Inaugural Dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht - Karls - University Heidelberg

Presented by

M.Sc. Paul Albert Wolfgang Schwerd-Kleine

born in **Würzburg**

Oral examination: **29 November 2024**

Dissecting the Cellular and Molecular Landscape of Inflammatory Breast Cancer through Single-Cell and Spatial Transcriptomics

Referees:

Prof. Dr. Andreas Trumpp

Prof. Dr. Peter Lichter

Declaration

The work presented in this dissertation was performed from November 2019 until November 2024 at the German Cancer Research Center (Deutsches Krebsforschungszentrum, DKFZ, Heidelberg) and the Heidelberg Institute for Stem Cell Technology and Experimental Medicine (HI-STEM gGmbH, Heidelberg) under the supervision of Prof. Dr. Andreas Trumpp and Dr. Martin Sprick.

Declarations according to § 8 (3) b) und c) of the Doctoral Degree Regulations:

(a) I hereby declare that I have written the submitted dissertation myself and, in this process, have used no other sources or materials than those explicitly indicated.

(b) I hereby declare that I have not applied to be examined at any other institution, nor have I used the dissertation in this way or any other form at any other institution as an examination paper nor submitted it to any other faculty as dissertation.

(c) I hereby consent to the verification of the dissertation by means of electronic data processing programs against standing scientific standards.

Abstract

Inflammatory Breast Cancer (IBC) is a distinct and aggressive form of breast cancer with rapid onset and poor prognosis. Despite extensive research, its molecular heterogeneity remains unclear. In this thesis, I present the first comprehensive single-cell and spatial transcriptomic analysis of IBC, comparing IBC to non-IBC patient samples. This approach provided new insights into IBC's unique biology and confirmed key disease hallmarks at the molecular level.

The study highlights the critical role of the tumour microenvironment (TME) in driving IBC's aggressive behaviour. Tumour-associated macrophages (TAMs) and cancer-associated fibroblasts (CAFs) foster an immune-suppressive, pro-angiogenic environment. Dysregulated pathways, including IL-1β expression and NF-κB signalling in myeloid cells, abundant T_{reg} cells and high TGF-β levels, and activated VEGF-PI3K signalling in endothelial cells, suggest potential therapeutic targets within the TME.

By mapping IBC's cellular composition, this study reveals immune suppression and chronic inflammation as key drivers of its aggressiveness. The abundance of suppressive T cells and immune evasion mechanisms suggests that immunotherapy, including immune checkpoint blockade, potentially in combination with TGF-β inhibition, could offer promising treatment options. Targeting key signalling pathways in CAFs and myeloid cells may further disrupt inflammatory feedback loops.

This research advances the understanding of IBC by providing a molecular map of its TME and identifying novel therapeutic avenues. Future studies should focus on validating these targets and developing models to explore combination therapies targeting both the tumour and its microenvironment, offering hope for improved treatment outcomes.

Zusammenfassung

Das inflammatorische Mammakarzinom (IBC) ist eine besondere und aggressive Form des Brustkrebses, die sich schnell entwickelt und eine schlechte Prognose aufweist. Trotz umfangreicher Forschung bleibt die molekulare Heterogenität des IBC unklar. In dieser Arbeit präsentiere ich die erste umfassende Einzelzell- und räumliche Transkriptomanalyse von IBC, wobei IBC mit nicht-IBC-Patientenproben verglichen wird. Dieser Ansatz lieferte neue Erkenntnisse über die einzigartige Biologie des IBC und bestätigte wichtige krankheitsspezifische Merkmale auf molekularer Ebene.

Die Studie unterstreicht die entscheidende Rolle der Tumormikroumgebung (TME) bei der Förderung des aggressiven Verhaltens von IBC. Tumor-assoziierte Makrophagen (TAMs) und krebsassoziierte Fibroblasten (CAFs) schaffen ein immununterdrückendes, proangiogenes Umfeld. Dysregulierte Signalwege, einschließlich der IL-1β-Expression und NF-κB-Signalübertragung in myeloiden Zellen, regulatorischen T-Zellen und aktivierten Endothelzellen, deuten auf potenzielle therapeutische Ziele innerhalb der TME hin.

Durch die Kartierung der zellulären Zusammensetzung des IBC zeigt diese Studie, dass die Immunsuppression und chronische Entzündungen Haupttreiber seiner Aggressivität sein könnten. Die Vielzahl von erschöpften T-Zellen und Mechanismen der Immunflucht legt nahe, dass Immuntherapien, einschließlich der Blockade von Immun-Checkpoints, möglicherweise in Kombination mit der Hemmung von TGF-β, vielversprechende Behandlungsoptionen bieten könnten. Die gezielte Beeinflussung wichtiger Signalwege in CAFs und TAMs könnte zudem entzündliche Rückkopplungsschleifen unterbrechen.

Diese Forschung verbessert das Verständnis von IBC, indem sie eine molekulare Karte seiner TME bereitstellt und neue therapeutische Ansätze identifiziert. Zukünftige Studien sollten sich darauf konzentrieren, diese Ziele zu validieren und Modelle zu entwickeln, um Kombinationstherapien zu erforschen, die sowohl auf den Tumor als auch auf seine Mikroumgebung abzielen, und so Hoffnung auf verbesserte Behandlungsergebnisse bieten.

Contents

1 Introduction **1**

1.1 Epidemiology of Breast Cancer 2

1.1.1 Demographics and Incidence 3

1.1.2 Mortality 3

1.1.3 Genetic Predisposition 4

1.1.4 Lifestyle and Environmental Influences 5

1.2 Subtypes of Breast Cancer 6

1.3 The Microenvironment of Breast Cancer 9

1.4 Breast Cancer Treatment 11

1.4.1 Surgery 12

1.4.2 Radiation Therapy 12

1.4.3 Systemic Therapies 13

1.4.4 Neoadjuvant and Adjuvant Therapy 14

1.4.5 Emerging Therapies 15

1.4.6 Treatment Resistance 16

1.5 Rare Breast Cancers 17

1.6 Pathophysiology of IBC 20

1.7 Management of IBC 21

1.8 Previous Studies of IBC 22

1.9 Single-Cell Transcriptomic Analysis 23

1.9.1 Technology 23

1.9.2 Computational Tools and Analysis Best Practices 25

1.10 Spatial Transcriptomics 27
1.10.1 Technology 28
1.10.2 Applications in Cancer Research 29
1.10.3 Computational Tools and Best Practices 30
1.10.4 Future Directions 32

2 Aim of This Dissertation **33**

3 Results **35**
3.1 IBC sample collection 36
3.2 Single-Cell Library Preparation and Sequencing 37
3.3 Data Pre-Processing 40
3.3.1 De-Multiplexing of Hashed Libraries 40
3.3.2 Background RNA Removal 43
3.3.3 Multiplet Removal 44
3.3.4 Quality Control 45
3.3.5 Preprocessing of Public Datasets 47
3.3.6 Integration of IBC and Non-IBC Datasets 50
3.3.7 Dimensionality Reduction 51
3.3.8 Sample Harmonisation 52
3.3.9 Cell Type Annotation 56
3.3.10 Differentiation of Malignant from Epithelial Cells 59
3.4 Features of Cell Populations in the Microenvironment of IBC 60
3.4.1 Cancer Associated Fibroblasts 61
3.4.2 Myeloid Cells 68
3.4.3 T/NK Cells 76
3.4.4 Endothelial Cells 86
3.5 Spatial Transcriptomics Analysis of IBC 94
3.5.1 Quality Control 95
3.5.2 Cell Type Assignment 96

3.5.3 Neighbourhood Enrichment 100
3.5.4 Spatial Proximity of T Cells and Cancer Cells 102

4 Discussion 105
4.1 Single-Cell and Spatial Transcriptomics Enable a New and More Detailed View on IBC 106
4.2 A Single-Cell Map of IBC 108
4.3 The Single-Cell and Spatial Data Help Understand the IBC Phenotype and Suggests New Therapeutic Avenues 110
4.4 Conclusion 113

5 Materials and Methods 115
5.1 Sample collection 116
5.2 Tissue digestion and biobanking 116
5.3 Single-cell library creation and sequencing 117
5.4 scRNAseq demultiplexing 117
5.5 scRNAseq data preprocessing, harmonisation and cell type assignment 118
5.6 Assignment of malignant cells 119
5.7 Single-cell data manipulation and visualisation 122
5.8 Dataset integration 122
5.9 Quantification of relative cell numbers 123
5.10 Ligand-receptor analysis using scRNA-seq data 123
5.11 Gene set enrichment analysis 125
5.12 Pseudobulk gene expression comparison 125
5.13 Pseudobulk expression heatmap for T/NK cells 126
5.14 Pathway activity analysis 127
5.15 Spatial transcriptomics 127
5.16 Use of Artificial Intelligence for Coding Assistance as well as Thesis Writing, Proofreading, and Typesetting 131

5.17 Graphics Software 131

6 Supplementary Material **133**

6.1 Xenium Gene Panel 134

6.2 List of Abbreviations 137

6.3 List of Figures 140

6.4 List of Tables 143

7 Bibliography **145**

8 Contributions **169**

9 Acknowledgements **173**

Chapter 1

Introduction

Breast cancer is the most frequent malignant tumour in women [1]. It can present in many different clinical, histological, and molecular forms, making it a very heterogeneous disease [2]. This heterogeneity implies that there is no one-size-fits-all approach to treatment [3]. Rather, ideal management needs to be individually tailored for optimal outcomes. Treatment decisions are driven by molecular- [4] or receptor-based suptypes [5], in combination with the tumour stage. Treatment of breast cancer is based on a combination of locoregional and systemic therapy [6, 7]. The former includes surgery and radiation and is especially important to remove the primary tumour before metastasis has occurred, but can be used as a palliative measure as well. Systemic therapy aims to shrink the tumour burden, as well as target dormant cancer cells or overt metastases at other locations of the body. As is the case with other diseases, breast cancer patients are best treated by a multidisciplinary team of physicians at specialised centres, improving measures like outcome and quality of life [8]. Global disparities exist at all levels, from disease prevention, to early detection and optimal care. These need to be addressed in order to achieve equitable outcomes for communities worldwide [9].

1.1 Epidemiology of Breast Cancer

Breast cancer is the most frequently diagnosed malignancy in women, with approximately 2.1 million new cases and over 600,000 deaths in 2018. Additionally, the number of diagnoses has been growing steadily over the past decade, owing to both growth and ageing of the population [10].

The epidemiology of breast cancer has been extensively studied, revealing variations in incidence, mortality, and the impact of genetic and environmental risk factors across different populations [11].

1.1.1 Demographics and Incidence

Breast cancer incidence rates vary widely across regions, influenced by factors such as age, geographic location, socioeconomic status, and access to healthcare [12]. It is most commonly diagnosed in women over the age of 50, though younger women are increasingly affected [13]. Incidence rates are highest in high-income countries, with North America, Australia, and Western Europe exhibiting the greatest rates. In contrast, lower incidence rates are observed in lower income regions such as parts of Africa and Asia, although these regions are experiencing a rising trend as they undergo economic development and the accompanying adaptation of lifestyle factors [14].

Age-standardised rates (ASR) of breast cancer incidence in high-income countries are generally higher due to better screening, healthcare access, and lifestyle factors such as delayed childbearing, lower fertility rates, and the prevalence of hormone replacement therapy (HRT) (See below, Lifestyle and Environmental Influences) [15]. However, improved awareness and screening practices have contributed to early detection, which not only influences incidence statistics but also improves survival outcomes [16, 17]

1.1.2 Mortality

While breast cancer remains a leading cause of cancer-related death among women, mortality rates have been declining in many parts of the world, particularly in high-income countries [18]. This is largely due to advancements in early detection through screening programmes such as mammography [19] and improvements in treatment options, including targeted therapies, chemotherapy, and radiotherapy [20]. Nonetheless, significant disparities in mortality rates persist between high- and low-income countries. In low-resource settings, late-stage diagnosis and limited access to adequate treatment contribute to higher mortality rates [21].

Mortality rates are also influenced by ethnic and racial factors. For example, African American women in the United States have lower incidence rates of breast cancer than white women but experience higher mortality rates, a trend attributed to more aggressive tumour types, limited access to quality healthcare, and socioeconomic inequalities [22]. Similar disparities are observed in other regions, with Indigenous and minority populations often bearing a disproportionate burden of mortality [23].

1.1.3 Genetic Predisposition

Breast cancer risk is influenced by a combination of genetic and environmental factors. Some 5–10% of all breast cancer cases are attributed to inherited genetic mutations. Mutations in the BRCA1 and BRCA2 genes are the most well-known genetic risk factors, significantly increasing the lifetime risk of breast and ovarian cancers [24]. Women with a BRCA1 mutation, for instance, have a 55–65% chance of developing breast cancer by age 70, while the risk for those with a BRCA2 mutation is estimated at around 45%.

Other genetic syndromes associated with elevated breast cancer risk include Li-Fraumeni syndrome (caused by an autosomal dominant mutation in the TP53 gene), Cowden syndrome (autosomal dominant mutations, mostly in PTEN), and Peutz-Jeghers syndrome (autosomal dominant mutation in the STK11 gene) [25].

In addition to these high-risk mutations, mutation screening studies have identified genes that are functionally related to BRCA1 and BRCA2, and confer an intermediate risk when mutated (CHEK2, ATM, BRIP1, and PALB2). Lastly, genome-wide association studies (GWAS) have identified a range of low-penetrance genetic variants that, when combined, may contribute to breast cancer risk in a polygenic manner [26].

1.1.4 Lifestyle and Environmental Influences

Beyond genetics, various lifestyle and environmental factors play a role in breast cancer development (Figure 1.1) [27, 28].

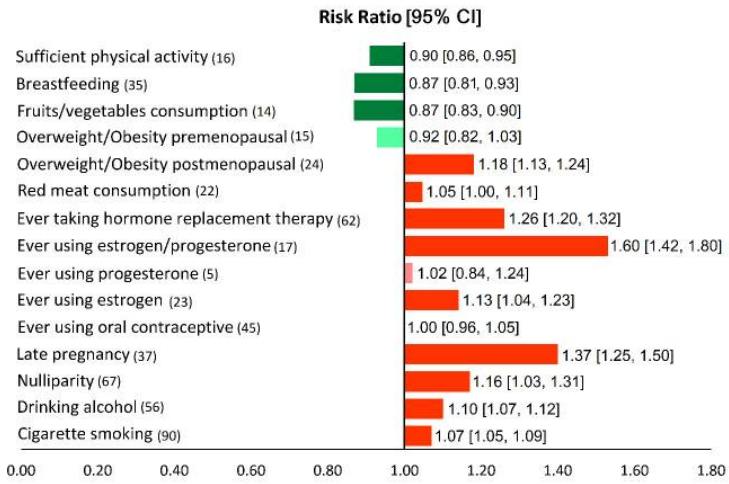


Figure 1.1: Modifiable lifestyle factors for breast cancer. Bar plot of protective factors and risk factors for the development of breast cancer. Original work published by Jalal Poorolajal *et al.* under Creative Commons License [28].

Modifiable risk factors include reproductive history, obesity, alcohol consumption, and physical activity levels [29, 30]. Due to the nature of breast cancer, many of these are related to hormone exposure levels. For instance, early menarche, late menopause, and nulliparity increase a woman's lifetime exposure to oestrogen, which is a known risk factor for breast cancer [31]. Conversely, factors such as breastfeeding and early childbirth have a protective effect by reducing exposure to circulating oestrogen.

Obesity, particularly post-menopausal obesity, is associated with an increased risk of breast cancer, as adipose tissue becomes the primary site of oestrogen production after menopause [32, 33]. Furthermore, excessive alcohol consump-

tion has been shown to raise breast cancer risk by altering oestrogen levels and other hormonal mechanisms [34]. Regular physical activity, on the other hand, is protective and has been associated with a lower risk of breast cancer across numerous studies [35].

Besides exposure to reproductive hormones, environmental factors, such as ionising radiation, particularly during childhood or adolescence, are also linked to an increased breast cancer risk [36]. Additionally, long-term hormone replacement therapy (HRT) with oestrogen-progestin combinations has been found to elevate breast cancer risk, particularly in women who use it for extended periods [37].

In summary, breast cancer epidemiology is shaped by a complex interplay of genetic predisposition, environmental and lifestyle factors. Understanding the patterns of incidence and mortality across different populations, and recognising the impact of modifiable risk factors, is crucial for developing targeted prevention strategies and improving breast cancer outcomes worldwide.

1.2 Subtypes of Breast Cancer

Breast cancer is a heterogeneous disease that can be sub-classified in order to tailor therapy to the individual patient's disease. There are two main approaches for classifying breast cancer, molecular subtypes based on gene-expression profiling and receptor status-based subtypes [38].

Perou, Sørlie and colleagues first reported a classification of breast cancers into intrinsic subtypes in 2000, causing a paradigm shift away from strategies founded solely on tumour burden towards biology-based approaches [4, 2]. Determination of these molecular subtypes is done by gene expression profiling [39], oftentimes based on a panel containing 50 genes, the PAM50 panel [40].

Using this method, breast cancer is typically subdivided into the following **molecular subtypes** [41]:

1. **Luminal A** Luminal A is the most common subtype and generally has the

best prognosis. It is characterised by being ER+ and/or PR+, HER2-, and has a low Ki-67 index, which is a marker of cell proliferation. Luminal A tumours tend to grow slowly and respond well to hormonal therapies. They are more frequently found in postmenopausal women.

2. **Luminal B** Luminal B tumours are also ER+ and/or PR+, but they tend to have a higher Ki-67 index and can be either HER2+ or HER2-. This subtype is more aggressive than Luminal A, with poorer prognosis and higher recurrence rates. Treatment usually involves a combination of hormonal therapies and chemotherapy, and HER2-targeted therapies if applicable.

3. **HER2-Enriched** The HER2-enriched subtype is characterised by high expression of the HER2 gene and the absence of ER and PR. This molecular subtype overlaps with HER2-positive breast cancer defined by receptor status, and includes cases that are particularly aggressive. HER2-enriched breast cancers tend to grow and spread more rapidly but are often highly responsive to HER2-targeted therapies.

4. **Basal-Like** Basal-like breast cancer closely overlaps with triple-negative breast cancer, although not all basal-like cancers are triple-negative. These tumours often express markers associated with the basal cells of the breast, such as cytokeratins 5 and 6. Basal-like breast cancer is typically aggressive and has a poor prognosis, often affecting younger women and those with BRCA1 mutations.

In clinical practice, however, a surrogate of this signature is most commonly used, which is based on histology and immunohistochemistry analysis of expression of some key proteins [42]. These include the hormone receptors (HR) oestrogen receptor (ER), the progesterone receptor (PR), as well as human epidermal growth factor receptor 2 (HER2), and Ki67, a marker expressed in proliferating

cells. Based on these proteins, classification into **receptor status subtypes** is possible [41]:

1. **Hormone Receptor-Positive (ER+ and/or PR+)** This subtype includes breast cancers that express oestrogen or progesterone receptors. These tumours rely on oestrogen and/or progesterone to grow, making them responsive to hormonal therapies such as tamoxifen or aromatase inhibitors. Hormone receptor-positive cancers tend to have a better prognosis and are more common in older women. They often grow more slowly than other subtypes.

2. **HER2-Positive** HER2-positive breast cancer is characterised by the overexpression of the HER2 protein, a receptor involved in cell growth. This subtype accounts for approximately 15-20% of breast cancer cases and is associated with a more aggressive disease course. However, the advent of targeted therapies, such as trastuzumab (Herceptin), has significantly improved outcomes for patients with HER2-positive breast cancer.

3. **Triple-Negative Breast Cancer (TNBC)** Triple-negative breast cancer lacks the expression of ER, PR, and HER2, making it more challenging to treat, as it does not respond to hormonal or HER2-targeted therapies. TNBC is more likely to be aggressive, with a higher likelihood of early metastasis and recurrence. It disproportionately affects younger women and women of African or Hispanic descent. While chemotherapy remains the mainstay of treatment, recent advances in immunotherapy have provided new options for some patients with TNBC.

The classification of breast cancer into these subtypes has profound implications for treatment and prognosis [43]. Hormone receptor-positive and HER2-positive cancers benefit from targeted therapies, which can significantly improve

outcomes [44]. In contrast, for triple-negative and basal-like subtypes, treatment options are more limited, as the main strategies of targeted therapy against ER, PR and HER2 fail in this setting [45]. These often require more aggressive treatments like chemotherapy and they are associated with higher rates of recurrence and mortality.

In recent years, precision medicine has been increasingly integrated into breast cancer care, with genomic testing further refining treatment decisions. Tests such as Oncotype DX and MammaPrint are being used to assess the risk of recurrence in hormone receptor-positive breast cancer, helping to guide the use of chemotherapy in addition to hormonal therapy [46, 47]. Additionally, clinical trials are assessing of biomarker-guided precision medicine [48, 49].

Understanding the heterogeneity of breast cancer subtypes enables clinicians to provide personalised treatments, leading to better outcomes and more effective management of the disease. Ongoing research into the molecular underpinnings of these subtypes continues to drive the development of novel therapies, particularly for aggressive subtypes such as triple-negative and HER2-enriched breast cancers.

1.3 The Microenvironment of Breast Cancer

Breast cancer cells do not exist on their own, but are surrounded by a wide variety of immune and stromal cells. These profoundly influence tumour development, response to therapy and prognosis (Figure 1.2) [2, 50].

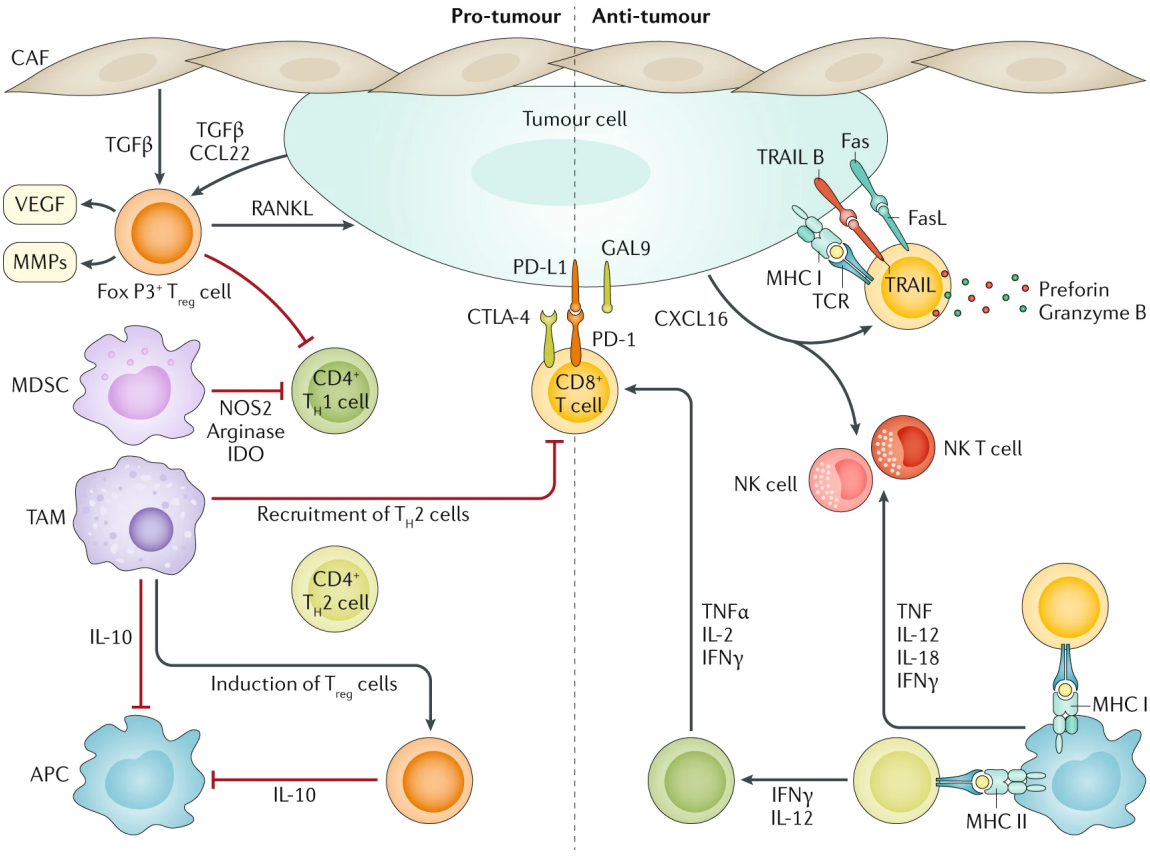


Figure 1.2: The Breast Cancer Microenvironment. Overview of cells in the microenvironment of breast cancer, with a focus on immune cells and their interactions, eliciting pro- and anti-tumour effects. Original work published by Nadia Harbeck *et al.* and reproduced under license from Springer Nature under license number 5875321241984 [2].

Cancer-associated fibroblasts are crucially shaping the extracellular matrix (ECM), and act as signalling hubs, attracting specific immune-cell populations by the release of cyto- and chemokines. There are various subtypes serving these different functions, including mCAFs and iCAFs. The composition of fibroblast states has a profound influence on tumour progression [51].

In addition, there are many different immune cells, which can exert opposing effects on tumour cells. Pro-tumourigenic effector functions are primarily ex-

erted by cells that suppress an effective anti-tumour immune response. Foxp3$^+$ regulatory T cells (T_{reg}) suppress the immune response by four complementary mechanisms. These are: Immune suppression by inhibitory cytokines, including IL-10 and TGFβ, suppression by cytolysis, suppression by metabolic disruption and suppression by modulation of DC maturation and function via CTLA4 and LAG3 [52].

Myeloid-derived suppressor cells (MDSCs) limit the anti-cancer response by inhibiting the activity of T cells and natural killer (NK) cells. They do so by releasing immunosuppressive factors like arginase-1, reactive oxygen species (ROS), and nitric oxide (NO), which can deplete essential nutrients needed by T cells, interfere with their signalling, or cause direct inhibition [53].

Immune checkpoints like PD-L1/PD-1 and CTLA-4 can prevent effective tumour cell killing by CD8 T cells. However, tumour reactive CD8 T cells, T_H1, NK and NK T cells, when stimulated by MHC class I, Trail B and Fas on cancer cells, or by MHC class I and II on antigen-presenting cells (APCs), have the capacity of mounting an effective anti-tumour response. It is for this reason, that the number of tumour-infiltrating leukocytes (TILs), is a key measure predicting response to immunotherapy [54].

1.4 Breast Cancer Treatment

The treatment of breast cancer has evolved significantly over the past few decades, with therapies now tailored to the specific characteristics of the tumour and the individual patient [55]. The primary goals of breast cancer treatment are to eradicate the tumour, prevent recurrence, and improve survival, while minimising adverse effects. Treatment strategies typically involve a combination of surgery, systemic therapies, and radiation, and are informed by the subtype, stage, and receptor status of the cancer, as well as patient-specific factors such as age, overall health, and personal preferences.

1.4.1 Surgery

Surgery is often the first line of treatment for early-stage breast cancer and aims to remove the tumour along with surrounding healthy tissue to achieve local control of the disease. The two main surgical options are Breast-conserving surgery (lumpectomy) and mastectomy.

