosomal glycoproteins.* Blood, 1993. **81**(6): p. 1607-13.
434. Yang, J., et al., *The role of tumor-associated macrophages in breast carcinoma invasion and metastasis.* Int J Clin Exp Pathol, 2015. **8**(6): p. 6656-64.
435. Jeong, H., et al., *Tumor-Associated Macrophages as Potential Prognostic Biomarkers of Invasive Breast Cancer.* J Breast Cancer, 2019. **22**(1): p. 38-51.
436. Szpor, J., et al., *Dendritic Cells Are Associated with Prognosis and Survival in Breast Cancer.* Diagnostics (Basel), 2021. **11**(4).
437. Talmadge, J.E., M. Donkor, and E. Scholar, *Inflammatory cell infiltration of tumors: Jekyll or Hyde.* Cancer Metastasis Rev, 2007. **26**(3-4): p. 373-400.
438. Jansen, M.P., et al., *HOXB13-to-IL17BR expression ratio is related with tumor aggressiveness and response to tamoxifen of recurrent breast cancer: a retrospective study.* J Clin Oncol, 2007. **25**(6): p. 662-8.
439. Ma, X.J., et al., *A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen.* Cancer Cell, 2004. **5**(6): p. 607-16.
440. Jerevall, P.L., et al., *Predictive relevance of HOXB13 protein expression for tamoxifen benefit in breast cancer.* Breast Cancer Res, 2010. **12**(4): p. R53.
441. Shah, N., et al., *HOXB13 mediates tamoxifen resistance and invasiveness in human breast cancer by suppressing ERalpha and inducing IL-6 expression.* Cancer Res, 2013. **73**(17): p. 5449-58.
442. Neo, S.Y. and A. Lundqvist, *The Multifaceted Roles of CXCL9 Within the Tumor Microenvironment.* Adv Exp Med Biol, 2020. **1231**: p. 45-51.
443. Liu, J., et al., *Multi-Omics Analyses Revealed GOLT1B as a Potential Prognostic Gene in Breast Cancer Probably Regulating the Immune Microenvironment.* Front Oncol, 2021. **11**: p. 805273.
444. Caiado, F., B. Silva-Santos, and H. Norell, *Intra-tumour heterogeneity - going beyond genetics.* FEBS J, 2016. **283**(12): p. 2245-58.
445. Dagogo-Jack, I. and A.T. Shaw, *Tumour heterogeneity and resistance to cancer therapies.* Nat Rev Clin Oncol, 2018. **15**(2): p. 81-94.
446. Nik-Zainal, S., et al., *The life history of 21 breast cancers.* Cell, 2012. **149**(5): p. 994-1007.
447. Yates, L.R., et al., *Subclonal diversification of primary breast cancer revealed by multiregion sequencing.* Nat Med, 2015. **21**(7): p. 751-9.
448. Yates, L.R., et al., *Genomic Evolution of Breast Cancer Metastasis and Relapse.* Cancer Cell, 2017. **32**(2): p. 169-184 e7.

449. Brastianos, P.K., et al., *Genomic Characterization of Brain Metastases Reveals Branched Evolution and Potential Therapeutic Targets.* Cancer Discov, 2015. **5**(11): p. 1164-1177.
450. De Mattos-Arruda, L., et al., *Capturing intra-tumor genetic heterogeneity by de novo mutation profiling of circulating cell-free tumor DNA: a proof-of-principle.* Ann Oncol, 2018. **29**(11): p. 2268.
451. Ding, L., et al., *Genome remodelling in a basal-like breast cancer metastasis and xenograft.* Nature, 2010. **464**(7291): p. 999-1005.
452. Hoadley, K.A., et al., *Tumor Evolution in Two Patients with Basal-like Breast Cancer: A Retrospective Genomics Study of Multiple Metastases.* PLoS Med, 2016. **13**(12): p. e1002174.
453. Savas, P., et al., *The Subclonal Architecture of Metastatic Breast Cancer: Results from a Prospective Community-Based Rapid Autopsy Program "CASCADE".* PLoS Med, 2016. **13**(12): p. e1002204.
454. Fleming, J.M., et al., *Local regulation of human breast xenograft models.* J Cell Physiol, 2010. **224**(3): p. 795-806.
455. Sflomos, G., et al., *A Preclinical Model for ERalpha-Positive Breast Cancer Points to the Epithelial Microenvironment as Determinant of Luminal Phenotype and Hormone Response.* Cancer Cell, 2016. **29**(3): p. 407-422.
456. *Cellecta* 2023 [cited 2023 01.09.2023]; Available from: https://cellecta.com/collections/clonetracker-barcode-libraries-and-kits.
457. Bakken, T.E., et al., *Single-nucleus and single-cell transcriptomes compared in matched cortical cell types.* PLoS One, 2018. **13**(12): p. e0209648.
458. Zhang, H., et al., *Engagement of I-branching beta-1, 6-N-acetylglucosaminyltransferase 2 in breast cancer metastasis and TGF-beta signaling.* Cancer Res, 2011. **71**(14): p. 4846-56.
459. Gorchs, L., et al., *Chemokine Receptor Expression on T Cells Is Modulated by CAFs and Chemokines Affect the Spatial Distribution of T Cells in Pancreatic Tumors.* Cancers (Basel), 2022. **14**(15).