Breast-conserving surgery involves the removal of the tumour and a margin of surrounding healthy tissue while preserving most of the breast. This option is often preferred for smaller tumours and, in most cases, is followed by radiation therapy to reduce the risk of local recurrence. Clinical trials have shown that lumpectomy followed by radiation provides equivalent survival outcomes to mastectomy for early-stage breast cancer [56, 57].

Mastectomy involves the removal of the entire breast and may be necessary for larger or more advanced tumours, multicentric disease (multiple tumours in different quadrants of the breast), or when patients opt for this procedure based on personal preference. Mastectomy can be followed by breast reconstruction, either immediately or as a delayed procedure, using implants or autologous tissue from other parts of the body [58].

In addition to the removal of the tumour, sentinel lymph node biopsy or axillary lymph node dissection may be performed to assess whether the cancer has spread to the lymph nodes, which can inform decisions about the need for adjuvant therapy.

1.4.2 Radiation Therapy

Radiation therapy is commonly used after breast-conserving surgery to destroy any remaining cancer cells in the breast tissue, reducing the risk of local recurrence. It is also sometimes employed after mastectomy for patients with large tumours, involved lymph nodes, or other high-risk features. Advances in radiation techniques, such as intensity-modulated radiotherapy (IMRT) and partial breast ir-

radiation, have allowed for more precise targeting of cancerous tissue while sparing healthy structures, thereby minimising side effects [59].

1.4.3 Systemic Therapies

Systemic therapies, including chemotherapy, hormonal therapy, and targeted therapies, are used to treat cancer cells that may have spread beyond the breast and lymph nodes. The choice of systemic therapy is based on the molecular characteristics of the tumour, particularly receptor status, and the stage of the cancer.

Chemotherapy

Chemotherapy is a systemic treatment that uses cytotoxic drugs to kill rapidly dividing cancer cells. It is often given in the neoadjuvant setting (before surgery) to shrink tumours, particularly in larger or locally advanced breast cancers, or in the adjuvant setting (after surgery) to reduce the risk of recurrence. Chemotherapy is especially important for triple-negative breast cancer (TNBC), which lacks hormone receptors and HER2 expression, making it less responsive to hormonal or targeted therapies [60]. The choice of chemotherapy regimen depends on the tumour subtype and patient-specific factors [61].

Hormonal Therapy

Hormonal (endocrine) therapy is used for hormone receptor-positive breast cancers (ER+ and/or PR+) [62]. These therapies work by either blocking oestrogen receptors or reducing oestrogen production, thus starving the cancer cells of the hormones they need to grow. Common hormonal therapies include: Tamoxifen, a selective oestrogen receptor modulator (SERM), which blocks oestrogen receptors in breast tissue. Aromatase inhibitors (e.g., anastrozole, letrozole), which reduce the production of oestrogen by inhibiting the enzyme aromatase, particularly useful in postmenopausal women . Ovarian suppression, achieved through

surgery, radiation, or medication (e.g., goserelin), can also be used to reduce oestrogen levels in premenopausal women.

Targeted Therapies

Targeted therapies are designed to attack specific molecular targets involved in cancer growth and survival. The most notable example is HER2-targeted therapy for HER2-positive breast cancers [63]: Trastuzumab (Herceptin) is a monoclonal antibody that targets the HER2 protein and has dramatically improved outcomes for patients with HER2-positive breast cancer. Other HER2-targeted therapies include pertuzumab, trastuzumab emtansine (T-DM1), and lapatinib [64]. Recent advances in targeted therapy also include PARP inhibitors, such as olaparib, for patients with BRCA1/2 mutations, particularly in the context of triple-negative breast cancer [65].

Immunotherapy

Immunotherapy has emerged as a promising treatment, particularly for triple-negative breast cancer [66]. Drugs like atezolizumab and pembrolizumab, which are immune checkpoint inhibitors, work by enhancing the immune system's ability to recognise and attack cancer cells. These therapies are often combined with chemotherapy in certain settings [67]. TNBC is most likely to benefit from immunotherapy. Further factors predictive of response are PD-L1 positivity and a high number of tumour-infiltrating leukocytes (TILs). There is a positive linear correlation between TIL content at diagnosis and clinical outcomes under immunotherapy [66].

1.4.4 Neoadjuvant and Adjuvant Therapy

The use of neoadjuvant (before surgery) and adjuvant (after surgery) therapies depends on the stage and subtype of the breast cancer [2]. Neoadjuvant therapy

(NACT) is commonly used to shrink tumours, making breast-conserving surgery possible for larger tumours, or to assess tumour response to therapy. Adjuvant therapy is used to eliminate any residual cancer cells and reduce the risk of local recurrence. While NACT makes patients more likely to receive breast-conserving surgery, it is associated with a significantly higher risk of local recurrence. Measures like distant recurrence, breast cancer mortality and all-cause mortality are the same for both NACT and adjuvant chemotherapy [68]. The combination of surgery, radiation, chemotherapy, and/or hormonal and targeted therapies is tailored to the individual characteristics of the tumour and the patient's overall condition .

1.4.5 Emerging Therapies

Ongoing research is continually improving breast cancer treatment, with novel therapies targeting various pathways being under active development [2] (Figure 1.3). These include additional kinase inhibitors, new immune checkpoint inhibitors, and compounds targeting angiogenesis, like the VEGF-targeted antibody Bevacizumab.

In addition, antibody-drug conjugates (like sacituzumab-govitecan, targeted to TROP-2), and therapies targeting the tumour microenvironment are entering the arsenal as well [69].

Personalised medicine approaches, guided by genomic profiling of the tumour, such as the COGNITION trial [48], the phase II treatment trial COGNITION-GUIDE (NCT05332561) and CATCH trial at the NCT in Heidelberg, are also increasingly being used to predict treatment responses and tailor therapies to individual patients [49].

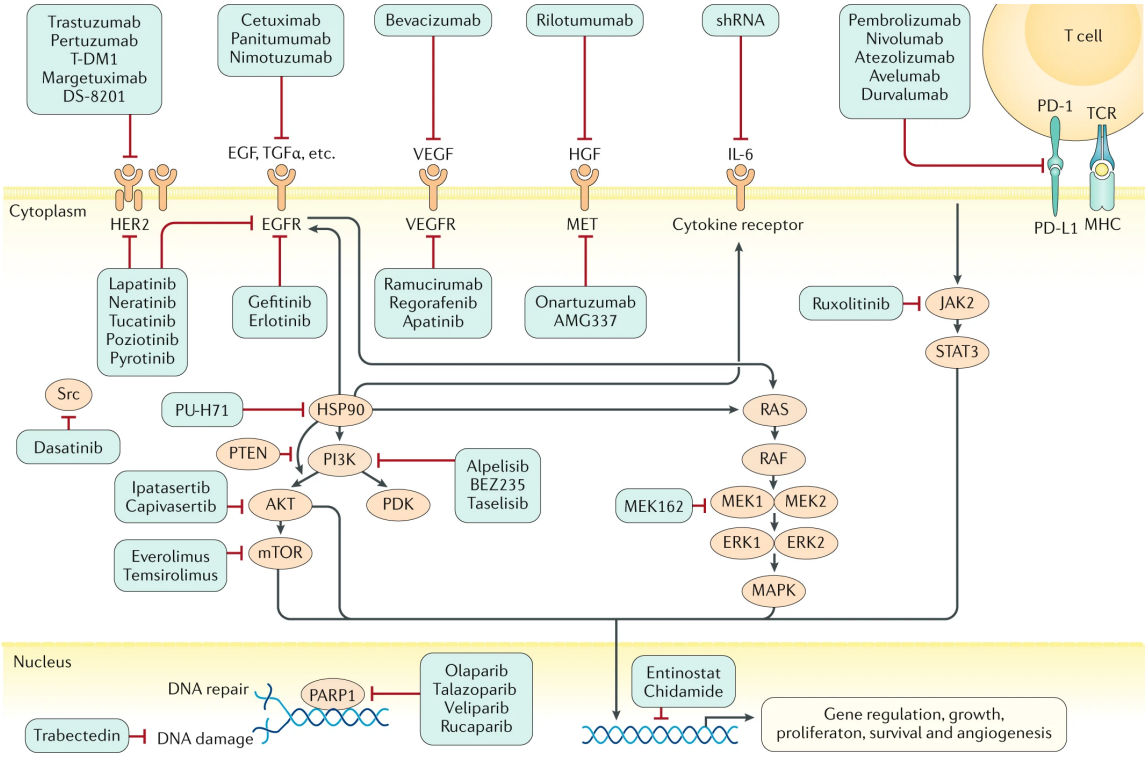


Figure 1.3: Emerging targeted therapies for breast cancer. Overview of molecular pathways that are targeted by emerging therapies for breast cancer. Original work published by Nadia Harbeck *et al.* and reproduced under license from Springer Nature under license number 5875321241984 [2].

1.4.6 Treatment Resistance

Successful treatment is unfortunately still often hampered by the development of therapeutic resistance and metastatic relapse. A variety of mechanisms have previously been described (reviewed in [70] and [71]). They include amplifications of driver genes (e.g. MYC, JAK2, MCL1 or PIM1), or loss of protective genes such an PTEN. It seems likely that genomic instability is the underlying mechanism allowing tumour cells to rapidly adapt to treatment pressure. But some links have also been made to the microenvironment of the primary tumour [72] or its metastatic site [73]. More research is needed in order to pin down specific resistance mecha-

nisms and find effective combination therapies [74], in order to overcome primary and acquired resistance to therapy.

In summary, breast cancer treatment is a multi-modal approach involving surgery, radiation, systemic therapies, and emerging personalised treatments. Advances in targeted therapies, immunotherapy, and genomic testing have revolutionised breast cancer management, improving outcomes and providing hope for even the most aggressive subtypes.

1.5 Rare Breast Cancers

Although invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC) account for the majority of breast cancer cases, a subset of breast cancers are considered rare. These rare subtypes of breast cancer exhibit unique biological characteristics, distinct clinical behaviours, and often require different therapeutic approaches. While each individual rare subtype may account for a small percentage of breast cancer diagnoses, collectively they represent a significant clinical challenge due to their diverse nature and often limited research data.

The different rare breast cancers subtypes can be classified by prognosis [75]. There are rare breast cancers with good and bad prognosis, the former of which can be further subclassified by hormone receptor expression (Table 1.1).

Rare subtypes of breast cancer I worked on as part of my doctoral research are metaplastic breast cancer and neuroendocrine breast cancer. Both of these subtypes are associated with therapy resistance and poor outcomes for the affected patients.

Metaplastic breast cancer (MpBC) is a rare and highly aggressive subtype of triple-negative breast cancer (TNBC), accounting for less than 1% of all breast

Table 1.1: Overview of rare breast cancers and their frequency relative to all breast cancers [75]

Tumor histology	Prognosis	Frequency
Mucinous adenocarcinoma*	Good	2–3%
Tubular carcinoma*	Good	1–2%
Cribriform*	Good	0.40%
Invasive papillary*	Good	0.50%
Medullary carcinoma**	Good	3–5%
Apocrine carcinoma**	Good	0.3–4%
Low-grade adenosquamous carcinoma**	Good	<0.1%
Secretory carcinoma**	Good	<1%
Pleomorphic lobular	Poor	<1%
Inflammatory carcinoma	Poor	2–4%
Metaplastic carcinoma	Poor	<1%
Neuroendocrine carcinomas of the breast	Poor	<1%
Micropapillary carcinoma	Poor	0.9–2
Lipid-rich carcinoma	Unclear	<0.1%
Oncocytic carcinoma	Unclear	<0.1%
Sebaceous carcinoma	Unclear	<0.1%
Glycogen-rich carcinoma	Unclear	0.9–3%
Mucinous cystadenocarcinoma	Unclear	<0.1%

* HR positive; ** HR low/negative

cancers [76]. This subtype is characterised by its mixed cellular composition, which may include squamous, spindle, and mesenchymal elements, making it histologically distinct from typical breast cancers [77]. MpBC is often resistant to conventional chemotherapy used for TNBC, and the prognosis tends to be worse than other forms of triple-negative breast cancer [78].

Treatment for MpBC typically involves surgery, followed by chemotherapy and radiation. However, due to its resistance to standard treatments, novel therapeutic approaches such as immunotherapy and targeted therapies are being explored in clinical trials. Genomic studies of MpBC have revealed a variety of mutations and alterations, which may offer future avenues for targeted treatment [79].

Neuroendocrine breast cancer (NEBC) is a rare subtype, accounting for 0.1–5% of all breast cancers. The uncertainty in this number is due to the infrequency with which breast cancer specimens are tested for neuroendocrine markers [80]. NEBC tumours display features of both breast epithelial cells and neuroendocrine cells, typically identified by the expression of markers like chromogranin A and synaptophysin. NEBC can be either well-differentiated, which is associated with more favourable outcomes, or poorly differentiated, which is more aggressive [81, 82].

Most NEBC cases are hormone receptor-positive (ER+ and/or PR+), and of either luminal A or luminal B molecular subtype, with the majority of poorly differentiated NEBC belonging to the latter subtype [83]. This allows for treatment with hormonal therapies such as tamoxifen or aromatase inhibitors. Surgery, typically followed by radiation, is the mainstay of treatment for localised NEBC, while chemotherapy is used for more advanced cases, especially poorly differentiated forms. Though HER2-positive NEBC is rare, HER2-targeted therapies like trastuzumab may be applicable if HER2 expression is present.

Due to its rarity, there is limited clinical data specific to NEBC, and most treatment approaches follow general breast cancer guidelines. Prognosis varies between studies, possibly in part owed to varying inclusion criteria, and based on the degree of differentiation, with well-differentiated NEBC typically having a more favourable outcome compared to poorly differentiated forms [81]. Further research is needed to tailor treatments specifically for this rare cancer type. Combination treatment approaches may be suitable to address some of the challenges posed by this type of breast cancer, and have been the subject of ongoing research in our lab.

Of note, inflammatory breast cancer, or IBC, has a poor prognosis and a comparatively high frequency of 2-4% of all breast cancers (Table 1.1). Thus, it caught my scientific interest, and I will go into more detail concerning its unique clinical

presentation and pathophysiology.

1.6 Pathophysiology of IBC

Inflammatory breast cancer is a highly aggressive form of breast cancer, accounting for 1–5% of all diagnosed breast cancer cases [84, 85]. Despite its relative rarity, IBC is responsible for 7–10% of breast cancer-related mortality [86, 87]. The clinical presentation of IBC is distinct from other forms of breast cancer, with characteristic signs and symptoms that resemble inflammatory conditions such as mastitis. These include swelling (edema) and redness (erythema) of the breast, accompanied by skin changes like peau d'orange (orange peel texture). The affected breast may be painful to the touch, and, in contrast to other breast cancers, a palpable mass is not always present. IBC involves the skin of the breast, leading to its classification as at least stage III (T4d) at the time of diagnosis [88, 89].

IBC is marked by its rapid clinical progression. Symptoms typically develop quickly, and the disease often metastasises early. Between 17 and 36% of patients with IBC present with metastatic disease at diagnosis, a stark contrast to the 5% incidence of metastasis in non-IBC cases [90]. The progression of metastasis continues aggressively, with up to 67% of IBC patients developing metastases within the first year post-diagnosis [90]. Furthermore, the overall survival of patients with IBC is significantly shorter compared to those with non-IBC, even at similar stages of the disease. For stage III patients, median survival is 4.75 years for IBC versus 13.4 years for non-IBC, while for stage IV patients, median survival drops to 2.27 years for IBC compared to 3.40 years for non-IBC [91, 92, 93, 94].

The diagnosis of IBC is primarily clinical, based on the aforementioned symptoms, and may be associated with certain histopathological features. However, histological findings alone are neither required nor sufficient for diagnosis, reinforcing the notion that IBC is a clinicopathological entity [90]. Importantly, IBC is not tied to a specific histologic subtype of breast cancer [95, 88]. It can exhibit any

of the molecular subtypes used in breast cancer classification [4, 40, 38], though there is an overrepresentation of hormone receptor-negative subtypes compared to non-IBC [96]. Key histopathological features include scattered tumour emboli, invasion of dermal lymphatic vessels, and lymphangiogenesis [97, 98]. The blockade of lymphatic drainage by tumour emboli is thought to contribute to the clinical appearance of IBC, with lymphatic fluid accumulation leading to the characteristic redness and swelling [99, 100, 95]. These emboli also facilitate rapid metastasis to lymph nodes and distant organs, and they may promote therapeutic resistance through mechanisms such as dormancy [101, 102].

1.7 Management of IBC

Given the aggressive nature of IBC, prompt diagnosis and rapid initiation of treatment are critical to improving patient outcomes. However, the clinical diagnosis of IBC poses several challenges. A key difficulty is the frequent absence of a distinct tumour mass, which complicates detection in routine mammograms. Nevertheless, the severity and rapid onset of clinical symptoms typically prompt patients to seek medical attention early, with differential diagnosis from mastitis being an essential step in the diagnostic process [90, 103]. This is because some of the key symptoms are similar, including redness and swelling, lending the name to inflammatory breast cancer as well.

The treatment of IBC usually involves a multimodal approach, combining chemotherapy, surgery, and radiation therapy. Due to its advanced stage at presentation, neoadjuvant chemotherapy is often initiated to reduce tumour size and address both local and distant disease. Surgical intervention, typically mastectomy, is then performed, followed by adjuvant radiation therapy. Despite advances in these treatment strategies, IBC remains challenging to manage, and outcomes remain poorer than those for non-IBC breast cancer.

1.8 Previous Studies of IBC

Understanding the molecular drivers of IBC has been the focus of numerous studies, including efforts from international research consortia [104]. However, despite extensive research, identifying a specific gene expression signature that definitively explains the aggressive nature of IBC has proven elusive [105, 106, 104]. Earlier attempts to achieve this included methods such as microdissection, aimed at reducing the confounding influence of the tumour microenvironment on gene expression analysis [106], and correlating gene signatures with clinical responses to neoadjuvant chemotherapy [107]. These efforts have uncovered dysregulated signalling pathways associated with IBC, including increased stemness markers (e.g. ALDH1 and CD44 [108]), epithelial-to-mesenchymal transition (EMT), VEGF, RAS, PI3K, JAK-STAT, and TGFβ pathways [109, 110, 111]. Genomic studies have similarly failed to reveal distinctive mutations unique to IBC, instead demonstrating patterns of alteration similar to those observed in non-IBC breast cancers [112, 113].

Emerging evidence suggests that the tumour microenvironment (TME) plays a pivotal role in shaping the unique phenotype of IBC [114]. Studies consistently highlight the significance of inflammatory and immunosuppressive pathways in the TME. Tumour-associated macrophages (TAMs), for instance, have been implicated in enhancing the metastatic potential and invasiveness of IBC cells [110]. Additionally, T cell suppression within the TME suggests a potential role for immune checkpoint inhibitors in IBC therapy [115]. The interaction between IBC cells and the TME, through both cytokine-mediated communication and direct cell-to-cell contact, contributes to lymphangiogenesis, metastasis, and immune evasion [116, 117, 118].

Despite the critical role of the microenvironment in IBC biology and its aggressive clinical course, a comprehensive single-cell analysis of IBC and its microenvironment remains lacking. Given the potential therapeutic implications of targeting the TME, there is an urgent need for studies that elucidate the cellular and molecu-

lar interactions within this environment, with single-cell and spatial transcriptomics analyses seeming to be uniquely suited to address this challenge.

1.9 Single-Cell Transcriptomic Analysis

Single-cell analysis has revolutionised the study of cancer biology, enabling researchers to investigate cellular heterogeneity at an unprecedented resolution. This technology allows for the detailed characterisation of tumour cells, immune cells, and the tumour microenvironment, leading to insights into cancer progression, metastasis, and treatment resistance. By profiling individual cells, it is possible to identify rare cell populations, elucidate developmental trajectories, and uncover the molecular mechanisms driving tumour heterogeneity.

1.9.1 Technology

Several technological platforms have been developed for single-cell analysis, each offering different advantages in terms of sensitivity, throughput, and cell capture efficiency. Two of the most prominent methods are plate-based and droplet-based single-cell RNA sequencing (scRNA-seq) [119].

Plate-based methods, such as SMART-seq2, involve isolating individual cells into separate wells of a plate (e.g., a 96- or 384-well plate), followed by library preparation and RNA sequencing [120]. This approach offers high sensitivity, allowing for the detection of full-length transcripts and more detailed gene expression profiles, particularly for lowly expressed genes. However, plate-based methods are labour-intensive, expensive, and typically lower throughput compared to droplet-based methods. Despite these limitations, plate-based techniques are still useful when high sensitivity and full-length transcript data are required, particularly in studies where the number of cells is small, such as rare cell populations or fine dissection of tumour biology.

Droplet-based platforms, such as 10x Genomics Chromium, have become the dominant technology for single-cell RNA sequencing due to their scalability and cost-effectiveness. In droplet-based scRNA-seq, individual cells are encapsulated into nanolitre-sized droplets, along with reagents for reverse transcription, allowing thousands of cells to be processed simultaneously. Barcoding each cell within the droplet enables the downstream identification of individual cell transcripts after sequencing [121]. While droplet-based approaches offer high throughput, capturing tens of thousands of cells in a single experiment, they provide 3' or 5' end reads rather than full-length transcripts, which limits their resolution compared to plate-based methods. Despite this, droplet-based platforms are invaluable for studying large-scale cell populations and uncovering cellular heterogeneity in complex tissues like breast tumours.

When using droplet-based scRNA-seq platforms, cell hashing is a technique used to address the challenge of high multiplet rates, which occur when more than one cell is encapsulated in the same droplet, leading to mixed signals and data misinterpretation [122]. Multiplet rates increase as the number of cells processed increases, limiting the number of cells than can be processed simultaneously. In addition, the droplet technology on its own does not allow to identify cells from pooled samples without additional measures taken to that end.

Cell hashing mitigates the issue of increasing multiplet rate and allows for a higher number of cells loaded per well of a microfluidics device by labelling cells with oligonucleotide barcodes before samples are pooled. Each sample is assigned a unique barcode, enabling the identification of cells from different samples after sequencing. This method significantly reduces multiplet error rates and reduces batch-related variation, allowing for more accurate single-cell analyses even when processing large numbers of cells in a single experiment.

1.9.2 Computational Tools and Analysis Best Practices

The complexity and scale of single-cell data require robust computational tools for accurate interpretation. As single-cell RNA sequencing has grown in popularity, a wide array of bioinformatics pipelines and software have emerged to manage data preprocessing, quality control, dimensionality reduction, and biological interpretation. Those include the Seurat Suite of tools developed for use in R and Scanpy, which is the most popular suite of tools in python. Best practices in single-cell data analysis emphasise careful quality control, batch effect correction, normalisation and clustering, followed by specific downstream analyses such as reference mapping and cluster annotation. Based on those, specific mechanistic studies are possible, including differential gene expression, gene set enrichment, cell type composition and cell-cell interaction studies.

Adhering to best practices in single-cell RNA sequencing [123] is crucial for producing reliable and reproducible results, particularly in complex analyses. Rigorous quality control is essential, with low-quality cells, such as those showing high mitochondrial gene expression or low total RNA content, needing to be filtered out to prevent artefacts in the data.

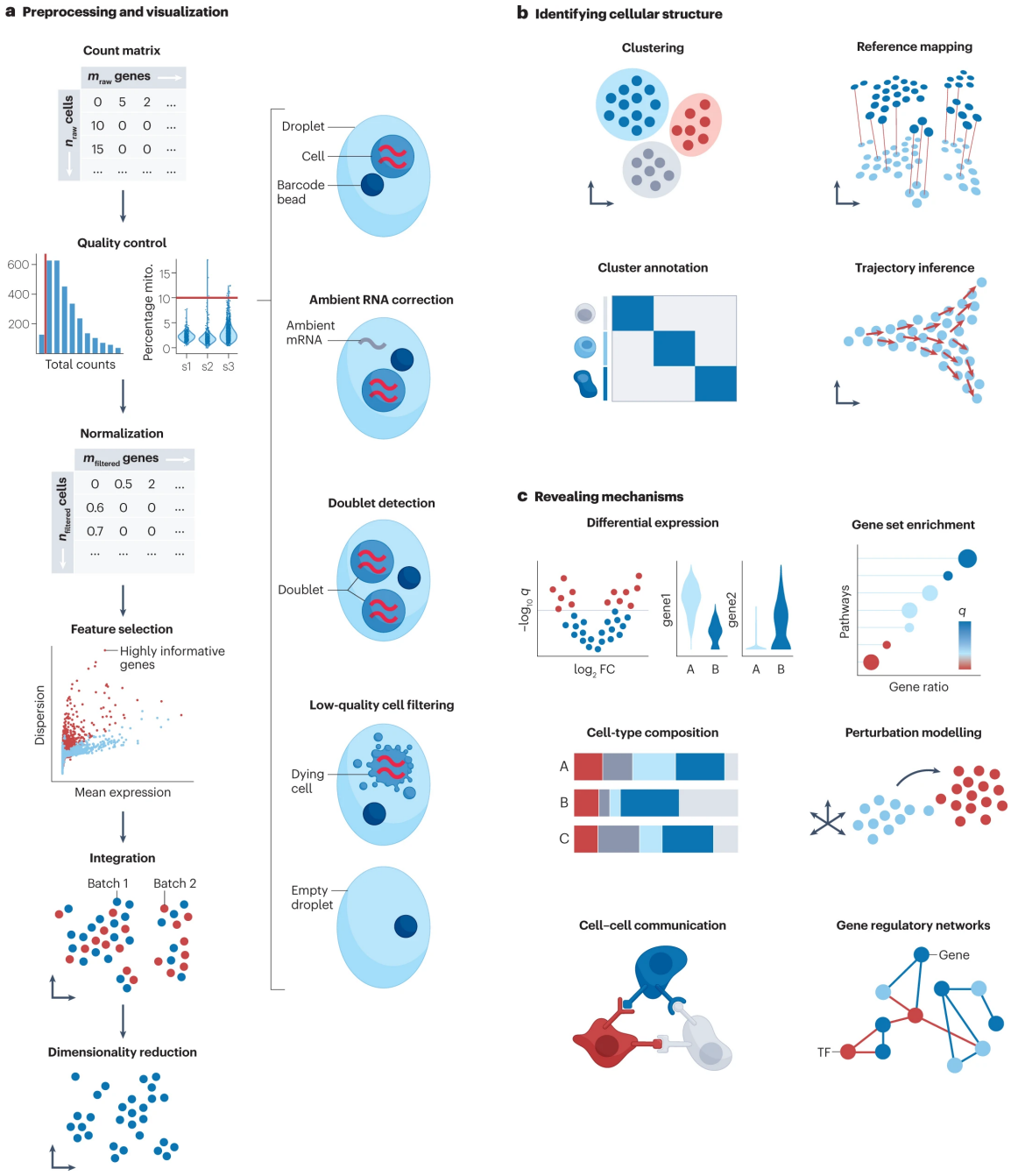


Figure 1.4: Single cell analysis best practices. Overview of the workflow for best practices assuring data quality for single-cell transcriptomics. Original work published by Lukas Heumos *et al.* and reproduced under license from Springer Nature under license number 5875441346575 [123].

Addressing batch effects is equally important, especially when samples have been processed in multiple batches, or data from various studies are integrated. Methods like Harmony [124] or Seurat's integration [125] are widely used to correct for technical variations, ensuring that the biological differences between cells, rather than technical noise, drive clustering and downstream analyses.

Dimensionality reduction techniques, such as PCA, UMAP, and t-SNE, help simplify the complexity of single-cell data, allowing distinct cell populations to be identified more easily. Clustering algorithms like Louvain and Leiden are then applied to group cells with similar gene expression profiles.

Once the initial data processing is complete, downstream analyses, including differential expression and gene set enrichment analysis help contextualise which pathways are active in different cell populations.

By following these best practices and employing advanced computational tools, single-cell data can yield meaningful biological insights, leading to new discoveries such as understanding tumour microenvironment heterogeneity, and uncovering novel therapeutic targets.

1.10 Spatial Transcriptomics

Spatial transcriptomics is an emerging technology that allows researchers to map gene expression within the spatial context of tissues, providing insight into how the location of cells within a tissue influences their function and behaviour. Unlike traditional transcriptomics, which dissociates tissues into individual cells or extracts bulk RNA from tissues, spatial transcriptomics preserves the spatial architecture of the tissue. This is particularly valuable in understanding complex diseases like breast cancer, where tumour heterogeneity and interactions between cancer cells and the tumour microenvironment are crucial drivers of disease progression and therapeutic resistance.

1.10.1 Technology

Spatial transcriptomics integrates histological tissue imaging with RNA sequencing. Tissue sections are placed on specialised slides containing spatially barcoded oligonucleotides. Once tissue-specific transcripts are captured, the spatial barcodes allow the reconstruction of gene expression patterns corresponding to their original positions in the tissue. At the time I performed my research, there were two main approaches for this (Figure 1.5). However, it is a field under rapid development, and new tools and methods are likely to become available and further transform the way spatial transcriptomics is done.

The first of the two main approaches is array-based spatial transcriptomics, where transcriptomes are recorded for an array of spots. Each spot measures the transcript levels of several adjacent cells at the same time. Typically, a deconvolution approach is used as part of the analysis, in order to analyse which cell types are likely to be included in each spot [123].

The second main approach is image-based spatial transcriptomics, which measures the transcript levels by directly imaging the tissue section of interest, combined with one of several approaches to visualise transcripts. This is typically not an unbiased measurement. Instead, a panel of genes of interest is selected in advance and only the levels of those genes are being measured.

Here is a small selection of the multitude of platforms that have, and continue to, become available:

10x Genomics Visium uses capture areas on glass slides to spatially resolve gene expression, allowing the profiling of tissue sections with high throughput. This platform balances spatial resolution and transcriptome-wide analysis, making it a popular choice for spatial studies in oncology. However, being an array-based method, its resolution is not down to single cell level, but rather spots including several adjacent cells.

In Situ Hybridisation (ISH)-based methods like MERFISH (multiplexed error-robust fluorescence in situ hybridisation) and seqFISH (sequential fluorescence in

situ hybridisation) enable the direct visualisation of RNA transcripts in tissue sections, providing higher resolution, but are often limited to predefined gene panels due to cost and complexity.

The 10x Xenium platform uses highly multiplexed RNA probes to hybridise with target transcripts, followed by rolling circle amplification. Then in successive rounds of fluorescent probe hybridisation, imaging and probe removal to detect and quantify these transcripts at single-cell resolution. The limitations of this approach include the limited panel size of approximately 400 genes.

1.10.2 Applications in Cancer Research

In cancer research, spatial transcriptomics has been instrumental in revealing the intricate relationships between tumour cells and their microenvironment. The spatial context allows researchers to investigate how interactions between tumour cells, immune cells, stromal cells, and vasculature contribute to tumour progression, immune evasion, and resistance to therapy.

Tumour Heterogeneity: Solid cancers are highly heterogeneous, with different regions of the tumour often harbouring distinct molecular profiles. Spatial transcriptomics has been used to map this heterogeneity, revealing subpopulations of cancer cells that may drive metastasis or respond differently to treatment. For example, Berglund et al. [126] used spatial transcriptomics to map gene expression in prostate cancer tissues and identified distinct regions of tumour cells that exhibited differential immune evasion signatures, providing clues to why certain regions of the tumour responded differently to immunotherapy.

Tumour Microenvironment (TME): The tumour microenvironment plays a pivotal role in cancer progression, influencing tumour growth, metastasis, and response to therapies. Spatial transcriptomics allows for the precise mapping of interactions between tumour cells and surrounding stromal, immune, and vascular cells. For example, Andersson et al. [127] developed a method for spatial transcriptomics to reveal the localisation of cells in the microenvironment.

Immune Cell Infiltration: Spatial transcriptomics has enabled researchers to profile immune cell populations within the tumour microenvironment with spatial resolution. This is critical for understanding the immune landscape of tumours, which can vary significantly across different regions of the tumour. In a study by [128], a deconvolution approach of single-cell and spatial transcriptomics identified populations within pancreatic cancer samples, revealing spatial variations in immune cell distribution and cell-type co-enrichment.

1.10.3 Computational Tools and Best Practices

The analysis of spatial transcriptomics data requires specialised computational tools and approaches due to the complexity of combining gene expression data with spatial information. Several frameworks have been developed to process and interpret these datasets.

Squidpy and Seurat V5 are widely used tools that integrate spatial information with single-cell gene expression data, enabling clustering of spatial gene expression patterns and identification of spatially variable genes.

Best practices in spatial transcriptomics include careful selection of tissue sections, ensuring that both the tumour and relevant microenvironmental regions (such as invasive margins or immune hotspots) are included. Additionally, integrating histopathological annotations with gene expression data is critical to interpret spatial patterns accurately. Data normalisation, clustering, and dimensionality reduction methods are essential for identifying spatial domains and understanding cellular interactions.

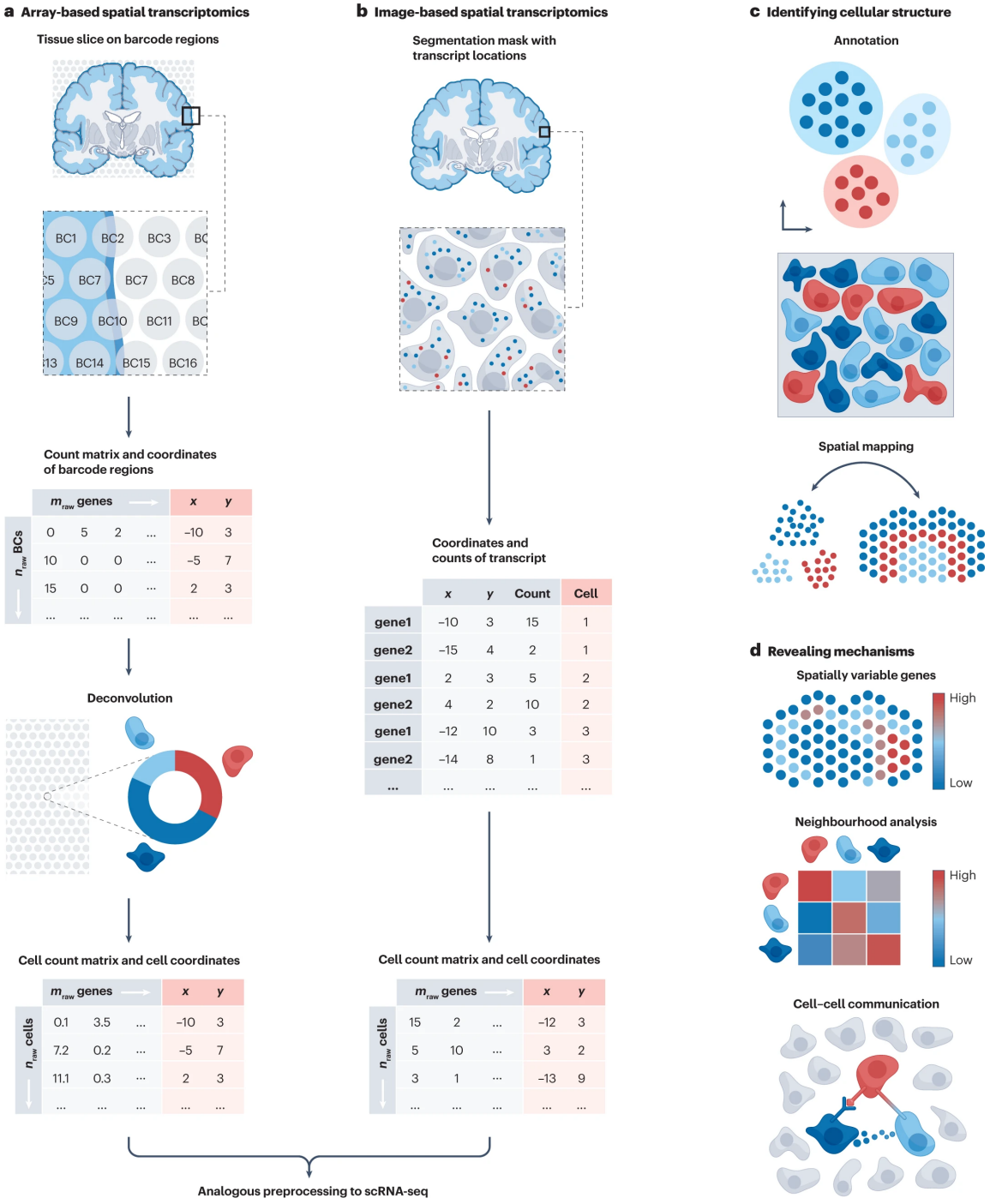


Figure 1.5: Spatial transcriptomics analysis best practices. Overview of the workflow for best practices assuring data quality for spatial transcriptomics. Original work published by Lukas Heumos *et al.* and reproduced under license from Springer Nature under license number 5875441346575 [123].

1.10.4 Future Directions

As the technology for spatial transcriptomics continues to evolve, we can expect even higher resolution and greater throughput, enabling deeper insights into cancer biology. Combining spatial transcriptomics with other modalities, such as spatial proteomics (like MIBI-TOF [129]) , will provide a more comprehensive picture of the tumour microenvironment and its dynamic interactions. Furthermore, integrating spatial transcriptomics with longitudinal tissue sampling, as is the goal of the SATURN3 consortium [130], may allow researchers to track the evolution of breast cancer in response to therapies, offering potential biomarkers for predicting treatment outcomes.

Chapter 2

Aim of This Dissertation

Inflammatory breast cancer (IBC) is a rare but aggressive subtype of breast cancer. Due to its unusual presentation, it is often diagnosed late, contributing to a worse prognosis compared to non-inflammatory breast cancer. Additionally, there are neither specific mutations nor a transcriptional signature associated with IBC, nor are there therapy guidelines targeting the specific biology of IBC.

It has previously been recognised that the microenvironment of IBC is essential to many of the disease's unique features. In recent years, the development of single-cell and spatial omics technologies has revolutionised the way scientists can probe cell types and pathways shaping the biology of physiological and pathological states.

Therefore, the objectives of this dissertation were to utilise emerging technologies, such as single-cell and spatial transcriptomics to delineate cell types, states, and cellular interaction networks contributing to IBC biology. To that end, the specific goals were to:

1. Curate a collection of tissue samples of IBC and non-IBC, create single-cell libraries, and sequence them, producing a high-quality dataset.
2. Assess and adapt available tools and analysis methods, in order to identify specific differences in the major cell types between IBC and non-IBC.
3. Integrate the single-cell data with spatial transcriptomics data to corroborate the gained insights using an orthogonal method.

In summary, through this dissertation, I aimed to contribute to our understanding of IBC biology and to identify potential therapeutic targets.

Chapter 3

Results

3.1 IBC sample collection

I curated a collection of samples from patients who were treated with an inflammatory breast cancer (IBC) diagnosis at the National Center for Tumor Diseases (NCT). The final collective of 9 biopsies from 9 individual patients included Luminal B, HER2-enriched and triple negative subtypes. I additionally included five samples of non-inflammatory breast cancer (non-IBC), reflecting all three subtypes (Table 3.1). All samples were from treatment-naïve early breast cancer and enrolled in the COGNITION trial [48].
As there are a number of published single-cell datasets of breast cancer, I decided to include those in the control (non-IBC) group for a more unbiased comparison and to achieve a balance in terms of molecular subtypes (see Table 3.3).

Table 3.1: Patient IDs, IBC status and molecular subtypes and processing batch of sample collective

Patient ID	Tumour type	Subtype by IHC	Batch
ARXDAG	IBC	TNBC	1
2DDCDC	IBC	LumB	1
FRGFMA	IBC	TNBC	2
JP6945	IBC	LumB	2
RM4MZJ	IBC	LumB	2
KQB5A9	IBC	LumB	2
C23KEF	IBC	HER2+	3
ZUV2AW	IBC	HER2+	3
*GXCLCM**	*IBC*	*HER2+*	*3*
*45XD3D**	*non-IBC*	*TNBC*	*3*
JRXVAL	non-IBC	LumB	3
JSYFKF	non-IBC	HER2+	3
XZS19C	non-IBC	HER2+	3
YW12FM	non-IBC	LumB	3
2DDCDC	IBC	LumB	4
KQB5A9	IBC	LumB	4
ARXDAG	IBC	TNBC	4

*10X chip failure

3.2 Single-Cell Library Preparation and Sequencing

The single-cell library preparation and sequencing were performed using the 10X Chromium 3.1 platform.

I needed to set up and optimise the protocol for the type of samples and conditions in our laboratory. Because of the material cost of 10X Genomics reagents and the rarity of IBC samples, I thawed them in batches and tried several methods to optimise cell recovery (Table 3.2).

Table 3.2: Cell numbers during pre-processing for single-cell library preparation. Recovery was based on number of sorted cells, where available, otherwise the initial cell count was used.

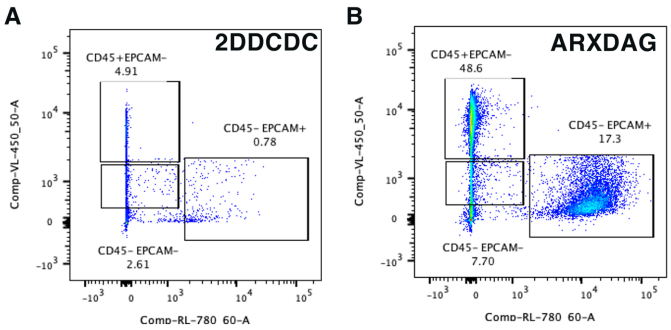
Patient	Count	Live sorted cells	Method	Cell Number After QC	Recovery
ARXDAG	1,300,000	40,000	Single	3,969	9.9%
2DDCDC	89,000	3,223	Single	583	18.1%
FRGFMA	1,500,000	7,060	Single	1,649	23.4%
JP6945	*na*	7,735	Single	573	7.4%
RM4MZJ	500,000	14,376	Single	5,472	38.1%
KQB5A9	400,000	100,000	Single	10,505	10.5%
C23KEF	24,000	*not sorted*	Hashed	1,399	5.8%
*GXCLCM**	14,250	*not sorted*	Hashed	-	-
*45XD3D**	25,000	*not sorted*	Hashed	-	-
JRXVAL	122,250	*not sorted*	Hashed	1057	0.9%
XZS19C	214,000	*not sorted*	Hashed	3495	1.6%
YW12FM	110,000	*not sorted*	Hashed	733	0.7%
JSYFKF	*na*	*na*	Hashed	93	*na*
ZUV2AW	*na*	14,250	Hashed	703	4.9%
2DDCDC	*na*	*na*	Hashed	629	*na*
KQB5A9	*na*	5,433	Hashed	755	13.9%
ARXDAG	*na*	13,860	Hashed	1,620	11.7%

*10X chip failure

To increase yield, I sorted live cells by flow cytometry for Batches 1, 2 and 4

(ARXDAG and KQB5A9)), wherever sufficient cells were present. Batch 1 was thawed and sorted with the help of Tasneem Cheytan. Because lymphocytes can more readily be isolated from tissue samples and tend to survive the process of digestion, washing, staining and single-cell encapsulation better, I tried to enrich for EpCAM-positive populations by flow cytometric sorting. A representative analysis of the sorting scheme used is shown for Batch 1 in Figure 3.1.

For Batch 3, in an attempt to further optimise the protocol, I tried to maximise yield by directly using unsorted cells and using barcode hashing. Since sample hashing with antibody-coupled barcodes allow the loading of a higher number of cells per well, I assumed that the total cell number loaded would not be the limiting factor. Thus, I accepted a potentially higher rate of non-viable cells as a trade-off to avoid the inefficient sorting procedure, which typically also comes with low recovery rates.



C

	2DDCDC	ARXDAG
CD45+	1897	16000
double negative	326	8000
EPCAM+	1000	16000
total	3223	40000

Figure 3.1: Flow cytometric sorting of primary samples for single-cell library preparation.
(A-B) Representative analysis of flow cytometric sorting of two samples to increase yield and balance cell type representation for single-cell library preparation.
(C) Table of cell numbers sorted for different markers of the two samples from A and B.
This experiment was performed in collaboration with Tasneem Cheytan.

Overall, the mean recovery rate, that is the ratio of cells included in the final dataset after all quality control procedures over the initial number of cells was 6.7% (SD 10.5%). Where available, the initial number of cells was defined as the count of viable sorted cells, otherwise on the initial cell count (Table 3.2).

After preparation of the single-cell suspension, individual cells were encapsulated into droplets using the 10X Chromium Controller, ensuring precise capture and isolation of each cell. RNA from each cell was then reverse transcribed and amplified within the droplets, leveraging the robust chemistry of the Chromium

Single Cell 3' Reagent Kits v3.1, which enhances the capture efficiency and sensitivity of transcriptomic profiles. The amplified cDNA libraries were subsequently indexed and pooled for sequencing on an Illumina NovaSeq 6000 platform for 100 paired-end cycles with S4 chemistry, achieving a median read depth of 90,850 reads per cell.

3.3 Data Pre-Processing

The single-cell RNA sequencing data generated from the 10X Chromium 3.1 platform were processed using the Cell Ranger software suite, version 7.1.0, to ensure accurate and comprehensive analysis of the transcriptomic profiles. I did not use the Cell Ranger mkfastq function to demultiplex the raw base call files (BCL) into FASTQ files, but rather used the fastq files provided by our in-house pipeline of the Omics IT and Data Management Core Facility. Based on those, the cellranger count function was utilised to align the sequencing reads to the reference genome (GRCh38), and to generate gene-barcode matrices. This step involved several critical processes, including read alignment using STAR, filtering of low-quality reads, barcode and UMI (unique molecular identifier) counting, and gene expression quantification.

3.3.1 De-Multiplexing of Hashed Libraries

The samples in Batches 3 and 4 were labelled with hashing antibodies prior to pooling and single-cell encapsulation. This labelling process was intended to facilitate later demultiplexing, allowing us to trace cells back to their original samples. However, the efficiency of this method proved suboptimal, resulting in low recovery rates during downstream demultiplexing. In Batch 3, the fraction of cells where no tag molecule could be assigned was > 30%.

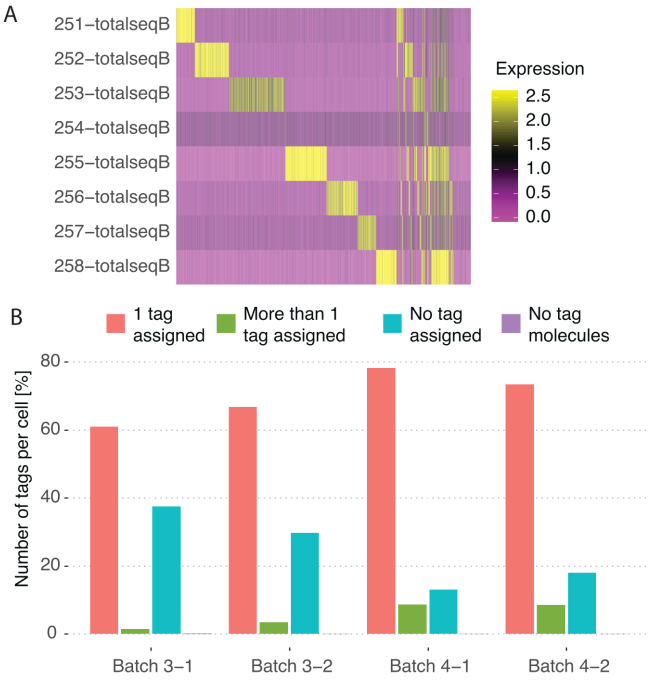


Figure 3.2: Labelling efficiency of cells with hashing antibodies.
(A) Heatmap of Batch 4-2 showing normalised and scaled counts of each hashing antibody.
(B) Bar graph showing the percentage of cells per batch which were labelled unambiguously, ambiguously, or were left unlabelled.

To address this challenge, I developed a custom workflow aimed at enhancing the demultiplexing of pooled samples labelled with hashing antibodies. This innovative approach leverages the unique genetic makeup of each sample, as each originates from a different patient and therefore exhibits distinct patterns of single nucleotide polymorphisms (SNPs) and mutations. By exploiting these genetic differences, it is possible to accurately cluster cells from different samples that were pooled prior to single-cell encapsulation and sequencing. Unlike previous methods that require prior knowledge of each individual's SNP patterns, my approach labels samples with hashing antibodies, clusters cells based on their SNP profiles, and assigns samples by identifying the most frequent hashtag within each cluster.

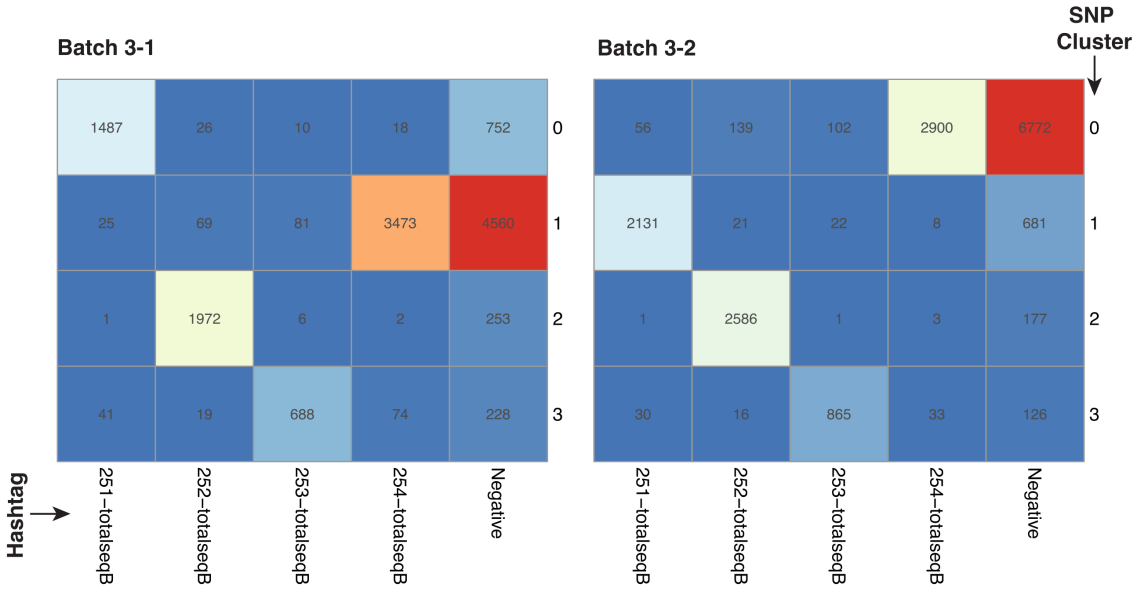


Figure 3.3: Additional cell recovery resulting from combined hashing and SNP-based demultiplexing. For samples in Batch 3, numbers of cells labelled with each hashing antibody (x) are plotted along with their SNP-based cluster assignment (y) in a tile plot. The sum of hashtag-negative cells are those that can be additionally recovered by using this method.

The sum of previously negative cells recovered additionally by using this approach was 5,793 for Batch 3-1 and 7,756 for Batch 3-2 (Figure 3.3). The SNP-based clustering also always assigned a low number of cells tagged with hashing antibodies of the non-majority to each cluster. However, those were consistently much fewer than the number of cells gained by including those that are hashtag-negative.

In addition to gaining additional cells that were not labelled effectively, heterotypic doublets or multiplets (from different patients) were identified and excluded using this method. Remaining duplicates (homotypic doublets, i.e. from the same patient), cannot be identified on a genetic basis and are addressed below using the modelling approach from the scDblFinder package.

3.3.2 Background RNA Removal

The next step of the pre-processing was to remove background RNA contamination from my single-cell RNA sequencing (scRNA-seq) data. Background RNA, originating from ambient RNA or from cell lysis during the experiment, can significantly interfere with the accurate profiling of individual cells.

To address this issue, I implemented a series of background RNA removal techniques. Initially, I utilised a computational approach that involved the application of the SoupX algorithm [131]. This algorithm estimates the background RNA contamination levels by analysing the expression of genes across empty droplets and adjusts the expression values in the actual cell-containing droplets accordingly.

As the contamination value is estimated independently for each gene, the result is a probability distribution. Since accurate estimates cluster around the real contamination value, the most common estimate should be close to the actual value. For example for Batch 3-2, the most likely contamination level is determined at 0.01 (Figure 3.4). Contaminating counts are then subtracted from cells one cluster at a time, by multiplying the sum of counts of each gene within the cluster with the determined contamination level. The resulting contaminating counts are then distributed across all cells in the cluster and subtracted.

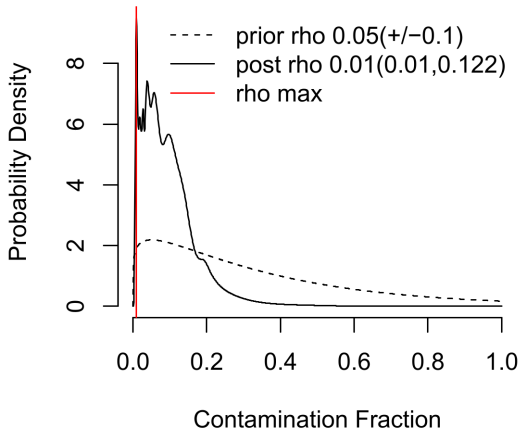


Figure 3.4: Exemplary output plot of the SoupX function. To estimate the background RNA contamination of Batch 3-2, prior probability density is plotted as dashed line, while the probability found by the algorithm is plotted as a solid line. Most likely RNA contamination value is indicated by a vertical red line.

3.3.3 Multiplet Removal

My custom SNP-based demultiplexing approach previously removed doublets of different patients of origin within hashed samples. Next, for both hashed and unhashed samples, I used the scDblFinder package to remove homotypic doublets originating from the same patient. I did this on a per-patient basis, meaning after demultiplexing for the hashed samples.

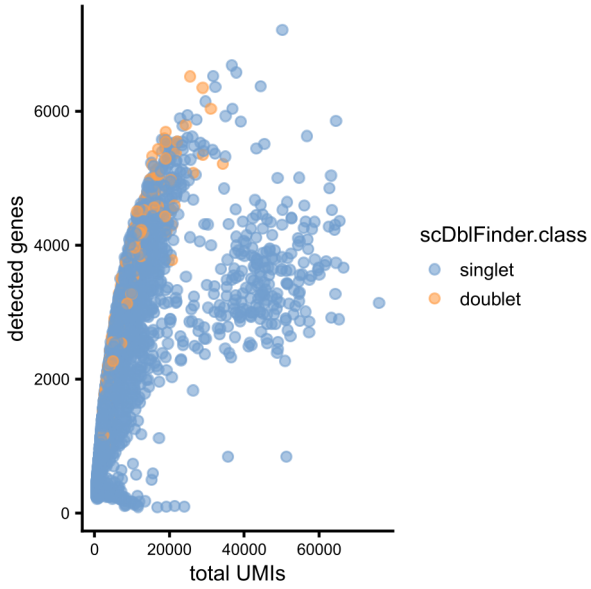


Figure 3.5: Homotypic doublet identification. Results of same-patient doublet filtering using the scDblFinder package. The total UMIs and the number of detected genes per cell are plotted on the X and Y axes respectively.

To assign a likelihood of a droplet to be a doublet or multiplet, scDblFinder calculates artificial doublets and correlates each droplet's profile to the artificial doublets' profile. To achieve that, a classifier is iteratively trained on an embedding of actual cells and artificially created doublets. The doublet rate (dbr) was not specifically assigned, but calculated by the algorithm, which the developers claim is a valid approach for 10X data [132].
The identified doublets tend to have higher numbers of detected genes, but rather not exceptionally high UMI counts (Figure 3.5).

3.3.4 Quality Control

To ensure high-quality data for downstream analyses, I filtered out cells with exceptionally high UMI levels and mitochondrial counts using a combined approach of median absolute deviation (MAD) and a hard cutoff. First, I calculated the de-

viation of each cell's respective value from the median of its sample. Then, I computed the median absolute deviation (MAD) for these deviations. Finally, any cell with a UMI count or mitochondrial percentage exceeding three times the MAD was excluded from the dataset. In addition, a hard cut-off of 25% mitochondrial genes was applied, excluding any cell above that threshold. The distribution of the values used for quality control were plotted as violin plots (Figure 3.6).

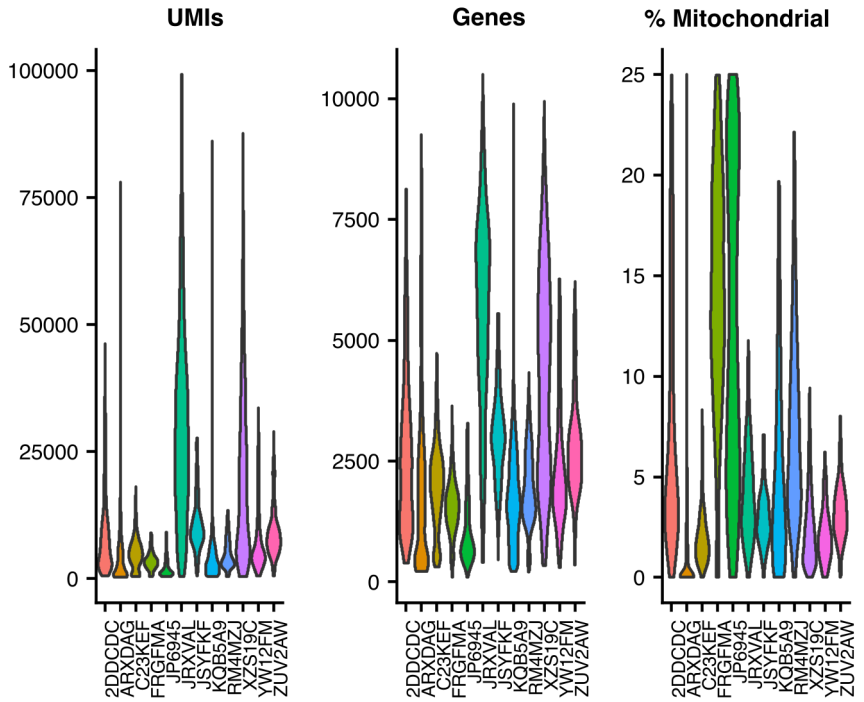


Figure 3.6: Single-Cell Data Quality Control. Violin plots showing the distribution of unique molecular identifiers (UMIs), detected genes, and the percentage of mitochondrial content across individual cells. Each violin represents a sample, with the width indicating the density of cells at different levels. Outliers with exceptionally high UMI counts and mitochondrial percentages were filtered out based on median absolute deviation (MAD) criteria to ensure data quality for downstream analyses.

3.3.5 Preprocessing of Public Datasets

For this study, I utilised published single-cell RNA sequencing data from two prominent breast cancer studies, focusing here on the dataset provided by Wu et al., published in Nature Genetics in 2021 [133]. This landmark study was the first to deliver a comprehensive single-cell and spatially resolved atlas of human breast cancers. The data unveiled significant tumour heterogeneity and intricate interactions within the tumour microenvironment. Although this study also incorporated spatially resolved transcriptomics using the 10X Visium platform for six samples, those specific data were not included in the analyses presented in this thesis.

The dataset was obtained from the Single Cell Portal hosted by the Broad Institute. The data were provided in 10X Genomics format, consisting of a count matrix in .mtx format, along with corresponding features and barcodes in .tsv format. A metadata table was also included in .csv format. To begin the analysis, I constructed a Seurat object and proceeded with the standard preprocessing steps. This included log normalization, identification of highly variable features, and scaling of the data. Principal component analysis (PCA) was performed, followed by Uniform Manifold Approximation and Projection (UMAP) for dimensionality reduction and visualisation.

The UMAP visualisation (Figure 3.7A) revealed distinct clustering of cells by cell type, indicating that the major cell lineages within the breast cancer samples were well-represented in the dataset. Further, when cells were coloured by sample ID (Figure 3.7B), the distribution of cells across different patient samples demonstrated the preservation of sample-specific heterogeneity within the subsampled dataset. Additionally, a stacked bar plot (Figure 3.7C) illustrates the proportional distribution of various cell types across different patient samples, highlighting the variable cell type composition within each tumour, which is a key characteristic of breast cancer's cellular diversity.

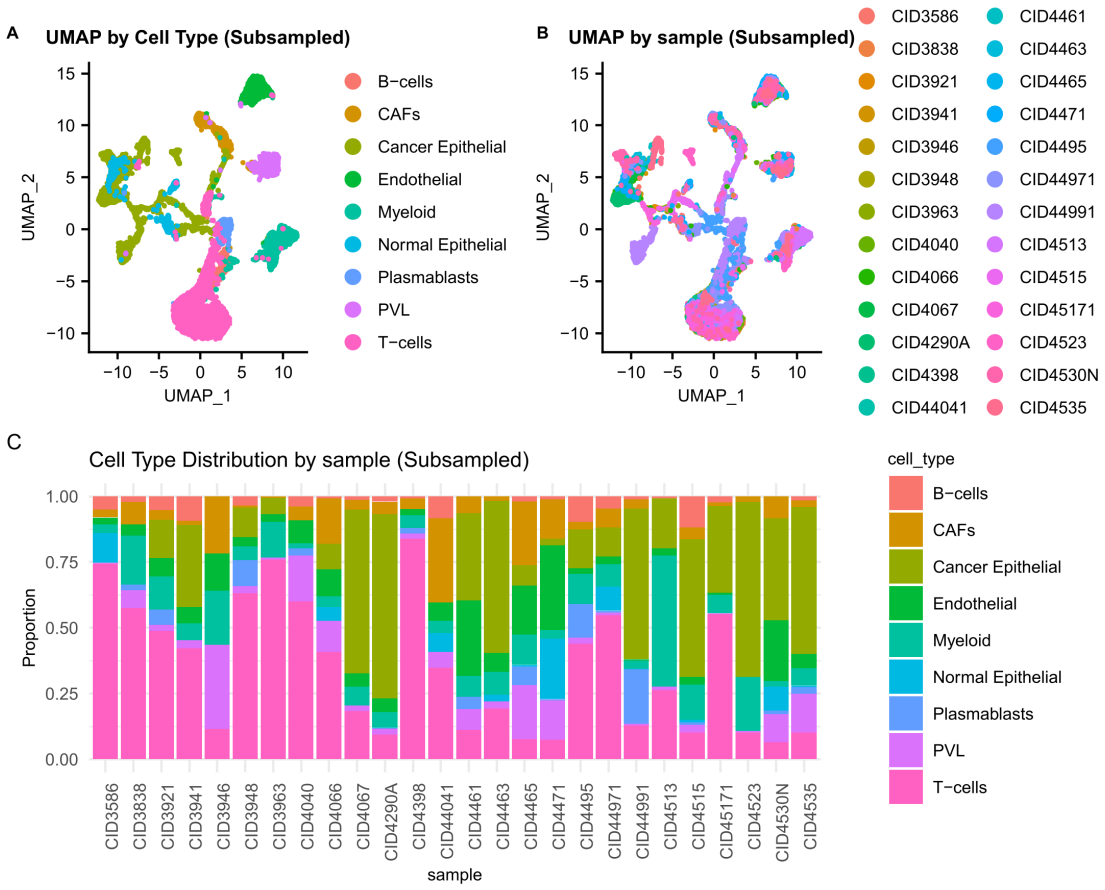


Figure 3.7: Overview of the Wu dataset's properties. The dataset was subsampled to 10% of cells for plotting.
(A) UMAP visualisation of cells colored by cell type, showing the clustering of different cell types in the subsampled dataset.
(B) UMAP visualisation of cells coloured by sample ID, illustrating the distribution of cells across different patients in the subsampled dataset.
(C) Stacked bar plot showing the proportional distribution of cell types across different patient samples, providing insight into the cell type composition within each sample.

In addition to the Wu et al. dataset, I incorporated data from another significant study conducted by Tietscher et al., published in Nature Communications in 2023 [134]. This study presented a comprehensive single-cell map of human breast

cancer, with a particular focus on T cells and markers of their exhaustion. The research highlighted the cellular interactions and signalling pathways contributing to immune suppression in the tumour microenvironment, revealing potential avenues for therapeutic intervention.

For this analysis, I retrieved the dataset from the EBI ArrayExpress repository under the accession number E-MTAB-10607. The dataset was freely accessible as it contained no identifiable patient information. The data included both counts and metadata tables, from which I created individual Seurat objects. I subsequently merged these objects, resulting in a comprehensive dataset comprising 110,729 labelled cells. The dataset encapsulated all major cell types present in the tumour microenvironment and exhibited clear clustering by cell type (Figure 3.8).

The UMAP visualisation (Figure 3.8A) displays the distinct clustering of cells according to their respective cell types, confirming the accurate annotation and representation of the major immune and stromal cell populations. Additionally, when cells were coloured by sample ID (Figure 3.8B), the visualisation illustrated the distribution of cells across various patient samples, indicating the preservation of patient-specific cellular diversity within the subsampled dataset. The stacked bar plot (Figure 3.8C) further details the proportional distribution of different cell types across the patient samples, providing insights into the variability in cellular composition among different tumours, a critical factor in understanding the immune landscape of breast cancer.

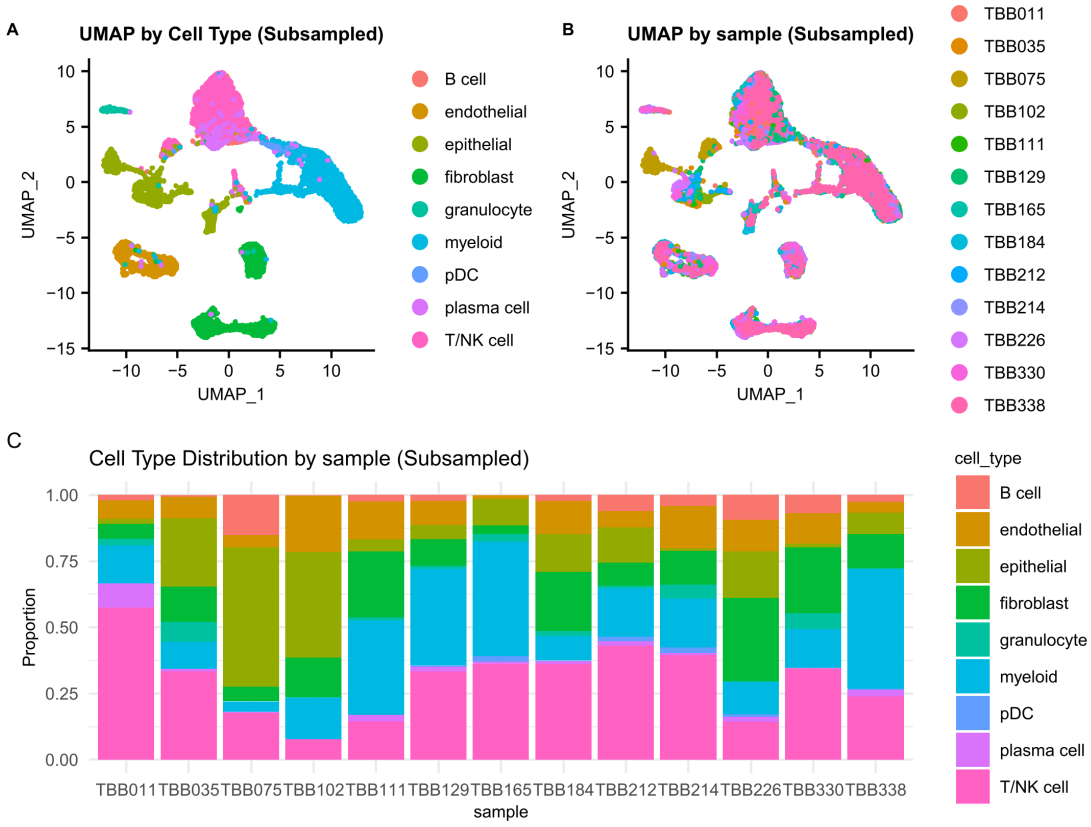


Figure 3.8: Overview of the Tietscher dataset's properties. The dataset was subsampled to 10% of cells for plotting.
(A) UMAP visualisation of cells colored by cell type, showing the clustering of different cell types in the subsampled dataset.
(B) UMAP visualisation of cells colored by sample ID, illustrating the distribution of cells across different patients in the subsampled dataset.
(C) Stacked bar plot showing the proportional distribution of cell types across different patient samples, providing insight into the cell type composition within each sample.

3.3.6 Integration of IBC and Non-IBC Datasets

After preprocessing the public datasets, I proceeded to create an integrated dataset combining all IBC and non-IBC cells sequenced in my study. To ensure

a balanced and representative analysis, I supplemented the non-IBC cells using the published datasets. The final integrated dataset was meticulously curated to include an equal number of cells for each molecular subtype across both IBC and non-IBC categories, thereby enabling a direct and unbiased comparison between the two conditions.

As shown in Table 3.3, the final dataset comprises an equal distribution of cells across the three major molecular subtypes: HER2-positive (HER2+), hormone receptor-positive (HR+), and triple-negative breast cancer (TNBC). This balanced design is crucial for comparative analyses, as it minimises potential biases arising from unequal representation of subtypes, thereby enhancing the robustness of subsequent analyses focused on identifying subtype-specific and condition-specific cellular characteristics.

Table 3.3: Numbers of cells per molecular subtype in the final dataset of IBC and non-IBC.

	HER2+	**HR+**	**TNBC**
IBC	703	19,916	7,238
non-IBC	703	19,916	7,238

3.3.7 Dimensionality Reduction

Dimensionality reduction is a critical step in the analysis of single-cell RNA sequencing data, enabling the visualisation of high-dimensional data in a more interpretable form. After merging of the data, they were projected onto a two-dimensional coordinate system to facilitate the identification of patterns and clusters corresponding to different cell types and states.

Initially, I performed PCA, using the Seurat package to identify the principal components of variation within the data. PCA is a linear dimensionality reduction technique that transforms the original high-dimensional data into a set of orthogonal (uncorrelated) components ranked by the amount of variance they capture

from the data. By focusing on the top principal components, I was able to capture the most significant sources of variation while reducing noise from less informative components.

Following PCA, I applied UMAP also using the Seurat package. UMAP is a non-linear dimensionality reduction technique that preserves both local and global data structures more effectively than PCA in lower dimensions. UMAP constructs a high-dimensional graph representing the data's local relationships and then optimises a low-dimensional projection of this graph. The result is a two-dimensional representation that maintains the relative distances between cells, thereby enabling the visualisation of complex cellular relationships and clusters.

The combination of PCA and UMAP is a standard approach in scRNA-seq data analysis. PCA serves to reduce the dimensionality of the data while retaining most of the variance, and UMAP further reduces the dimensions to two while preserving the overall structure and relationships within the data. This approach facilitates the identification of distinct cell populations and provides insights into the underlying biology, as visualised in UMAP plots.

3.3.8 Sample Harmonisation

Initially, I observed significant study-related heterogeneity in the data. Although the primary clustering in a UMAP embedding of the merged data was by cell type (Figure 3.9A), within those clusters, cells from the external datasets occupied specific spaces and were not evenly distributed (Figure 3.9C). This uneven distribution indicates that batch effects from the different studies were influencing the clustering, potentially confounding biological interpretations. Specifically, cells from each dataset tended to form subclusters within the broader cell type clusters, suggesting that technical differences between the studies were driving some of the observed variation.

To address this issue, I applied the Harmony batch correction algorithm, which is designed to mitigate these batch effects while preserving biological variability.

After applying Harmony, the UMAP embedding revealed a more homogeneous distribution of cells from the different datasets across each cell type cluster (Figure 3.9B). This adjustment resulted in cells from all datasets being more evenly dispersed within each cell type cluster, as shown in Figure 3.9D. The improved integration of data from different studies highlights the effectiveness of Harmony in reducing batch effects, allowing for a more accurate comparison of cell types across datasets.

The resulting homogeneity within cell type clusters post-Harmony correction suggests that the observed cellular differences are more likely to reflect true biological variation rather than technical artefacts. This step was crucial for ensuring the validity of downstream analyses, such as differential expression and pathway enrichment, which rely on the accurate representation of cellular heterogeneity across studies.

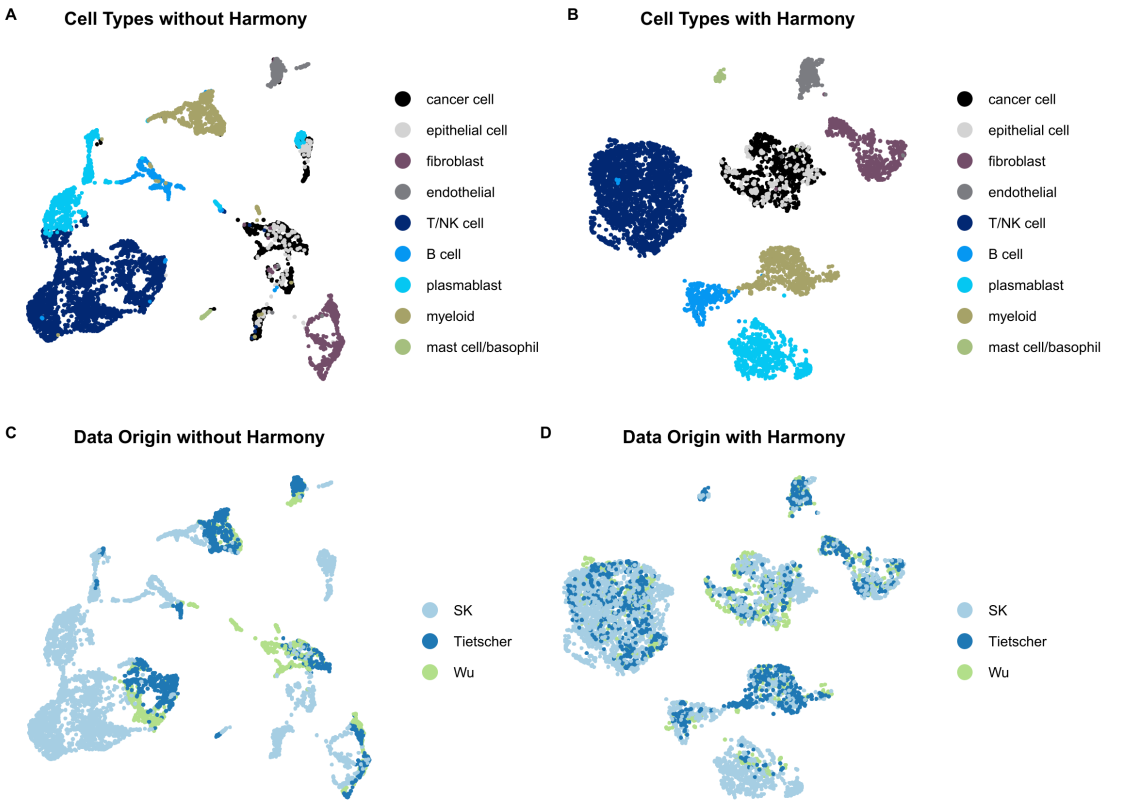


Figure 3.9: Data harmonisation using Harmony. UMAP projections of single-cell RNA sequencing data showing the distribution of various cell types (cancer cells, epithelial cells, fibroblasts, endothelial cells, T/NK cells, B cells, plasmablasts, myeloid cells, and mast cells/basophils) before and after Harmony batch correction.
(A,B) Cell type distributions without and with Harmony, respectively.
(C,D) Data originating from different datasets (SK, Tietscher, Wu) without and with Harmony correction, respectively.

Another approach, that is typically successful for complex datasets is single-cell Variational Inference (scVI) [135]. It is part of a suit of packages implemented in Python called 'scvi-tools', aimed at brining deep-learning functionality to single-cell analysis. In contrast to Harmony and related methods like Canonical Correlation Analysis (CCA), scVI uses a probabilistic deep-learning frame work to learn

a low dimensional embedding of the data. A variational autoencoder (VAE) model is trained to translate the gene expression profile of a cell to a point in a low-dimensional 'latent space' and vice-versa. The advantage of this method is that it can handle technical noise well and explicitly account for batch-related variation. I also tested this approach successfully to harmonise both the batch effects in my single-cell data and public datasets at the same time (Figure 3.10). The data were aligned well in latent space, resulting in a homogenous distribution among the clusters (Figure 3.10, bottom left).

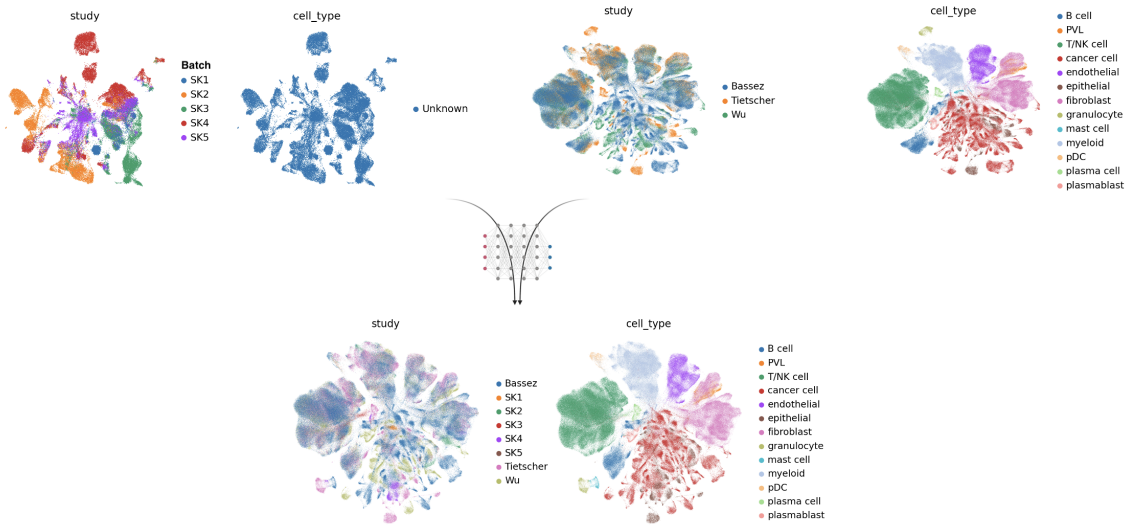


Figure 3.10: Deep Learning-Based Harmonisation. scVI harmonisation of single-cell RNA-seq data across multiple studies and cell types. The upper row displays UMAP visualisations before scVI harmonisation, with cells coloured by study (left) and cell type (right). The lower row shows the UMAP visualisations after applying scVI, with cells coloured by study (left) and cell type (right). scVI successfully integrates the datasets, resulting in a more homogenous distribution among the clusters, effectively reducing batch effects and aligning the data in the latent space.

In addition, cell types could directly be annotated in an automated fashion. This is very advantageous, as it saves a large amount of manual work, to optimally assign cell types to clusters in a new dataset. However, due to its ease of use,

and still very good harmonisation results, I decided to proceed with Harmony and base all downstream analyses on it.

3.3.9 Cell Type Annotation

After deciding to proceed with the Harmony embedding of the data, I still had to annotate cell types to the resulting clusters. To do so, I used canonical markers of the major cell lineages expected in the microenvironment of breast cancer. Each of the major clusters could clearly be assigned to one cell type based on the expression of these marker genes (Figure 3.11).

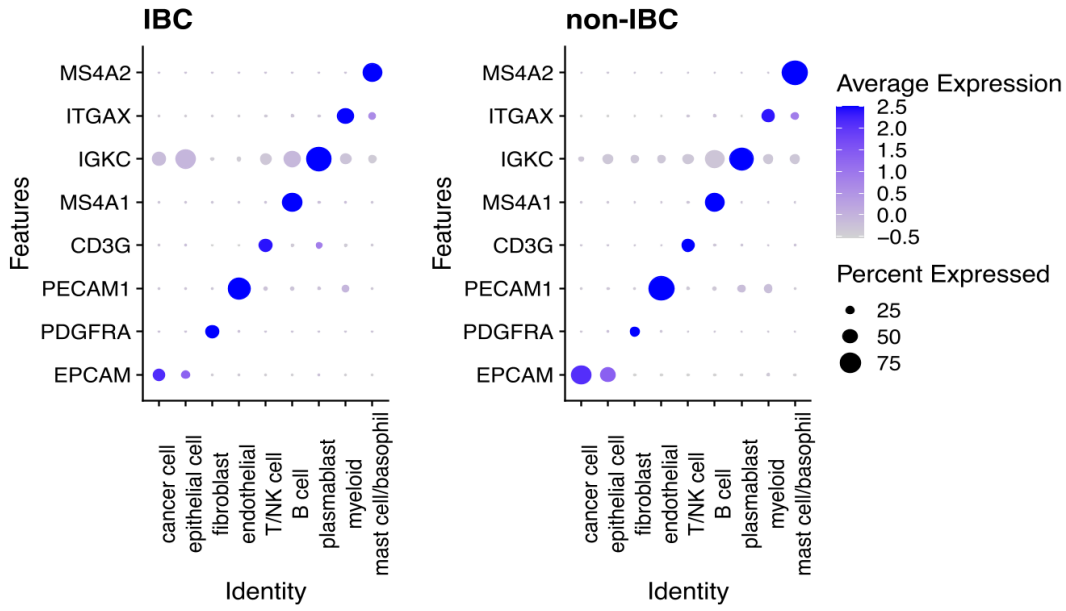


Figure 3.11: Celltype Marker Dotplot. Expression patterns of canonical markers used to assign cell types to clusters in the breast cancer microenvironment. Each marker corresponds to a specific cell type, facilitating the identification of major cell lineages in the dataset.

After annotating the cell types, I also quantified the per-sample percentage of each cell type, relative to the total number of cells recorded for each sample. To

do so, I wrote a custom function. The function filters the dataset to retain only samples with a minimum number of cells, defined by the cutoff parameter. It then calculates the proportions of each cell type per patient by creating a contingency table based on the combination of the metadata column representing the breast cancer type and patient ID and the subclusters.

The calculated proportions are further classified into "IBC" and "non-IBC" categories, based on the clinical classification of each patient. The function then generates a box plot to visualise these proportions, faceted by subcluster, using ggplot2. The plot includes statistical comparisons between the IBC and non-IBC groups using a t-test, with p-values annotated on the plot. This visualisation helps to identify differences in cell type proportions between the two groups (See Materials and Methods).

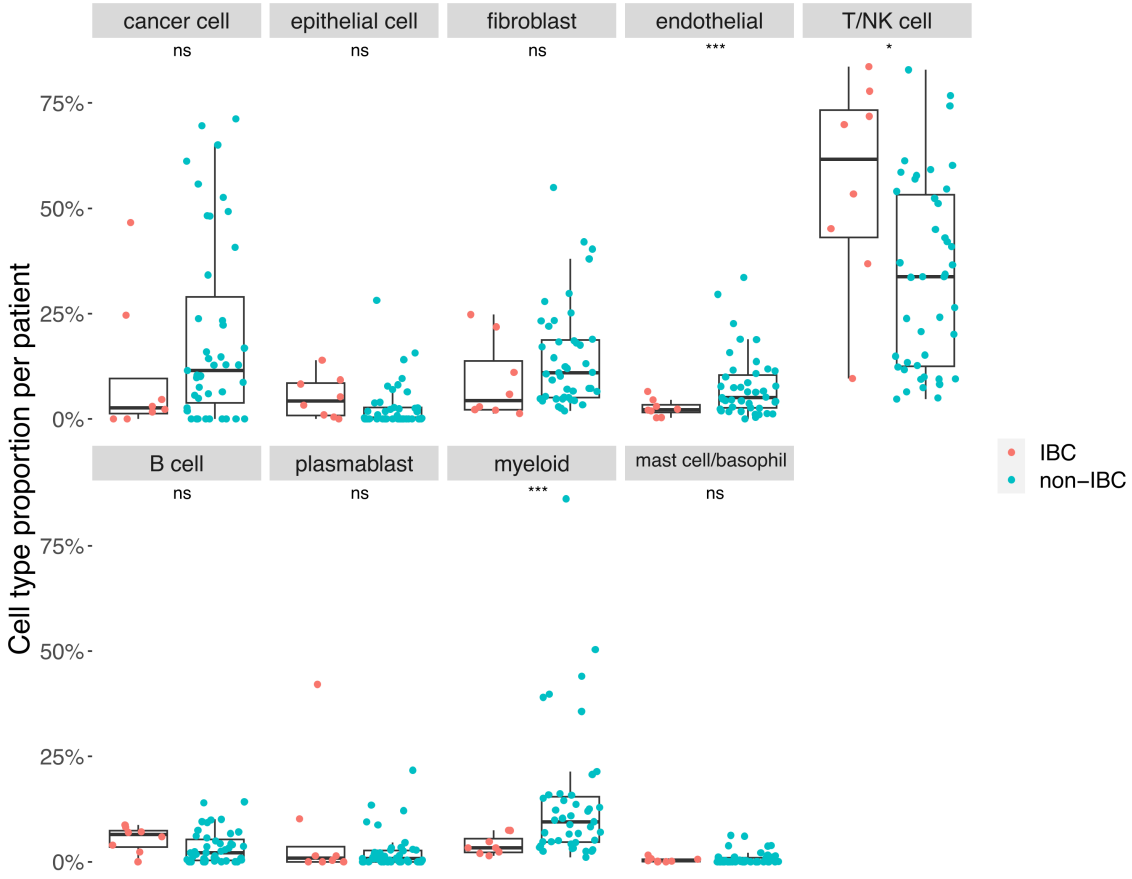


Figure 3.12: Cell Type Proportions. Comparison of cell type proportions between IBC and non-IBC samples. The box plots show the distribution of cell type proportions for various cell types (B cells, plasmablasts, myeloid cells, mast cells/basophils, cancer cells, epithelial cells, fibroblasts, endothelial cells, and T/NK cells) across patients. Statistical significance based on t-tests between IBC and non-IBC groups is indicated ($*p < 0.05$,$*** p < 0.001$, ns = not significant). Each dot represents a patient biopsy, and jittering is applied to improve visibility of overlapping points.

In this case, we see that there are significantly more endothelial and myeloid cells in the non-IBC samples, whereas the IBC samples on average contain more T and NK cells (Figure 3.12). However, the exact sampling procedure of the biopsies happened in the clinic, and is a variable I can't control for. In addition, the

distributions might be skewed by the fact that most of the non-IBC samples where processed by other labs, albeit following similar methods. Because of those differences, I would like to refrain from drawing conclusions from these data. Later on, I will be discussing differences of cell states within these clusters, which I hypothesise should be more robust to sampling biases.

3.3.10 Differentiation of Malignant from Epithelial Cells

A key question when analysing single-cell data of cancer tissue, is how to differentiate malignant from normal epithelial cells. These typically express similar marker genes including EPCAM, but display a high inter-patient variability. While in typical two dimensional embeddings of single-cell data, immune and stromal cells even from different individuals often cluster together, cancer and epithelial cells often form distinct per-patient clusters.

I thus implemented a method previously described by Wu *et al.* [133]. Since almost all breast cancers display copy number variation [136], it is an attractive approach to discern malignant from non-malignant cells. Their approach was to use inferred copy number alteration from inferCNV [137]. InferCNV uses a sliding-window of adjustable size (default: 100 genes), to detect concerted up- or down-regulation. Consecutive deviations from the average expression level of genes in a set of reference cells of the same sample (typically immune cells), spanning entire chromosomes or large portions thereof are then considered copy number events.

My implementation of the method included the following steps (See Materials and Methods): After running inferCNV on all samples using T/NK and myeloid cells as reference group, I added the resulting scores to the Seurat object containing the IBC and non-IBC single-cell data. I then selected the top 5 % most genetically unstable (that is, with the highest inferCNV CNA signal) of cells

of each sample, creating a malignant reference profile by averaging their scores across chromosomes. I then correlated the per-chromosome CNV signals of all other cells to that malignant score. Lastly, I assigned a malignant label to each epithelial/cancer cell that had (1) CNV signal > 0.02, and (2) correlation coefficient with the malignant CNA profile > 0.4. For visualisation purposes, I plotted the results as a UMAP coloured by the malignancy label (Figure 3.13).

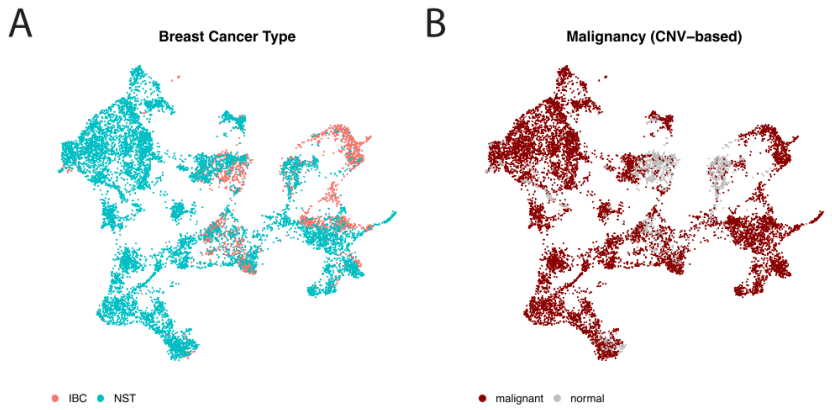


Figure 3.13: Assignment of Malignant Cells. I used a method (See Materials and Methods) described by [133], which is based on inferCNV, to differentiate between malignant and non-malignant epithelial cells.
(A) UMAP embeddings of only the cancer/epithelial cells, labelled by IBC/non-IBC.
(B) UMAP embeddings labelled by malignancy, as assigned by the described method.

3.4 Features of Cell Populations in the Microenvironment of IBC

After assigning each cell to the lineage it belongs to, I went ahead to analyse the makeup and transcriptional states of these cell populations separately. The goal was to find out how they differ between IBC and non-IBC, and if there are any

targetable pathways that could be further investigated to improve therapy options for IBC.

3.4.1 Cancer Associated Fibroblasts

Cancer associated fibroblasts, or CAFs, are a heterogeneous populations of structural cells. As discussed in the introduction, they can adopt different states, with opposing effects on the outcome of the disease [51].

CAF Cell States

To investigate CAFs more closely, I first created a subset of data containing only this type of cells (Figure 3.14A,B). I then proceeded to calculate a new UMAP embedding, using Harmony to correct for any batch effects. Based on the calculated embedding, I ran unsupervised clustering, which resulted in 3 main clusters (Figure 3.14C).
I then used the Seurat-native method of canonical correlation analysis (CCA), to perform a label transfer, using a published single-cell dataset of breast cancer fibroblasts as a reference. To create that dataset, Cords *et al.* [138] created a dataset of 16,000 stromal cells, sourced from breast cancer tissue of 16 individual patients. Based on these data, they defined CAF phenotypes in depth, proposing 9 different classes of cancer-associated fibroblasts and an additional population of pericytes.
Of these, the three main clusters of fibroblasts in our samples are mostly made up of Inflammatory iCAFs, myofibroblastic mCAFs, and vascular vCAFs, with the other phenotypes representing minor populations (Figure 3.14D).

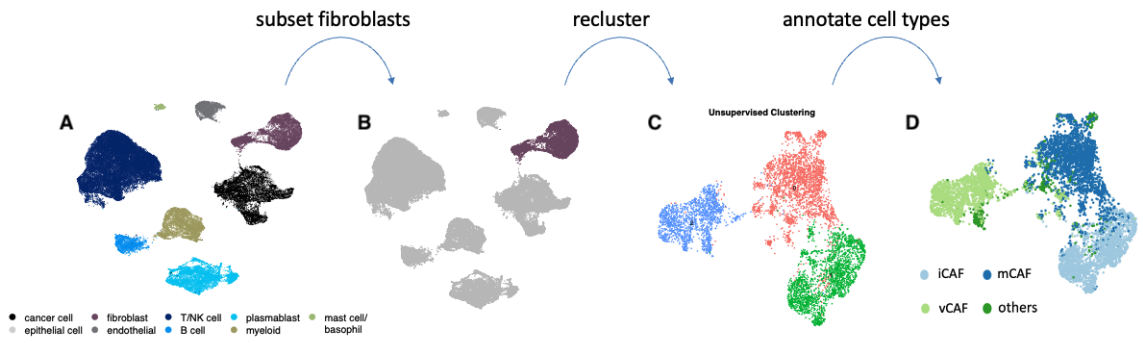


Figure 3.14: Overview of the fibroblast subsetting and reclustering process.
(A) UMAP plot showing the initial clustering of the various cell types within the tumour microenvironment, including fibroblasts, cancer cells, immune cells, and others.
(B) Fibroblasts are highlighted in grey and were subset to create a new dataset.
(C) Reclustering of the subsetted fibroblast population using unsupervised clustering.
(D) Annotated clusters of fibroblast subtypes, including iCAF, mCAF, and vCAFs.

CAF Cell State Proportions

Next, I investigated the proportions of CAF states in both IBC and non-IBC, using the same custom function described in Subsection 3.3.9 (Figure 3.15). Although the median proportion of iCAFs is higher in IBC, this difference was not significant when performing a t-test. On the other hand, the median share of mCAFs of the total CAFs per sample was higher in non-IBC, but also that difference was not significant.

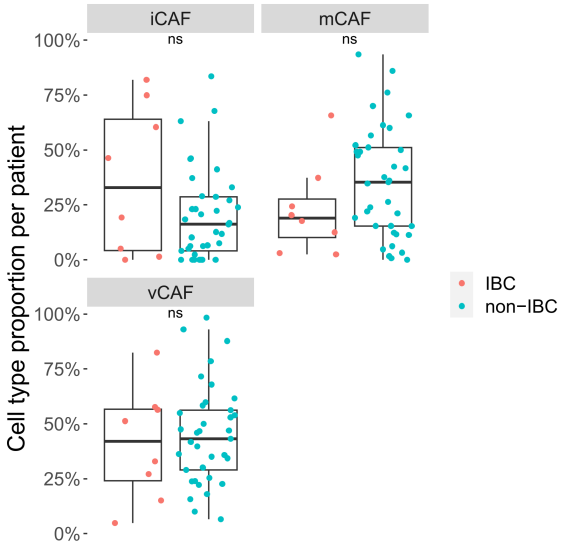


Figure 3.15: Comparison of cell type proportions between IBC and non-IBC samples. The box plots show the distribution of cell type proportions the three main CAF types (iCAFs, mCAFs, vCAFs) across patients. Statistical significance based on t-tests between IBC and non-IBC groups is indicated (ns = not significant). Each dot represents a patient biopsy, and jittering is applied to improve visibility of overlapping points.

CAFs as Signalling Hubs in IBC

In their capacity as structural cells, fibroblasts play an essential role in shaping the microenvironment of cancer [139]. Besides their essential function in shaping and remodelling the extracellular matrix (ECM), they play a key role as signalling hubs and can influence angiogenesis, immune response or evasion, and metastatic progress.

I thus proceeded to investigate cell-cell interactions between subsets of fibroblasts and the rest of the microenvironment using a number of computational and statistical frameworks. Using CellChat, I calculated differential interaction numbers and strengths between subsets of cells in IBC and non-IBC (Figure 3.16).

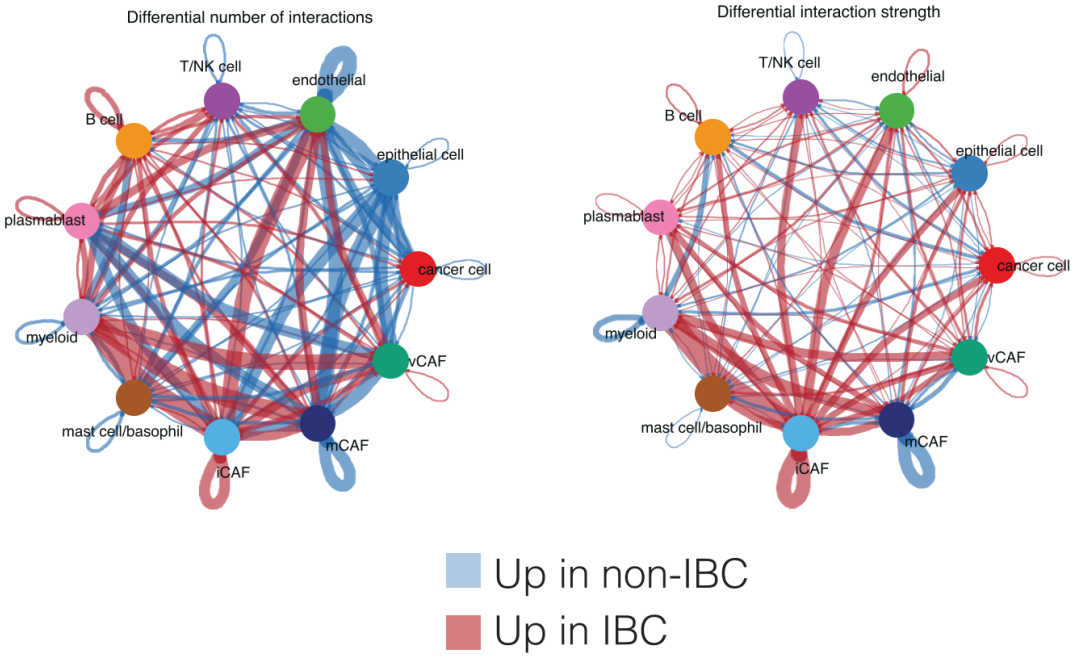


Figure 3.16: Differential interaction numbers and strengths in IBC and non-IBC. These plots show the number and strength of differential interactions between cell types, as predicted by CellChat. Each cell type or CAF state is represented by a node. The edges in between are coloured by whether they occur more frequently or at a higher strength in IBC or non-IBC. Their weight reflects how great this difference is.

Interestingly, the thickest edges in both plots of Figure 3.16 connect myeloid cells and the three CAF populations, hinting at a strong role this interaction may play in IBC compared to non-IBC.

I thus also investigated whether this altered signalling affected mainly incoming or outgoing signalling. Incoming signalling strength is reflected by higher expression of receptors, while increased outgoing signalling would lead to higher expression of ligands.
Still using CellChat, I plotted a scatter plot of incoming and outgoing interaction

strength in IBC and non-IBC (Figure 3.17).

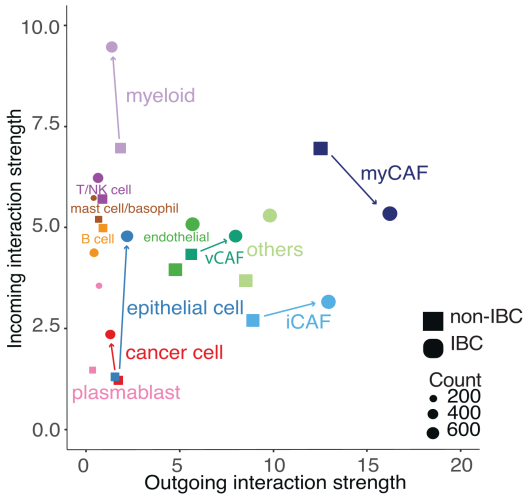


Figure 3.17: Outgoing and incoming interaction strength in IBC and non-IBC The scatter plot shows the interaction strength, with colours for each cell type and CAF state. Squares represent values for non-IBC and circles for IBC, and their size is proportional to the number of predicted interactions.

The x-axis represents the outgoing interaction strength, that is, a measure of how high the level of expression of the ligand in a predicted ligand-receptor pair is. The y-axis on the other hand shows the incoming interaction strength of each cell type, meaning how strongly the receptor of a predicted ligand-receptor pair is expressed.
Interestingly, mostly the outgoing interaction strength is increased in CAF populations of IBC as compared to non-IBC. Contrastingly, epithelial, cancer and myeloid cells seem to be mostly on the receiving end of the heightened signalling.

Signalling Pathways of CAF Populations in IBC

To further investigate which pathways may be affected, I used a combination of the tools Liana and NicheNet.

Liana [140] is a ligand-receptor analysis framework, that attempts to build on existing methods and prior-knowledge databases including CellPhoneDB, CellChatDB and others. It attempts to increase confidence by integrating the results of those databases and providing an aggregate score.
In addition, NicheNet considers downstream signalling pathways of receptors and their target genes as well. Thereby it tries to obtain more meaningful predictions of relevant ligand-receptor interactions.

Because the crosstalk from all three major CAF populations into the myeloid compartment was predicted to be upregulated, I designed the analysis using each of the CAF types as sender and holding myeloid cells constant as receivers. I investigated the downstream activation of multiple pathways (data not shown). I was however most interested to find out to what extent CAFs shape the inflammation-like phenotype observed in IBC. Thus I used the Hallmark Inflammatory Response gene set from MSigDB [141] as downstream pathway.
With NicheNet, I then compared the expression of genes contained in the Hallmark Inflammatory Response gene set to the average expression of background genes (that is, genes with measurable expression in at least 5% of cells). A pearson correlation coefficient was then calculated for the expression of ligand and receptor pairs and downstream upregulation of that pathway.

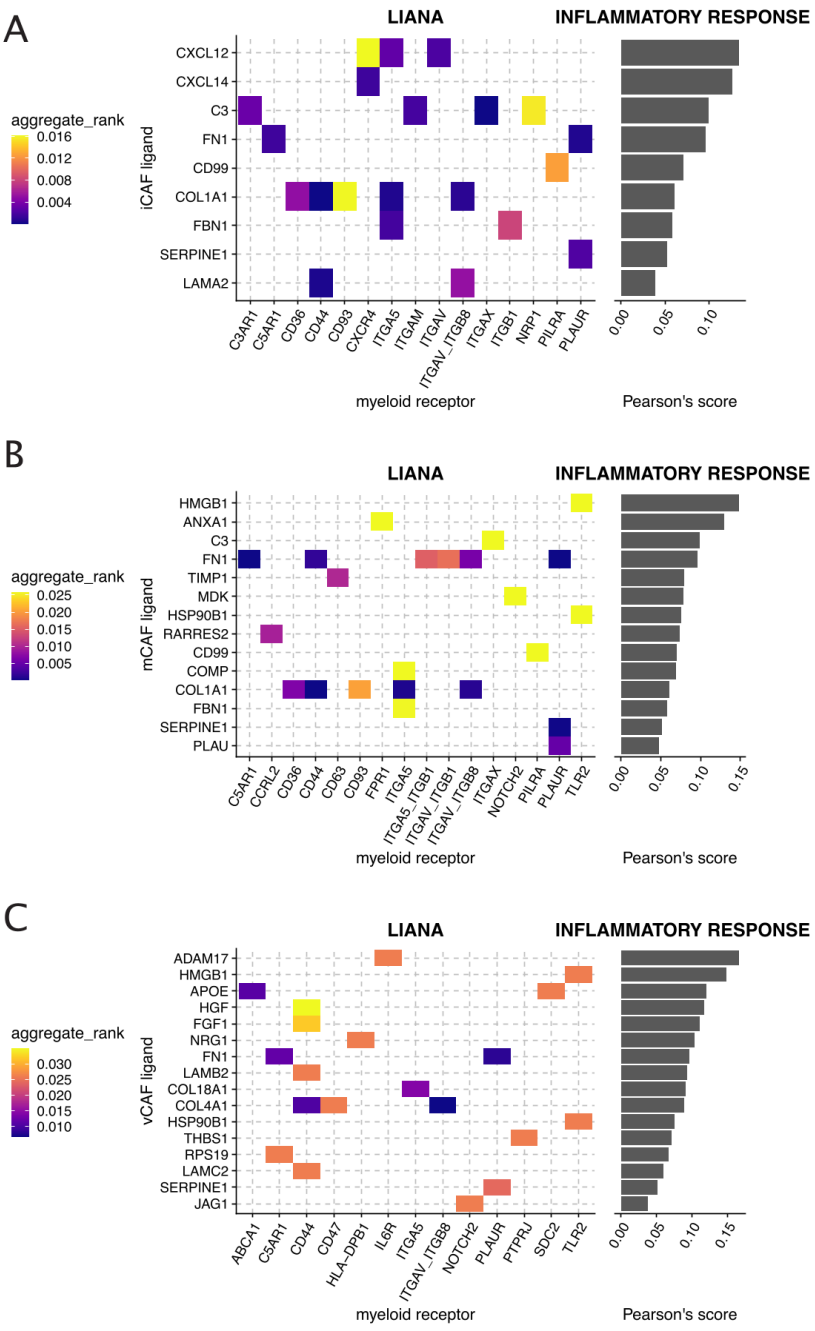

A
LIANA
INFLAMMATORY RESPONSE
aggregate_rank
0.016
0.012
0.008
0.004
iCAF ligand
CXCL12
CXCL14
C3
FN1
CD99
COL1A1
FBN1
SERPINE1
LAMA2
C3AR1
C5AR1
CD36
CD44
CD93
CXCR4
ITGA5
ITGAM
ITGAV
ITGAV_ITGB8
ITGAX
ITGB1
NRP1
PILRA
PLAUR
0.00
0.05
0.10
myeloid receptor
Pearson's score
B
LIANA
INFLAMMATORY RESPONSE
aggregate_rank
0.025
0.020
0.015
0.010
0.005
mCAF ligand
HMGB1
ANXA1
C3
FN1
TIMP1
MDK
HSP90B1
RARRES2
CD99
COMP
COL1A1
FBN1
SERPINE1
PLAU
C5AR1
CCRL2
CD36
CD44
CD63
CD93
FPR1
ITGA5
ITGA5_ITGB1
ITGAV_ITGB1
ITGAV_ITGB8
ITGAX
NOTCH2
PILRA
PLAUR
TLR2
0.00
0.05
0.10
0.15
myeloid receptor
Pearson's score
C
LIANA
INFLAMMATORY RESPONSE
aggregate_rank
0.030
0.025
0.020
0.015
0.010
vCAF ligand
ADAM17
HMGB1
APOE
HGF
FGF1
NRG1
FN1
LAMB2
COL18A1
COL4A1
HSP90B1
THBS1
RPS19
LAMC2
SERPINE1
JAG1
ABCA1
C5AR1
CD44
CD47
HLA-DPB1
IL6R
ITGA5
ITGAV_ITGB8
NOTCH2
PLAUR
PTPRJ
SDC2
TLR2
0.00
0.05
0.10
0.15
myeloid receptor
Pearson's score

Figure 3.18: CAF signalling induces an inflammatory response in myeloid cells Ligand-receptor analysis based on a combination of Liana and NicheNet of the major CAF types. The most significant signalling interactions according to an aggregate of multiple prior knowledge databases (Liana) were ranked by the activity of their downstream pathways (NicheNet) and plotted (see Materials and Methods). This was done for each of the CAF subsets with myeloid cells as the receiver cell type and the Hallmark Inflammatory Response [142] pathway for the NicheNet analysis of downstream signalling.

Interestingly, my findings indicate that iCAFs, or inflammatory CAFs, a subtype of CAFs, may stimulate an inflammatory response in the myeloid compartment of inflammatory breast cancer (Figure 3.18 A). They do so by releasing chemokines, including various CXCLs, of which the expression together with the expression of cognate receptors like CXCR14 correlates with an upregulation of inflammatory response genes in myeloid cells. This could cause an inflammatory reaction in the IBC microenvironment, which can further support tumour growth and immune evasion, making it a critical focus for potential treatments.

In addition, mCAFs and vCAFs are predicted to have a signalling function inducing an inflammatory response in myeloid cells as well. One major pathway this is predicted to occur by is the HMGB1-TLR2 axis (Figure 3.18 B, C). This pathway is crucial for the activation of the innate immune response, and have been linked to an assembly of the inflammasome, pyroptosis and ferroptosis [143].

3.4.2 Myeloid Cells

Myeloid cells are a diverse group of immune cells. They share a common origin from myeloid progenitors in the bone marrow. Amongst them are monocytes, dendritic cells, and macrophages. They are crucial for mounting an innate immune response of the organism against invading pathogens. Additionally, the serve a

bridging function. By processing and presenting antigens to cells of the lymphoid lineage, they help to engage the adaptive immune system as well.

Myeloid Cell Types and States

To investigate myeloid cells more closely, I first created a subset of data containing only this type of cell (Figure 3.19 A, B). I then proceeded to calculate a new UMAP embedding, using Harmony to correct for any batch effects.
I then utilised Seurat's canonical correlation analysis (CCA) method to perform a label transfer, using a published single-cell dataset of myeloid cells as a reference. We used an extensively annotated reference data set [144] by Mulder *et al.*. In a meta-analysis, they curated data from 41 datasets, collecting a total of 178,651 mononuclear phagocytes (MNPs) from 13 different tissues. Based on these data, they identified various myeloid phenotypes, describing key functional states.
In our analysis, the three main clusters of myeloid cells primarily consist of IL1B Monocytes, IL4I1 Macrophages, and general macrophages, with additional minor populations present (Figure 3.19 C).

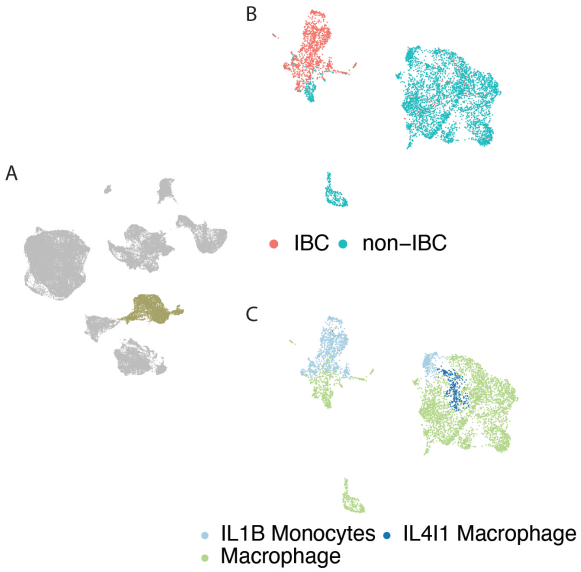


Figure 3.19: Overview of the Myeloid Subsetting and Reclustering Process
(A) UMAP plot showing the initial clustering of the various cell types within the tumour microenvironment, including fibroblasts, cancer cells, immune cells, and others. Myeloid cells are highlighted and were subset to create a new dataset.
(B) Batch-corrected embedding of the subsetted myeloid population using unsupervised clustering and coloured by IBC and non-IBC patients.
(C) Annotated clusters of myeloid subtypes, including IL1B Monocytes, IL4I1 Macrophages and Macrophages.

Myeloid Cell State Proportions

Next, I investigated the proportions of myeloid cell states in both IBC and non-IBC, using the same custom function described in Subsection 3.3.9 (Figure 3.20). In line with the distribution of IBC cells in the UMAP embedding (Figure 3.19 B, C), the median proportion of IL1B Monocytes is significantly higher in IBC cells, whereas non-IBC myeloid cells are mostly comprised of IL4I1 Macrophages and general macrophages.

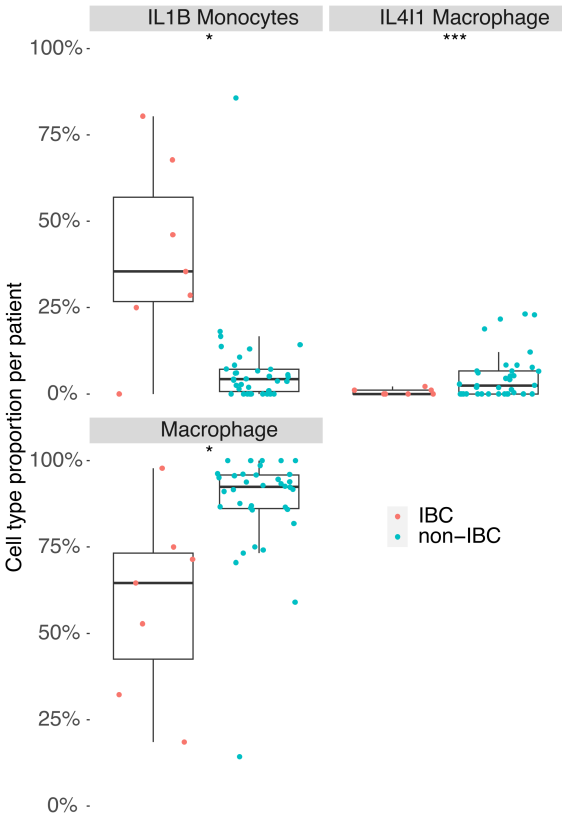


Figure 3.20: Comparison of cell type proportions between IBC and non-IBC samples. The box plots show the distribution of cell type proportions the three main myeloid cell types (IL1B Monocytes, IL4I1 Macrophages and Macrophages) across patients. Statistical significance based on t-tests between IBC and non-IBC groups is indicated ($*$ p < 0.05, $**$ p < 0.01,$***$ p < 0.001, ns = not significant). Each dot represents a patient biopsy, and jittering is applied to improve visibility of overlapping points.

IL1B Monocytes

The major subcluster of myeloid cells in IBC has been identified as IL1B monocytes. To define their properties and function better, I ran EnrichR Gene Set Enrichment Analysis (EnrichR GSEA) on this cluster of cells, comparing them to the remaining monocytes (Figure 3.21).

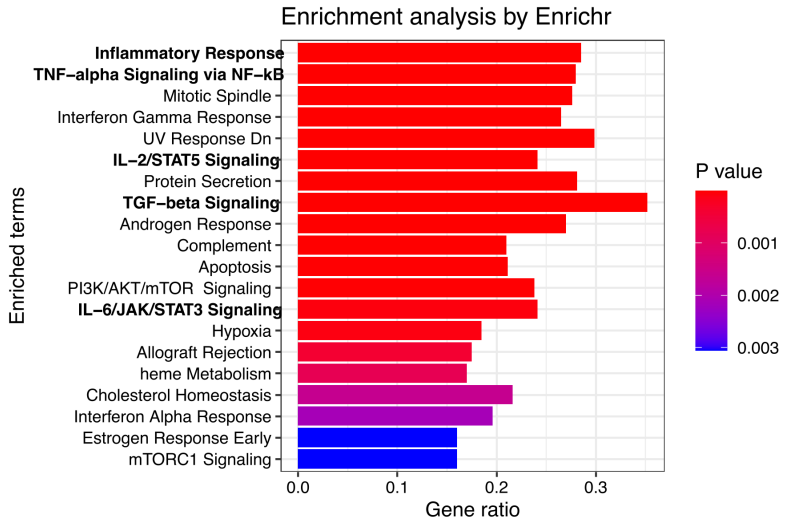


Figure 3.21: Properties of IL1B Monocytes EnrichR gene set enrichement analysis of IL1B monocytes, compared to remaining myeloid cells. Besides inflammatory response and NF-κB pathways, they upregulate immune-suppressive pathways (in bold). The x-axis represents the ratio of genes within the given pathway that are upregulated.

Interestingly, they most significantly upregulate inflammatory response genes and genes belonging to the TNF-α signalling via NF-κB. This is in line with the previous finding that CAFs induce an inflammatory response in myeloid cells in IBC. Another interesting finding is the fact that this main myeloid population in IBC has been labelled IL1B monocytes by comparison to the reference dataset. IL1-β is a pro-inflammatory cytokine and could be responsible for some of these effects [145].

Corroborating their inflammatory identity, they also express key regulators of the inflammatory response. After being primed by pathogen-associated molecular patterns (PAMPs), TNF or IL-1β signalling, the NLRP3 inflammsome can be activated by a second signal, which can be one of many pathogen- or damage-associated molecular patterns (DAMPs). This leads to its assembly, which in turn

can lead to pyroptosis and the release of inflammatory cytokines, including IL-1β. NF-κB is the master regulator of inflammation and mediates various associated effects, including the activation of pro-inflammatory genes, the production of cytokines, as well as proliferation, differentiation and angiogenesis [146].
In myeloid cells of IBC patients, both genes are significantly more expressed than in non-IBC cells (Figure 3.22).

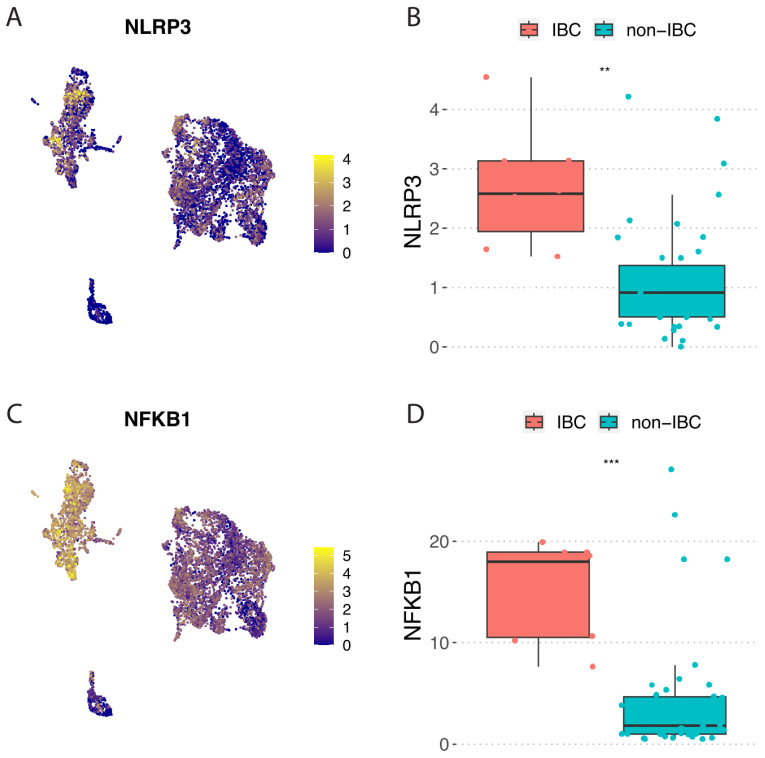


Figure 3.22: IBC Myeloid Cells Express Key Inflammation Regulators
(A) Feature plot with cells in a UMAP embedding, coloured by their level of expression of NLRP3.
(B) Box- and scatter plots of mean NLRP3 expression per patient within the monocyte/macrophage compartment.
(C) Feature plot with cells in a UMAP embedding, coloured by their level of expression of NFKB1.
(D) Box- and scatter plots of mean NFKB1 expression per patient within the monocyte/macrophage compartment.

Myeloid Ligand Receptor Analysis

Previously, we discussed how various CAF populations may induce an inflammatory response in myeloid cells. The above observation that the inflammasome might also be activated, could be explained by HMGB1 and HSP90 signalling

from IBC cells to Toll-Like-Receptors in IL1B Monocytes (Figure 3.23 A).

In the context of possible therapy approaches, it is of great interest how the the immune response is shaped in the IBC microenvironment. We found that IBC monocytes express certain ligands like INHBA and IL1B, that potentially downregulate T cell immunity by activating the TGF-β signalling pathway [147, 148] (Figure 3.23 B). This could explain why the immune system often struggles to mount an effective response against IBC, highlighting a potential target for therapy.

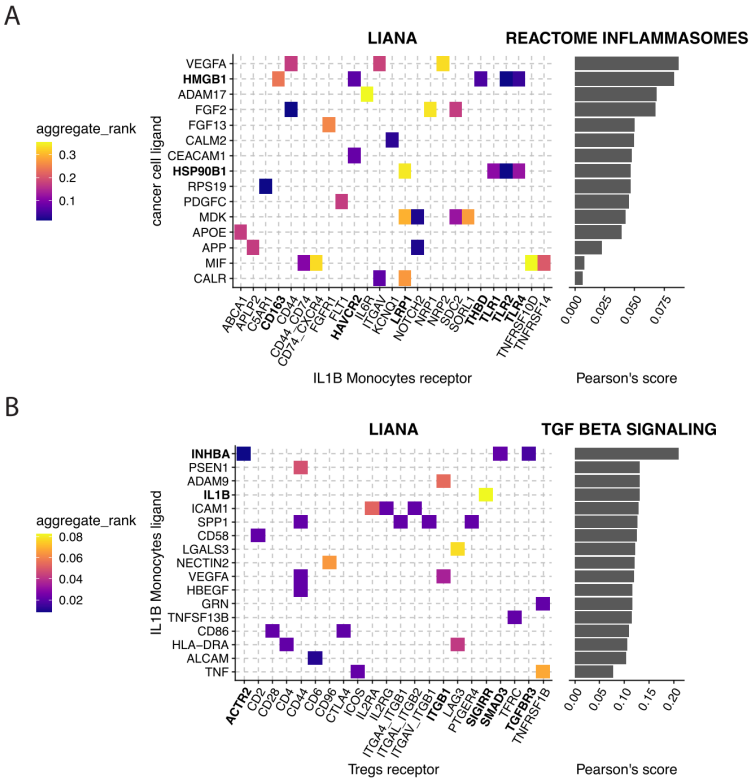


Figure 3.23: The Cancer Cell - IL1B Monocyte - Treg Signalling Axis Ligand-receptor analysis based on a combination of Liana and NicheNet of the major CAF types. The most significant signalling interactions according to an aggregate of multiple prior knowledge databases (Liana) were ranked by the activity of their downstream pathways (NicheNet) and plotted (see Materials and Methods).
(A) Ligands expressed by cancer cells, that could induce the 'Reactome Inflammasomes' pathway via the plotted receptors in IL1B Monocytes.
(B) Ligands expressed by IL1B Monocytes, that could induce the 'TGF-β Signalling' pathway via the plotted receptors in Tregs.

3.4.3 T/NK Cells

T cells and NK cells both arise from the lymphoid lineage of the haematopoietic system. While T cells are key players in the adaptive immune system, specifically in cellular immunity, NK cells serve an essential role in the innate immune

response, though they share some functional similarities with cytotoxic T cells.

T and NK Cell Types and States

In single-cell transcriptomic data from tumour tissues, T cells and NK cells often form a single cluster. At first glance, this may seem unexpected, given their distinct biology, functions, and marker expression. T cells are a critical component of adaptive immunity, while NK cells operate within the innate immune system. Despite these differences, both cell types share a cytotoxic function and express overlapping genes involved in this role, such as perforin and granzyme. As a result, their transcriptomic profiles can appear similar at a broader resolution.

However, they can be differentiated by the specific markers they express—T cells typically express CD3 (with subtypes expressing CD4 or CD8), while NK cells are characterised by markers such as CD56 and NKp46. Therefore, it is necessary to analyse their transcriptomic profiles in a more granular manner to accurately distinguish between these cell types and to explore their distinct subpopulations and activation states.

To investigate T and NK cells with sufficient granularity, I first created a subset of data containing only these types of cells (Figure 3.24 A, B). I then proceeded to calculate a new UMAP embedding, using Harmony to correct for any batch effects. I then utilised Seurat's canonical correlation analysis (CCA) method to perform a label transfer, using a published single-cell dataset as a reference. The dataset used is by Tietscher *et al.* [134]. They performed a detailed comparison of breast cancer with and without exhausted T cells. To that end, they measured a total of 120,000 cells from 14 tissue samples. Based on these data, they created a detailed mapping of T and NK cell states in the breast cancer microenvironment. I was able to map functional states to the T and NK cluster using this dataset as a reference (Figure 3.24 C).

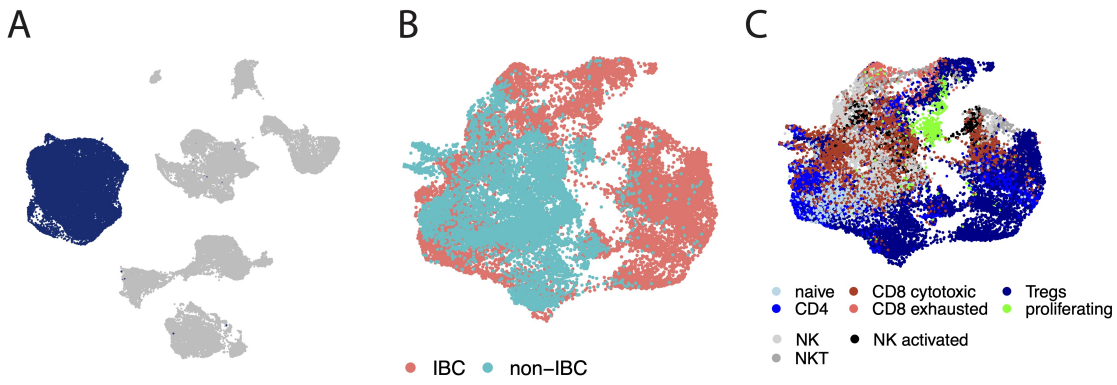


Figure 3.24: Overview of the Myeloid Subsetting and Reclustering Process
(A) UMAP plot showing the initial clustering of the various cell types within the tumour microenvironment, including fibroblasts, cancer cells, immune cells, and others. T and NK cells are highlighted and were subset to create a new dataset.
(B) Batch-corrected embedding of the subsetted T/NK population using unsupervised clustering and coloured by IBC and non-IBC patients.
(C) Annotated clusters of T/NK cell types.

To ensure accurate cell type assignments, I generated a heatmap displaying the normalised counts of key marker genes across the assigned clusters. The markers used are specific to different T and NK cell subpopulations, and their expression patterns are expected to align with the cluster labels.

From the heatmap, it is clear that CD3E, CD3G, and CD3D are strongly expressed in T cell clusters, confirming the presence of CD3+ T cells in these clusters. Specifically, CD4 is predominantly expressed in clusters corresponding to helper T cells, while CD8A shows high expression in cytotoxic T cell clusters, validating the identification of these subpopulations.

NK cells are clearly distinguished by the expression of NCAM1 (CD56) and NCR1, with high expression of GNLY (granulysin) confirming their cytotoxic nature. Notably, NK cell activation states are identifiable, as clusters marked by NCR1 and increased GZMA and PRF1 (perforin) expression represent activated NK cells, further differentiating them.

Clusters labelled as exhausted T cells are characterised by increased levels of TOX, while FOXP3 expression clearly distinguishes regulatory T cells (Tregs). Proliferating T cells are marked by strong expression of MKI67 and HIST1H4C, indicating active cell division.

These findings support that the marker gene expression profiles align well with the assigned cluster labels, supporting the accuracy of the cell type classification.

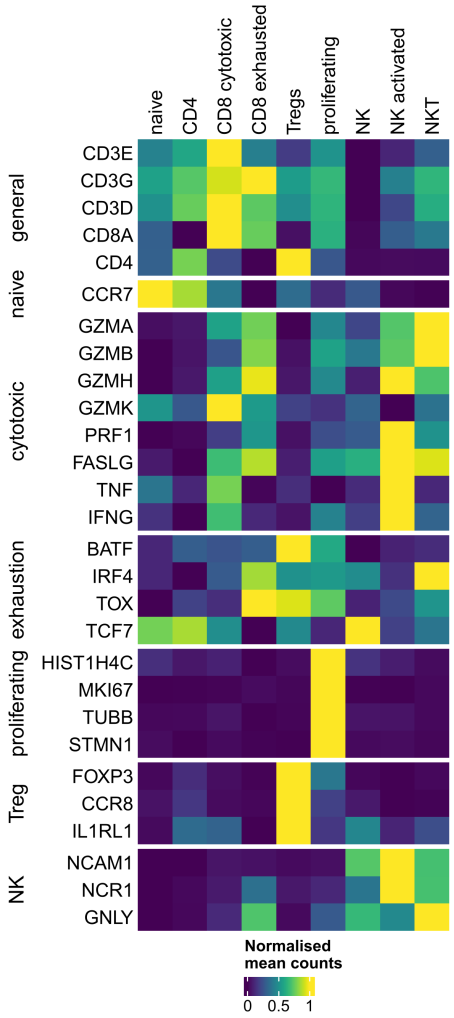


Figure 3.25: Heatmap of normalised mean counts for key marker genes across T and NK cell clusters. Distinct marker expression patterns validate the identity of major subpopulations, including CD4+ helper T cells, CD8+ cytotoxic T cells, NK cells, and regulatory T cells, as well as specific states such as exhaustion and proliferation.

T and NK Cell State Proportions

Continuing the analysis of T and NK cell types, I also compared the relative frequencies of these immune cells in the tumour microenvironment of IBC versus

non-IBC). Figure 3.26 depicts the proportions of key T and NK cell subtypes in individual patients from both cancer types, highlighting potential differences in immune cell distribution.

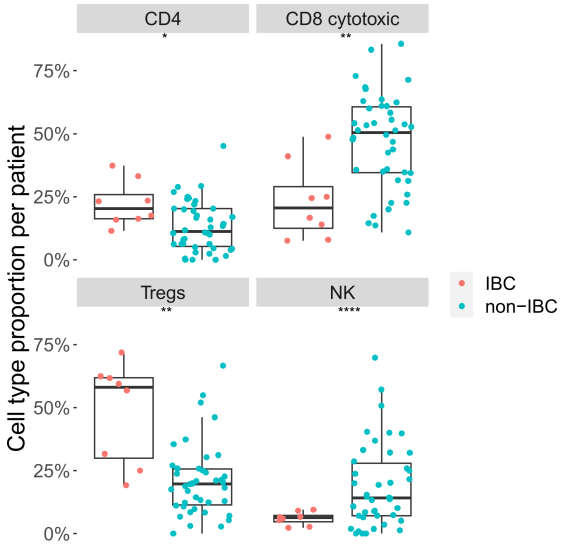


Figure 3.26: Comparison of T and NK cell type proportions between IBC and non-IBC samples. Box plots comparing the relative frequencies of Tregs, NK cells, CD4+ T cells, and CD8+ cytotoxic T cells in the T/NK compartment of IBC and non-IBC patients. IBC shows a higher proportion of Tregs and NK cells, while non-IBC is characterised by a greater presence of cytotoxic CD8+ T cells, suggesting differences in immune response dynamics between the two cancer types. Statistical significance based on t-tests between IBC and non-IBC groups is indicated ($*$ $p < 0.05$, $**$ $p < 0.01$,$***$ $p < 0.001$, ns = not significant). Each dot represents a patient biopsy, and jittering is applied to improve visibility of overlapping points.

The data suggest a striking contrast between IBC and non-IBC in terms of immune cell composition. Specifically, regulatory T cells (Tregs) and NK cells appear more frequent in IBC patients, as indicated by their increased proportions. This elevated presence of Tregs could suggest a more immunosuppressive microenvironment in IBC, potentially aiding tumour evasion by downregulating immune responses. The higher proportion of NK cells, particularly if activated, may represent an immune

attempt to counteract the tumour's aggressive nature, though their efficacy might be limited by the suppressive environment facilitated by Tregs.
Conversely, cytotoxic CD8+ T cells, essential for anti-tumour immunity, are more abundant in non-IBC patients. Their relatively reduced frequency in IBC suggests a diminished cytotoxic response in this more aggressive cancer subtype, possibly contributing to the immune escape mechanisms seen in IBC. CD4+ T cells, which are involved in coordinating immune responses, are distributed similarly across both cancer types, although their exact functional state (e.g., helper or exhausted) requires further investigation to draw definitive conclusions.
These differences in T and NK cell frequencies between IBC and non-IBC hint at the complex immune landscape of these cancers. The relative scarcity of cytotoxic T cells and the abundance of Tregs in IBC may contribute to its poorer prognosis by promoting immune tolerance, while the presence of NK cells might reflect ongoing immune surveillance, albeit potentially ineffective in the context of a suppressive tumour microenvironment.

T Cells Display Increased Exhaustion in IBC

To further characterise the functional state of cells previously labelled as 'exhausted' by reference mapping, and compare them between IBC and non-IBC, I assigned a module score to each cell based on a previously published exhaustion gene [149] using Seurat. This scoring allows for a quantitative assessment of the exhaustion levels in these cells, which is crucial for understanding their potential functional impairment in the tumour microenvironment.
The comparison of exhaustion scores between IBC and non-IBC patients reveals a notable difference in the extent of T cell exhaustion. As shown in Figure 3.27, cells in IBC exhibit significantly higher exhaustion scores compared to those in non-IBC. In addition, the number of cells labelled as exhausted is also much higher in IBC as compared to non-IBC.
This increased number of exhausted cells with an elevated exhaustion score in

IBC suggests that T cells in this context are more likely to be functionally impaired, limiting their ability to mount an effective anti-tumour response. Exhausted T cells are characterised by reduced proliferative capacity, diminished cytotoxic activity, and altered cytokine production, all of which can contribute to immune evasion by the tumour.

The higher exhaustion levels in IBC cells could be driven by chronic antigen exposure and persistent inflammatory signals within the tumour microenvironment, which are known factors promoting T cell exhaustion. This state of exhaustion likely plays a critical role in the aggressive nature of IBC, as it hampers the body's natural ability to control tumour progression through immune surveillance. On the other hand, the relatively lower exhaustion scores in non-IBC suggest a more robust and potentially effective T cell-mediated immune response, aligning with the observed higher proportions of cytotoxic CD8+ T cells in non-IBC patients.

These findings underscore the importance of targeting T cell exhaustion as a therapeutic strategy, particularly in IBC, where reversing exhaustion could restore T cell functionality and enhance anti-tumour immunity.

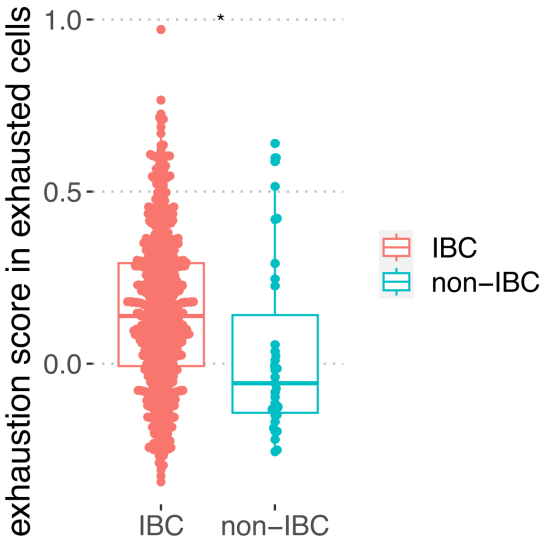


Figure 3.27: IBC T Cells Show More Pronounced Exhaustion. Box plot comparing exhaustion scores of T cells in IBC and non-IBC patients, calculated using a gene set from the referenced study with Seurat. The figure demonstrates higher exhaustion scores in T cells from IBC patients, indicating a more pronounced functional impairment compared to non-IBC, potentially contributing to the aggressive tumour behaviour observed in IBC. Statistical significance based on t-tests between IBC and non-IBC groups is indicated ($*$ $p < 0.05$). Each dot represents one cell previously labelled as exhausted, and jittering is applied to improve visibility of overlapping points.

Immune suppression in the IBC Microenvironment

I investigated the expression levels of the TGFB1 gene across the entire dataset, encompassing all cell types. TGFB1 is known to play a significant role in immune regulation and tumour progression by promoting an immunosuppressive microenvironment, which can lead to T cell exhaustion and hinder effective anti-tumour immune responses.

The Figure 3.28 displays the expression of TGFB1 mapped onto a UMAP projection of the dataset, revealing that TGFB1 expression is widespread but varies in intensity across different cell types. Notably, higher expression levels of TGFB1

are observed in areas predominantly occupied by T/NK, B, myeloid cells and plasmablasts, and comparatively at higher intensities in IBC. This pattern suggests that TGFB1 might be contributing to the more immunosuppressive and aggressive phenotype observed in IBC by influencing the tumour microenvironment to favour immune evasion mechanisms.

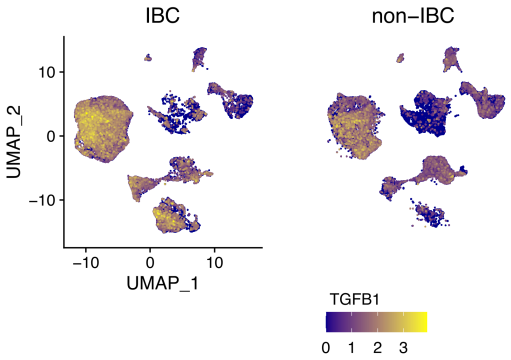


Figure 3.28: TGFβ Expression in the Microenvironments of IBC and non-IBC UMAP feature plot showing the expression levels of TGFB1 across all cell types in the dataset. The plot suggests higher TGFB1 expression in IBC, suggesting a role in creating an immunosuppressive tumour microenvironment.

To quantify the expression differences of TGFB1 between IBC and non-IBC, I generated a box plot (Figure 3.29), showing the distribution of TGFB1 pseudobulk expression levels per patient across both groups. The box plot indicates significantly higher TGFB1 expression in IBC patients compared to non-IBC, aligning with the observed higher levels of immune cell exhaustion and increased frequency of regulatory T cells in IBC. This suggests that TGFB1 may play a pivotal role in mediating the immunosuppressive conditions seen in IBC, potentially driving the functional impairment of immune cells and contributing to the overall tumour aggressiveness.

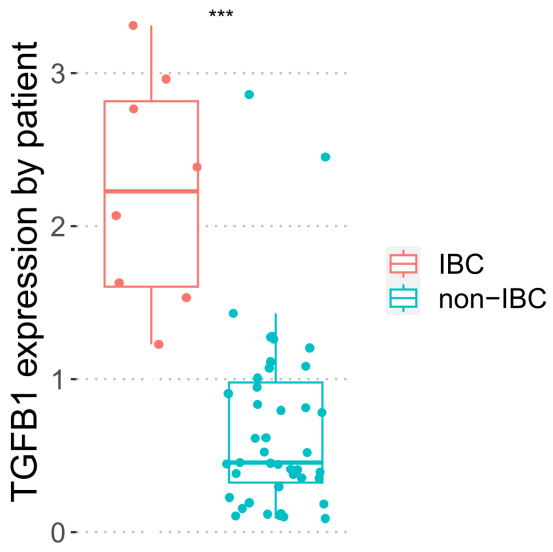


Figure 3.29: TGFβ Is More Highly Expressed in IBC Box plot comparing TGFB1 expression levels between IBC and non-IBC patients, demonstrating significantly higher expression in IBC. This elevated expression of TGFB1 in IBC is consistent with the observed increase in immune suppression and cell exhaustion, implicating TGFB1 as a key factor in the aggressive nature of IBC.

These findings highlight the importance of TGFB1 as a potential target for therapeutic intervention, aiming to mitigate its immunosuppressive effects and restore effective anti-tumour immunity in IBC patients.

3.4.4 Endothelial Cells

A key feature of inflammatory breast cancer is increased angiogenesis. Angiogenesis is a crucial hallmark of tumours [150], enabling their progression by supplying them with the necessary nutrients to meet their intense need for rapid growth and survival.

The clinical IBC diagnosis is not always correlated to histopathological features, but if any, blood vessel sprouting is the most likely feature to be observed by a

pathologist. I thus attempted to investigate the molecular pathways driving angiogenesis in IBC.

Endothelial Cell Types

There are two main types of endothelial cells: LYVE1$^+$ lymphatic endothelial cells, and blood endothelial cells, which are typically marked by CD31 and VE-cadherin. LYVE1$^+$ lymphatic endothelial cells are involved in the maintenance and function of the lymphatic system, while blood endothelial cells line the blood vessels and are crucial for vascular homeostasis, angiogenesis, and the regulation of blood flow and permeability [151].

To investigate endothelial cells, I thus proceeded to subset the IBC and non-IBC dataset (Figure 3.30 A). I then calculated a new UMAP embedding, addressing batch- and study-related variation with Harmony (Figure 3.30 B). I then used the distinct marker genes LYVE1 and PECAM1 to assign lymphatic and vascular endothelial cells to the two very distinct clusters (Figure 3.30 C, D).

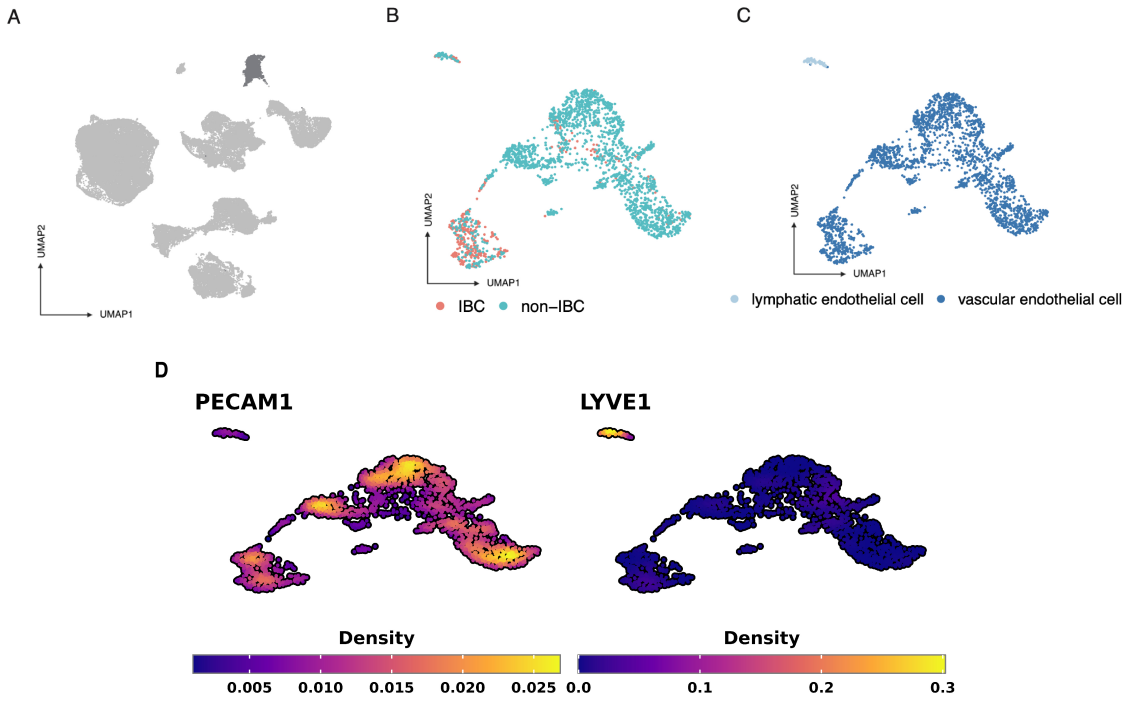


Figure 3.30: Subsetting and Cell Type Assignment of Endothelial Cells.
(A) Endothelial cells (highlighted) were subset from the whole dataset, creating an endothelial-only subset.
(B) A new UMAP embedding was calculated, addressing batch-effects with Harmony. IBC vascular endothelial cells occupy a specific space and likely differ in their transcriptomic profile.
(C) Annotation of endothelial cells with lymphatic and vascular subtypes. The marker genes LYVE1 and PECAM1. Light blue dots represent lymphatic endothelial cells, and dark blue dots represent vascular endothelial cells, highlighting the very distinct clustering of these subtypes within the dataset.
(D) Expression density of marker genes overlayed on UMAP embedding (nebulosa plot), showing the distinct expression of marker genes used to annotate clusters.

Interestingly, IBC blood endothelial cells occupy a distinct space within that cell cluster, hinting at an altered transcriptional state.

Unfortunately, only very few lymphatic endothelial cells, especially of IBC origin, were contained in the dataset. Since invasion and colonisation of lymphatic vessels is the causative hallmark resulting in the clinical presentation of IBC, it would have been of high interest to investigate this population of cells further. This however does not seem possible with 24 IBC and 44 non-IBC endothelial cells included in the dataset. Thus, I focussed on vascular endothelial cells, and the molecular pathways that may cause the angiogenic phenotype of IBC.

Ligand Receptor Interaction Stimulating Angiogenesis in IBC

Angiogenesis can be stimulated externally, by factors like VEGFs and FGFs.
I thus performed ligand-receptor analysis using the combination of Liana and Nichenet used and described before (Figure 3.31).

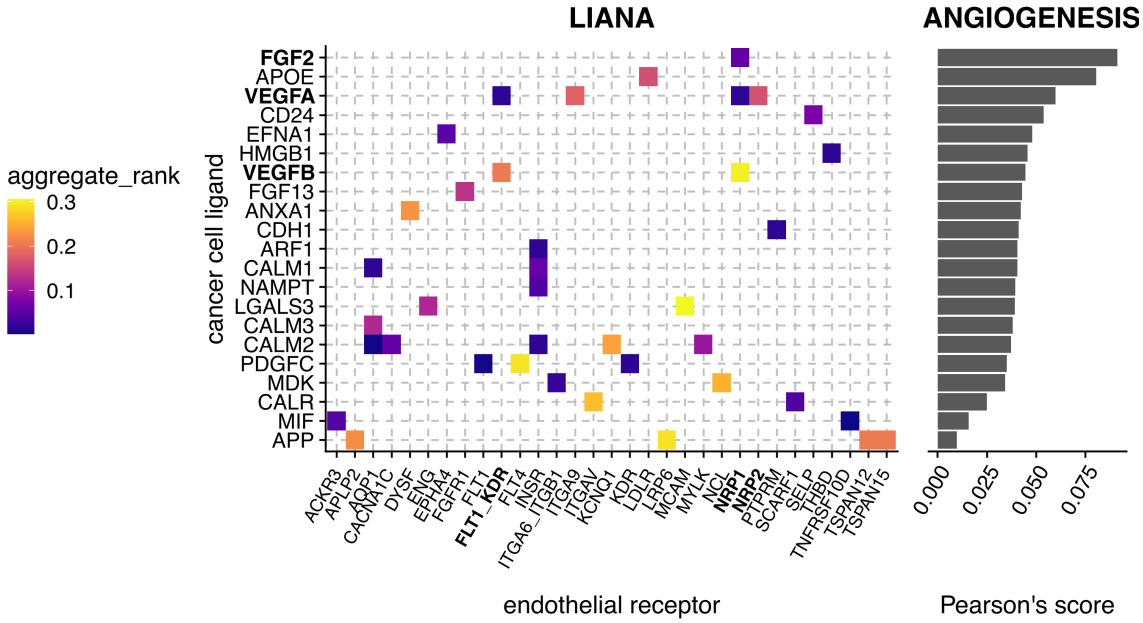


Figure 3.31: Endothelial Ligand-Receptor Analysis. Ligand-receptor analysis of angiogenesis-related interactions between IBC cells and endothelial cells.The analysis was performed using Liana and Nichenet with the Hallmark: Angiogenesis gene set from MSigDB.

Interestingly, I found that IBC cells express a variety of ligands that have the potential to activate angiogenesis. Specifically, the analysis identified several ligands, including VEGFA, VEGFB, FGF2, and other angiogenic factors like PDGFC and EFNA1, which are expressed by IBC cells. These ligands interact with corresponding endothelial receptors, such as FLT1, FLT4, KDR, and ITGAV, to promote angiogenic signalling pathways. Additionally, ApoE was identified as a ligand with potential angiogenic activity, highlighting the multifaceted roles of these molecules in the tumour microenvironment. These findings suggest that IBC cells may actively contribute to the pro-angiogenic environment by expressing a range of angiogenic ligands, which could facilitate tumour progression through enhanced blood vessel formation.

Mitogenic Pathways are Activated in IBC Endothelial Cells

I next set out to answer the question of what happens in response to the angiogenic cues received by the IBC endothelium. To that end, I performed pathway activity analysis implemented in the decoupleR package [152]. The tool allows to analyse pathway enrichment in omics data based on prior knowledge databases of gene regulatory networks including the sign and weights of interactions. For analysis of pathways, it comes with an implementation of PROGENy [153]. PROGENy is an acronym for Pathway RespOnsive GENes, and it is essentially a collection of publicly available perturbation experiments and their effects on gene expression.

I ran the pathway analysis based on the combination of these two tools and visualised the results split by IBC and non-IBC (Figure 3.32).

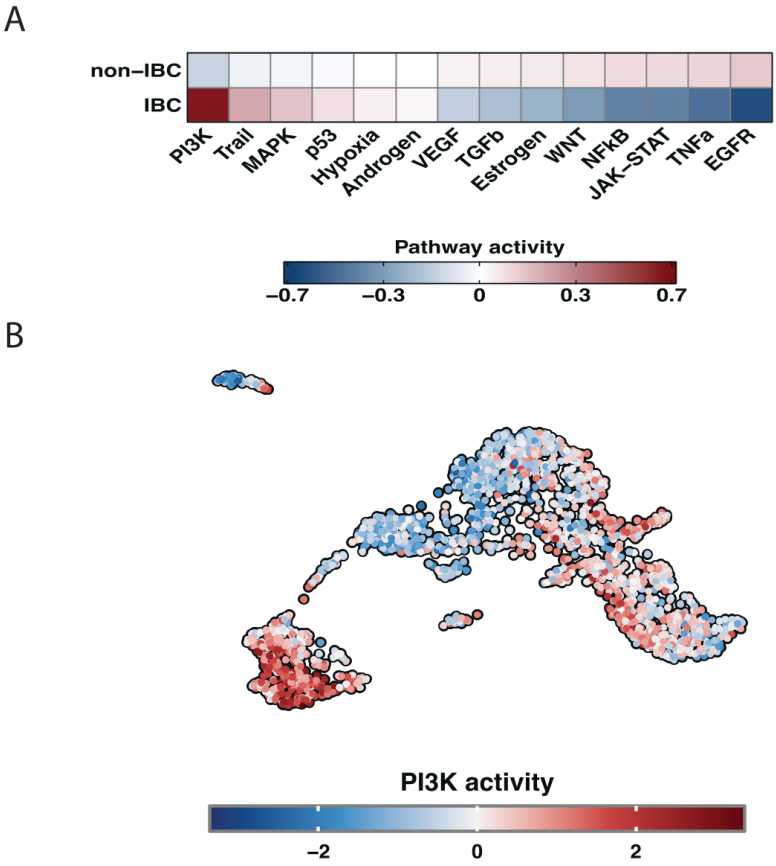


Figure 3.32: Endothelial Cells in IBC Upregulate Proliferation-Associated Pathways Pathway activity analysis in IBC versus non-IBC endothelial cells. The analysis was performed using the decoupleR package with PROGENy, highlighting the differential activity of various signalling pathways, possibly in response to angiogenic cues.

The analysis revealed differential pathway activity between IBC and non-IBC endothelial cells in response to angiogenic cues. Notably, pathways such as PI3K, VEGF, and MAPK showed elevated activity in IBC compared to non-IBC, suggesting that these pathways are more actively engaged in the angiogenic response of IBC endothelial cells. In contrast, pathways like p53 and TGFβ exhibited lower activity in IBC, indicating potential differences in regulatory mechanisms between

the two conditions. These findings suggest that IBC endothelial cells may indeed react to the previously identified angiogenic factors released by cancer cells (Figure 3.31), and preferentially activate specific signalling pathways to proliferate and enhance their angiogenic potential.

3.5 Spatial Transcriptomics Analysis of IBC

After in-depth analysis of the IBC and non-IBC microenvironment on single-cell level, I wanted to validate and supplement my findings using a spatial transcriptomics approach. Ewgenija Gutjahr thus assembled a small collective of IBC (n = 4) and non-IBC (n = 2) FFPE tissue samples from the pathology archive and created a tissue microarray. She sectioned the FFPE tissue block and mounted a section on a specific slide for Xenium spatial transcriptomics, obtained from 10X Genomics. An adjacent section was cut and mounted on a regular microscopy glass slide and H&E stained (Figure 3.33 A).
The DKFZ Single-Cell Open Lab, under the supervision of Dr. Jan-Philipp Mallm, then ran spatial transcriptomics using the 10X Xenium platform. To that end, Dr. Marc Zapatka provided the required gene panel, as part of the collaborative Saturn3 consortium. It was a breast cancer specific panel provided by the manufacturer (10X Genomics), with an additional 100 custom genes (Supplements 6.1).
The resulting data was then processed on the fly by the included software, performing cell segmentation based on radial nucleus expansion and clustering (Figure 3.33 B, C).

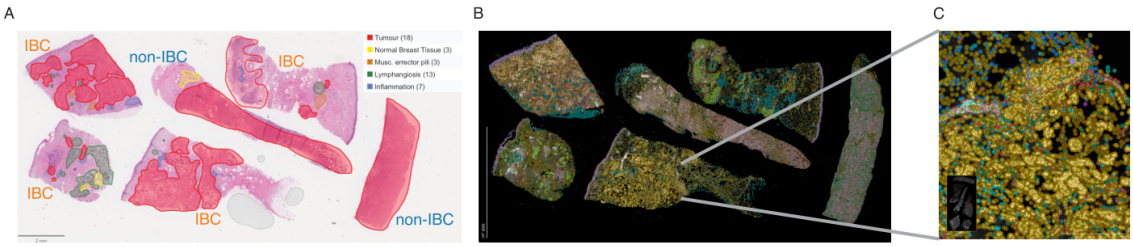


Figure 3.33: Xenium Slide Assembly and Data Recording
(A) A tissue micro array of 4 IBC and 2 non-IBC samples was created. Here, an adjacent section was H&E stained, imaged, and annotated by Ewgenija Gutjahr.
(B) Xenium data with cell segmentation and clustering performed by onboard software. Segmented cells are coloured by cluster.
(C) Detail view of the segmentation results in B.

3.5.1 Quality Control

I first assured sufficient data quality, by plotting a number of cellular and transcriptomic metrics (Figure 3.34). The distribution of total transcripts per cell (Figure 3.34 A) indicated a left-skewed distribution of transcript counts. The unique transcripts per cell distribution (Figure 3.34 A) follwed a almost normal distribution, with most cells expressing between 50 and 100 unique transcripts, pointing to consistent gene expression across the sample, which is the expected range.

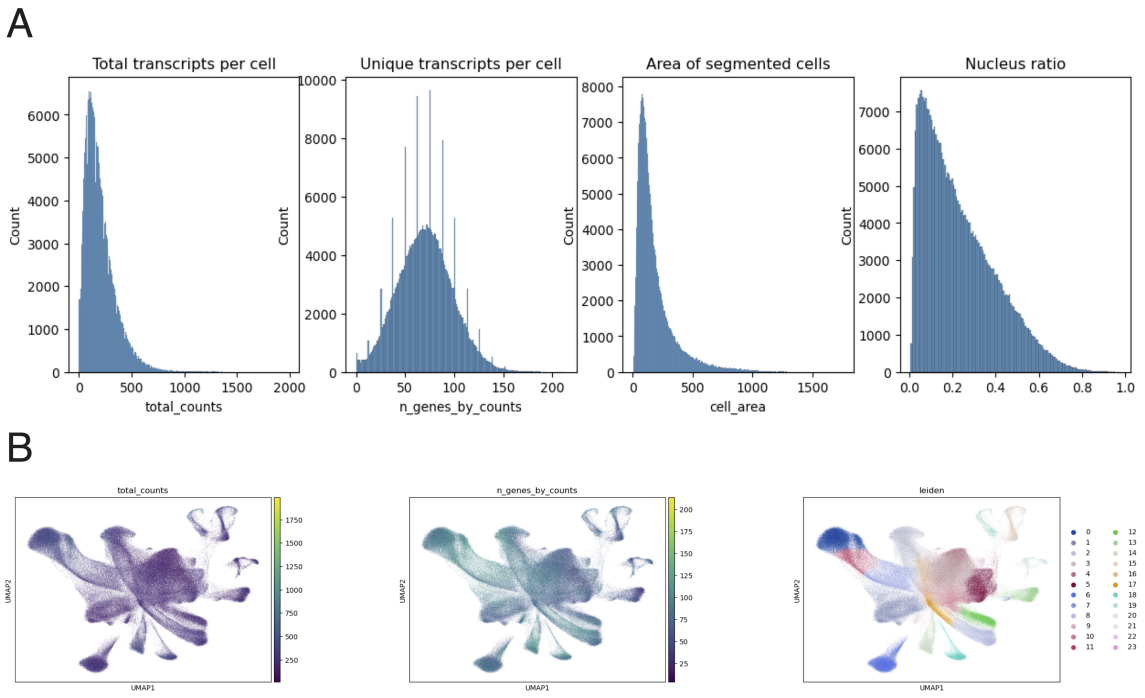


Figure 3.34: Xenium Data Quality Control
(A) Histograms showing (left to right) total transcripts per cell, unique transcripts per cell, segmented cell area, and nucleus-to-cell-area ratio.
(B) UMAP projections displaying total transcript counts, unique transcript counts, and Leiden clustering.

Cell size, represented by the segmented cell area (Figure 3.34 A), followed a left-skewed distribution, with most cells clustered around smaller sizes. The left-skew is to be expected with a nuclear expansion algorithm, when a minority of cells are located in sparse areas and thus are assigned a larger size. The nucleus-to-cell-area ratio (Figure 3.34 A) exhibited a skewed distribution as well, with the majority of cells having a low ratio, suggesting a successful radial expansion around the detected nuclei. In addition, this hints that there is a gain in information to be expected by using radial expansion, as compared to segmenting the nuclei only.

The UMAP projections (Figure 3.34, B) provided a spatial representation of transcriptomic diversity. Most importantly, based on the cell segmentation used, the algorithm creates clearly separated clusters. This is a very important quality attribute of a successful cell segmentation. The Leiden clustering (Figure 3.34 B, right) further demonstrated well-defined clusters, representing distinct cell populations, and underscoring the underlying transcriptional heterogeneity within the sample.

These results suggest robust data quality and a diverse cellular landscape, providing a solid foundation for downstream analyses.

3.5.2 Cell Type Assignment

I next assessed the clustering of cells. To do so, I plotted the cells in the previously calculated UMAP embedding, colouring them by sample, breast cancer type (bc_type, that is, IBC or non-IBC), and by cell type as assigned by two different methods (Figure 3.35).

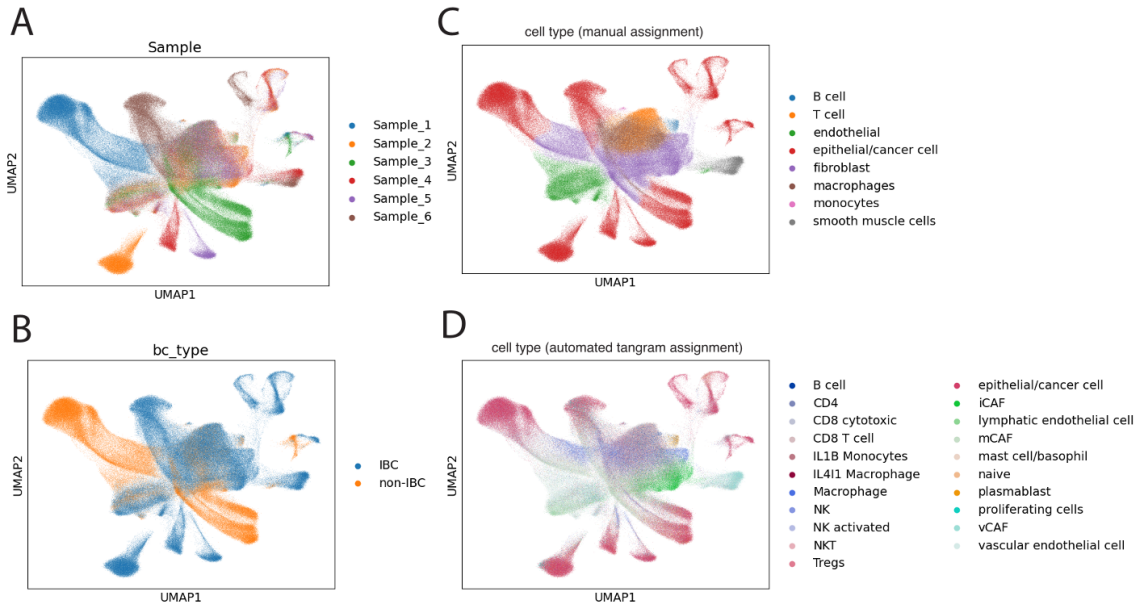


Figure 3.35: Clustering and Cell Type Assignment
(A) UMAP coloured by sample.
(B) UMAP coloured by breast cancer type (IBC / non-IBC).
(C) UMAP coloured by manually assigned cell type. The cell types were assigned based on the expression of cell-type specific genes, as annotated by 10X in the description of the gene panel.
(D) UMAP coloured by automatically assigned cell type labels. Tangram was used to deconvolve the cell type composition based on the single-cell data.

While there was a major cluster at the centre of the UMAP, which was made up of cells of all samples, around it there were also clusters exclusively comprised of cells from a single sample (Figure 3.35 A). This is also reflected by the distribution of IBC and non-IBC cells (Figure 3.35 B).

I used two different methods of cell type assignment. First, I manually assigned the cell types to the Leiden clusters. Each gene of the Xenium base panel from 10X came with a label indicating which cell type it mostly is expressed by. I thus created a heatmap of average expression of each gene per Leiden cluster (not

shown, as it is too large). Based on that I manually assigned a cell type to each cluster (Figure 3.35 C).

In addition, I used a tool called Tangram, which has been developed for cell type deconvolution of spatial transcriptomic data. It takes a single-cell dataset with cell type labels and a spatial dataset, both in the AnnData format as input. It then uses the expression of overlapping genes to assign cell type labels to the spatial dataset. (Figure 3.35 D).

To compare the results of both methods, I plotted some diagnostic plots, comparing the cell types assigned with each method and plotting the expression of typical marker genes on a per-cluster basis (Figure 3.36).

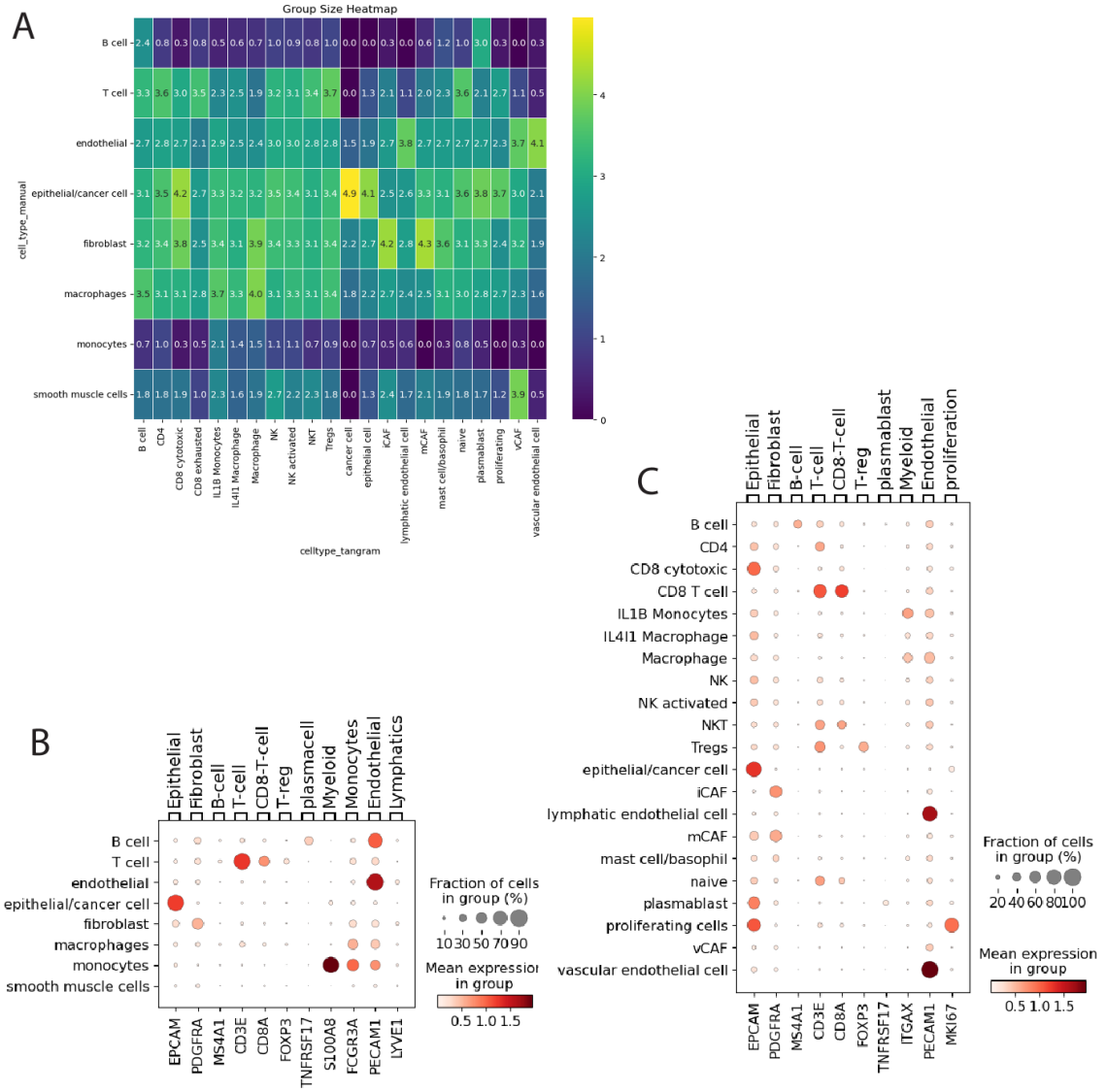


Figure 3.36: Comparison of Manual and Automated Cell Cype Assignment
(A) Heatmap showing the number of cells labelled with each method. It allows to assess if comparable cell types were given to the same cells using either method. Numbers and scale bar are log10(x+1), where x is the absolute number of cells assigned the label in x and y.
(B) Dotplot of typical lineage marker genes and the manually assigned cell types.
(C) Dotplot of typical lineage marker genes and the cell types projected from the single-cell data using tangram.

The types of cells assigned with the two methods are generally in good agreement (Figure 3.36 A). For example, most epithelial and cancer cells, as labelled by tangram (x axis), were manually labelled as epithelial/cancer cells (y axis). In addition, the cell clusters' average expression of the typical marker genes makes sense as well (Figure 3.36 B, C).

Expectedly, the clusters that are made up of cells from single samples are labelled as epithelial/cancer cells. As cancer cells typically display a higher degree of inter-patient transcriptomic variability, they often form distinct clusters.

3.5.3 Neighbourhood Enrichment

I next assessed cellular communities and compared them between IBC and non-IBC tissue samples. To that end, I used the Squidpy functions sq.gr.spatial_neighbors() and sq.gr.nhood_enrichment(). I used the the labels derived from the tangram projection of single-cell data, to obtain a more granular view of the cellular neighbourhoods. I did so for each sample individually. Then, for each element of a cell type x cell type matrix, I calculated the mean of IBC and non-IBC, and plotted that difference in means in a heatmap (Figure 3.37). In addition, I calculated a significance value using a t-test, for the difference between IBC and non-IBC for each field of the matrix.

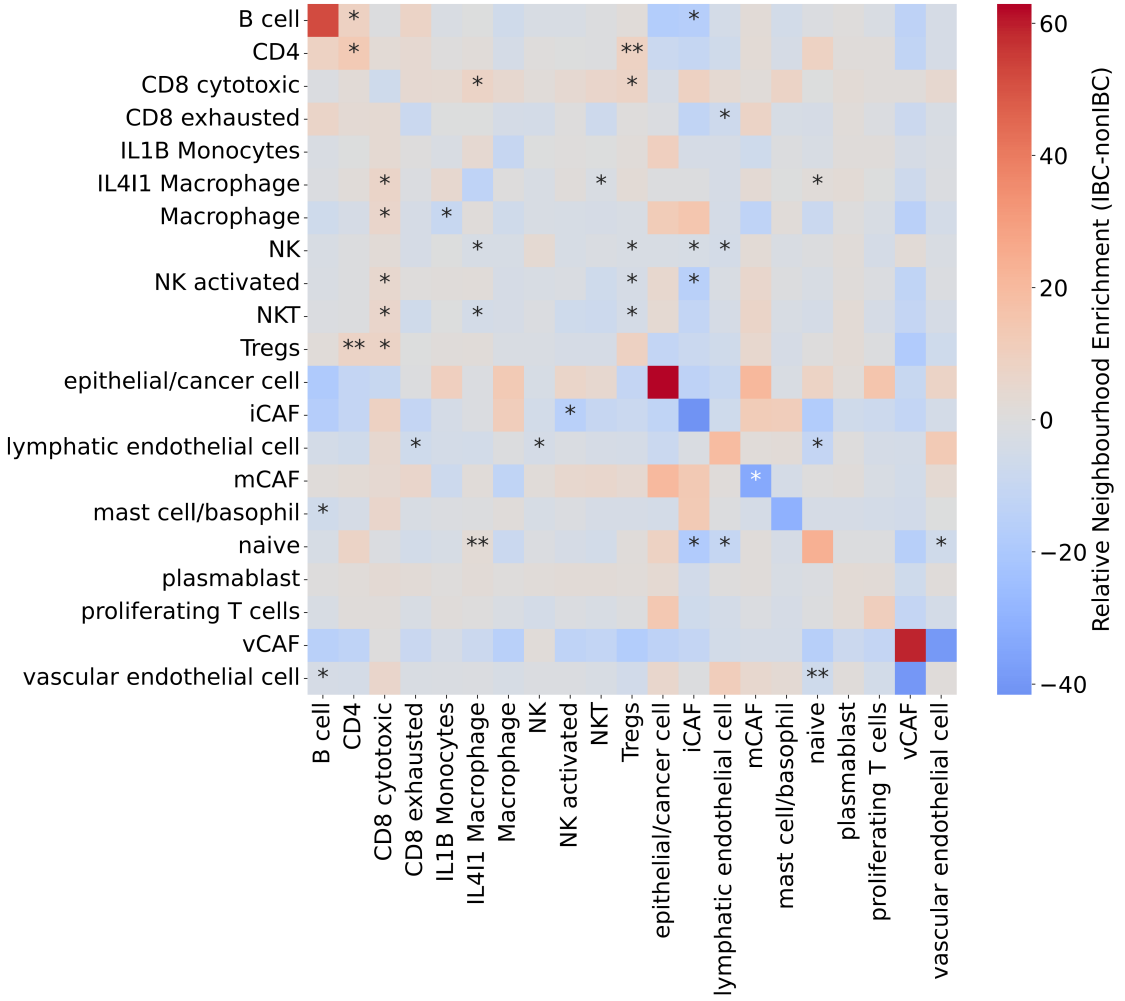


Figure 3.37: Relative neighbourhood enrichment in IBC compared to non-IBC Heatmap of the relative neighbourhood enrichment of cell types between IBC and non-IBC. $*p < 0.05$,$**$ $p < 0.01$.

Interestingly, there are significantly more regulatory T cells in the vicinity of CD8 cytotoxic and CD4 T cells in IBC, corroborating the notation that T cell immunity is actively suppressed by Tregs in IBC.

Additionally, macrophages, especially IL4/IL1 macrophages, appear to have increased interactions with NK cells and CD8 cytotoxic T cells in IBC, potentially highlighting a higher immune infiltration and diversity resembling inflammation in

these regions. This continuous inflammatory state may in the long run lead to the previously discussed immune exhaustion often noted in IBC.

3.5.4 Spatial Proximity of T Cells and Cancer Cells

The neighbourhoods of IBC cancer and epithelial cells were de-enriched in T cells, although these findings were not significant (Figure 3.37). But having previously described a suppressive immune microenvironment based on the single cell data, I wanted to investigate further if that was accompanied by a spatial difference as well. I thus calculated the mean shortest distance of epithelial/cancer cells to CD4 and cytotoxic CD8 cells (Figure 3.38)

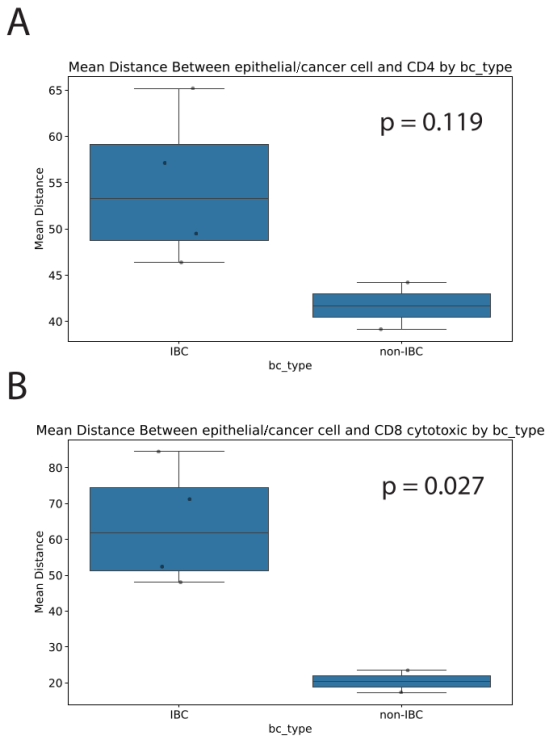


Figure 3.38: Spatial distribution of epithelial/cancer cells relative to T cells in IBC and non-IBC tissues.
(A) Mean distance between epithelial/cancer cells and CD4 T cells in IBC and non-IBC samples (p = 0.119).
(B) Mean distance between epithelial/cancer cells and CD8 cytotoxic T cells in IBC and non-IBC samples (p = 0.027).

This analysis highlights that there is indeed a greater distance between epithelial cells and their closest CD8 cytotoxic cancer cell (p = 0.027). There is a non-significant tendency for a greater distance to the nearest CD4 T cell as well (p = 0.119).

The increased distance to cytotoxic T cells is in support of the previous finding that IBC cells are under decreased immune surveillance and may be spatially located further away from T cells.

Chapter 4

Discussion

It has been widely accepted that the tumour microenvironment is the key feature defining inflammatory breast cancer [110]. The work presented in this thesis is the first in-depth, single-cell and spatial study of IBC. I created a sizeable single-cell dataset of IBC and non-IBC samples, using it to analyse the differences between the two across major cell lineages. Thus, I was able to draw a detailed map of IBC's key features - for the first time supporting the hallmarks of its biology by single-cell molecular data.

4.1 Single-Cell and Spatial Transcriptomics Enable a New and More Detailed View on IBC

The complexity of inflammatory breast cancer (IBC) has driven research efforts towards dissecting both tumour-intrinsic factors and the tumour microenvironment (TME), which collectively contribute to its aggressive behaviour. Previous methodologies, such as microdissection, sought to minimise the confounding influence of the TME on gene expression profiles [106]. However, these techniques have provided limited insights into the distinct molecular signatures of IBC. Instead, they have highlighted alterations in several well-known cancer-related pathways, including stemness markers like ALDH1 and CD44, which have been strongly associated with tumour-initiating properties and poor prognosis in breast cancer [108]. Other dysregulated pathways, including epithelial-to-mesenchymal transition (EMT), VEGF, RAS, PI3K, JAK-STAT, and TGFβ, further underscore the highly metastatic nature of IBC [109, 110, 111].

While genomic studies have not identified IBC-specific mutations, they have revealed similarities in the mutation landscape between IBC and non-IBC breast cancers [112, 113]. This finding suggests that the unique clinical presentation of IBC may not be driven by a distinct genetic profile, but rather by the context in which these mutations operate, particularly within the inflammatory and immune-

suppressive TME. The role of the TME, therefore, emerges as a critical factor in shaping the aggressive phenotype of IBC.

The tumour microenvironment, composed of stromal cells, immune cells, and the extracellular matrix, is known to significantly influence cancer progression. In IBC, myeloid cells have been implicated in promoting both lymphangiogenesis and metastasis through the secretion of cytokines and growth factors that modulate the invasive potential of IBC cells [110]. Myeloid cells, depending on their subtype, can promote tumour progression by fostering processes including immune suppression, as well as invasiveness and EMT [154]. The IL-1B expressing myeloid cells in IBC may contribute to its early and rapid lymphatic spread, which is characteristic of this disease.

Furthermore, the immune landscape within the TME has also garnered increasing attention. My data suggesting a suppression of T cell activity in IBC, points towards the potential for immune evasion mechanisms that could hinder an effective anti-tumour immune response. This finding may give a rationale for the therapeutic use of immune checkpoint inhibitors, such as those targeting the PD-1/PD-L1 axis, in the treatment of IBC [115]. However, clinical data on the efficacy of these inhibitors in IBC remains limited, highlighting the need for further investigation into how immune checkpoint blockade might be integrated into the treatment paradigm for this disease.

In addition to immune suppression, the cytokine-rich environment of the TME appears to facilitate communication between IBC cells and surrounding stromal and immune cells. This interaction was previously proposed to foster processes such as lymphangiogenesis, which promotes the dissemination of cancer cells via the lymphatic system, and immune evasion, allowing the tumour to escape immune detection [116, 117, 118]. Together, these processes underscore the pivotal role of the TME in driving both the local invasiveness and metastatic potential of IBC.

Given these insights, it becomes evident that future therapeutic strategies should not solely focus on targeting IBC cells themselves, but also consider the TME as

a crucial component of disease progression.
The advent of single-cell and spatial methods has revolutionised our ability to study the molecular basis of how the microenvironment is shaped in health and disease [155, 156]. Especially the ability to assess stromal and immune cells in the microenvironment of tumours allows researchers to gain a better understanding of how these cell types contribute to disease progression as well as therapy response [157], making them an ideal tool to study the microenvironment in IBC.
Single-cell transcriptomics enables the characterisation of cell-states and activities, as well as the analysis of shifts in various pathologic settings, making these methods suitable to compare cell states in the microenvironment of IBC to non-IBC.
In addition, spatial transcriptomics adds an additional dimension to the data, allowing more precise statements on molecular interactions and local communication circuits and the description of specific cell communities and neighbourhoods [158].

4.2 A Single-Cell Map of IBC

In the course of the research towards this thesis, I created and curated single-cell and spatial transcriptomics data from IBC and non-IBC patient samples. Here, I would like to discuss some of the strengths and weaknesses of the resulting data set.
The resulting single-cell transcriptomics dataset comprised 55,714 cells, made up in equal parts by IBC and non-IBC. Owing to developments in technology, the scale of published single-cell datasets has rapidly increased and reached millions of cells. Comparing the size of our dataset to the trajectory of development over the past years, it is clear that it is of the size of a median dataset of approximately 3-4 years ago (Figure 4.1) [159].

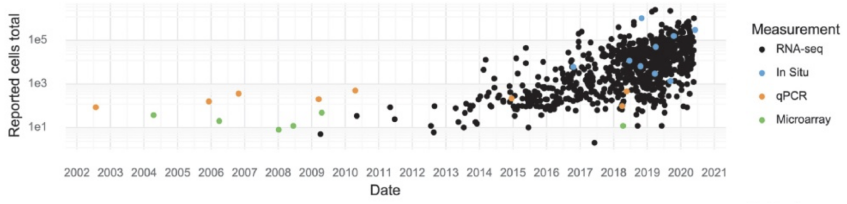


Figure 4.1: Single-Cell Dataset Size Over Time A scatter plot of dataset size (total number of cells) over time. Created and published under a creative commons license by Valentine Svensson *et al* [159].

In addition, the number of IBC patients included is limited at only 9 individuals. Also, the representation of breast cancer subtypes is not balanced (Table 3.3). However, considering that inflammatory breast cancer is a rare entity and the period of sample collection was limited, this number is certainly not to be considered insignificant. The fact that a whole-cell protocol was used for the single-cell data generation meant that only freshly digested cells could be used. This in turn meant that we could not fall back on large collections of archived biomaterial, like for example FFPE tissue blocks.

In the future, using methods like single-nucleus sequencing (nuc-seq)or spatial transcriptomics will enable researchers to tap into these vast resources and generate data on a much larger and more representative scale.

An additional benefit will be the fact that sampling artifacts will be addressable. Currently, the spatial information is lost at the time point of sampling. Because of the small size of biopsy samples and the low yield of digestion methods, they have to be used in their entirety. Thus, it is not possible to reconstruct the exact location or (tumour) tissue content of the biopsies.

In particular, spatial transcriptomics has the huge advantage over single-cell sequencing, that it comes with spatial context, and usually a reference H&E staining of either the exact same, or an adjacent tissue section. Thus, it is possible to contextualise the cellular communities identified as part of the analysis, and thus extract much more reliable findings.

4.3 The Single-Cell and Spatial Data Help Understand the IBC Phenotype and Suggests New Therapeutic Avenues

In this thesis, I described in detail the molecular cell states of cancer-associated fibroblasts, tumour associated myeloid cells, tumour infiltrating lymphocytes and endothelial cells, and how they contribute to the aggressive behaviour, rapid metastatic progression and therapy resistance of IBC (Figure 4.2).

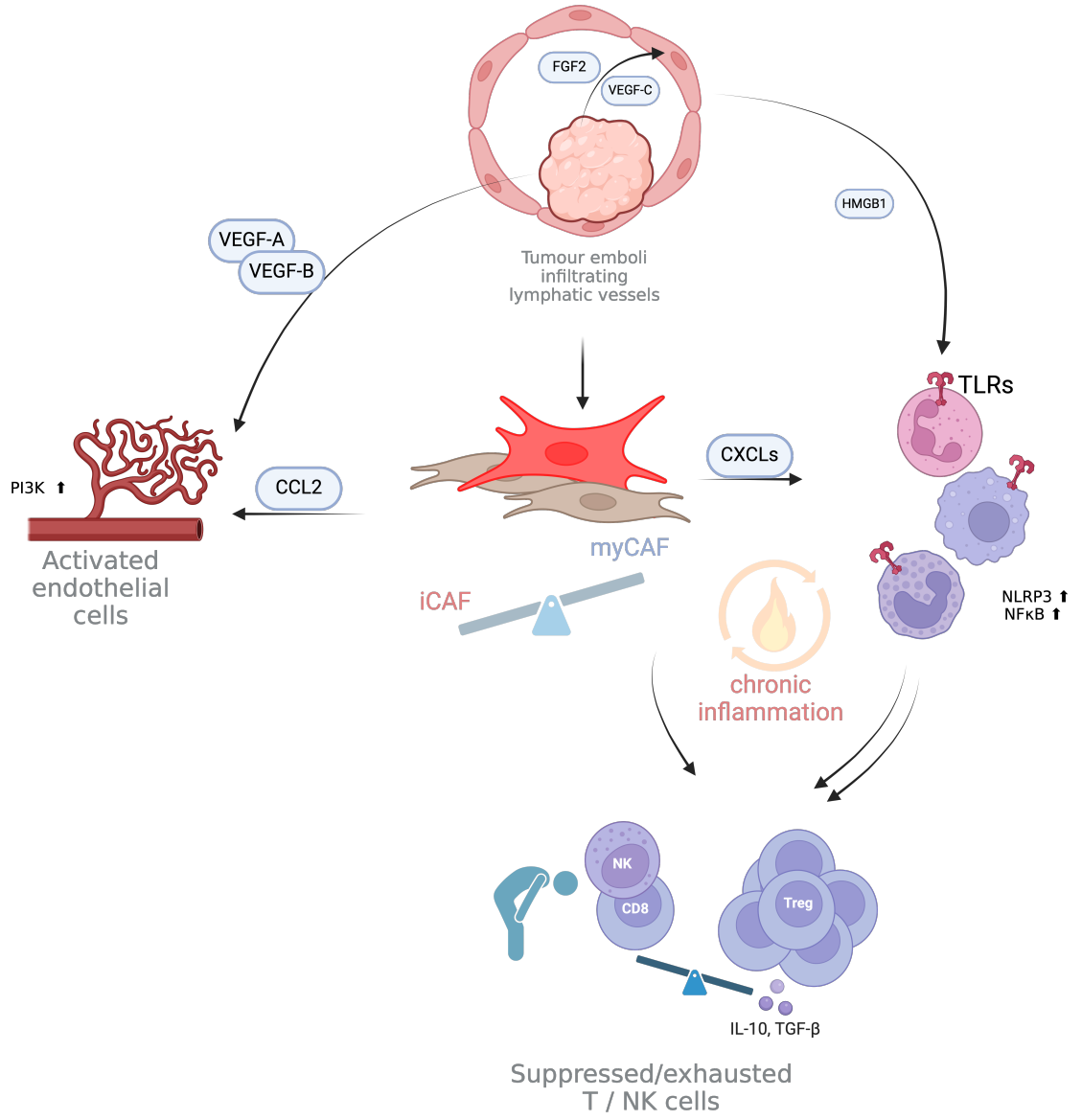


Figure 4.2: Suggested Model of IBC Biology. A graphical overview of findings, which suggest a model of a self-sustaining inflammatory response, driven by fibroblast signalling, an inflammatory state of myeloid cells, and suppression and exhaustion of T and NK cells. Tumour growth and progression is supported and enabled by angiogenesis, trough the activation of endothelial cells. I created the figure in Biorender using a paid license under my supervisor Martin Sprick's name.

I propose that cancer associated fibroblasts (CAFs) contribute to the shaping of an inflammatory tumour microenvironment in IBC by acting as signalling hubs. Inflammatory iCAFs, myofibroblastic mCAFs, and vascular vCAFs, all cooperate to signal into the myeloid compartment, possibly inducing and sustaining an inflammatory response. iCAFs typically elicit tumour-promoting effects by secreting pro-inflammatory cytokines and growth factors [160]. We demonstrated that in IBC, fibroblasts signalling to the microenvironment is strongly increased, with activated signalling pathways, such as CXCLs, IGFs, and FGFs.
Targeting tumour-promoting pathways of CAFs, specifically iCAFs could thus be a viable therapeutic option [161]. Inhibition of IL-1 signalling, with the aim of suppressing iCAFs, has previously shown some promising preclinical results in the context of pancreatic cancer [162]. Additionally, inhibition of CXCL12 was shown to increase T cell infiltration and anti-tumour immunity in pancreatic cancer [163].

A major cell population that my data predict to be affected by altered CAF signalling are tumour associated macrophages and monocytes. IL-1B monocytes, enriched in IBC, have previously been reported to sustain inflammatory response in health and disease [145]. Mechanistically, an interplay of damage associated molecular pattern signalling from cancer cells (via HMGB1/HSP90B1-TLRs), as well as IL-1β from the monocytes themselves, could prime the activation of the NF-κB signalling. This in turn, via the NLRP3 inflammasome, could sustain IL-1β secretion and an inflammatory response in the microenvironment.
Hadadi *et al.* previously showed that IL-1β levels in inflammatory conditions was modulated by the Hsp27 pathway, and could thus be a viable target to modulate inflammatory response [145].

A defining feature of IBC is its angiogenic phenotype that facilitates rapid cancer progression. I showed that both lymphatic and vascular endothelial cells are ac-

tivated in IBC and display upregulation of mitogenic pathways. My data suggest that this likely happens by stimulation through myeloid cell-derived NF-κB and cancer cell derived FGF2, VEGFA and VEGFB, stimulating the downstream PI3K pathway.

The anti-VEGF antibody bevacizumab was previously FDA-approved in combination with paclitaxel for metastatic HER2-negative breast cancer, but was later withdrawn due to an insufficient positive to adverse effects ratio [164]. However, the European EMA still sees an indication for first-line therapy in metastatic breast cancer, in combination with paclitaxel or capecitabine [165]. Possibly, using bevacizumab for IBC, where angiogenesis is a key feature, could be a strategy worth investigating, especially as VEGF signalling may play a role causing it. It could help tip the balance towards more beneficial outcomes outweighing the adverse effects.

In immune-mediated inflammatory diseases with chronic inflammation, T cells are continuously stimulated by the inflamed microenvironment and eventually lose their cytotoxic function and enter an exhausted state [166]. I show that inflammatory breast cancer is a prime example of this, reflected by very high level of T cell infiltration, but mostly in either a regulatory or exhausted state. An overall increased expression of TGF-β in the microenvironment is likely to be both cause and effect of this in a self-sustaining manner. A possible implication is that reactivation of these T cells by immunotherapy might be a promising angle to challenge this hard-to-treat disease. In addition, TGF-β targeted treatment could be worth exploring, to alleviate some of its immune-suppressive effect [167].

4.4 Conclusion

Taken together, my results draw a clear picture of the key role of the inflammatory microenvironment of IBC, suggesting it is causal for the disease's defining features. Importantly, there are no therapies specifically targeting molecular path-

ways that are used in practice to treat IBC so far. My findings suggest a number of possible avenues for exploration for specific therapeutic targeting, informed by these detailed molecular data.

In conclusion, while significant progress has been made in elucidating the pathways involved in IBC pathogenesis, the TME plays an indispensable role in shaping its unique and aggressive clinical behaviour. The identification of inflammatory and immunosuppressive pathways within the TME, along with the presence of TAMs and T cell suppression, highlights the importance of developing therapies that target both tumour cells and their microenvironment. As research continues, future studies should focus on understanding the interplay between tumour genetics and the TME, as well as the potential of combination therapies that address both intrinsic and extrinsic drivers of IBC. To go beyond hypothetical target pathways, it will be necessary to develop and experimentally validate *in vitro* and *in vivo* model systems, that faithfully reflect the biology and major hallmarks of IBC.

This work should be regarded as just a further step in the direction of elucidating IBC. The road up to the successful treatment of IBC is still a long one, but as new methods and model systems become available, and our knowledge continuous to grow, we can continue building towards the long-term goal of bringing therapies to patients suffering from inflammatory breast cancer.

Chapter 5

Materials and Methods

5.1 Sample collection

Biopsy specimen from clinically diagnosed inflammatory breast cancer patients, as well as control samples of non-inflammatory breast cancer, were obtained via vacuum-assisted biopsy from consenting patients that were treated at the National Centre for Tumour Diseases (NCT) in Heidelberg, Germany. Samples were collected as part of the COGNITION trial by clinical collaborators under the supervision of Prof. Dr. Andreas Schneeweiss. All human participants gave explicit written consent, and samples were collected following ethical guidelines that were reviewed and approved by the institutional review board (Heidelberg University). After collection, samples were immediately transferred into transport medium consisting of CO2 independent medium (Life Technologies), 1% bovine serum albumine (ThermoFisher), Glutamax (ThermoFisher) and Penicillin/Streptomycin (ThermoFisher), and then transported at ambient temperature to our lab at Deutsches Krebsforschungszentrum (DKFZ), also in Heidelberg, Germany.

5.2 Tissue digestion and biobanking

Tumour biopsies were dissociated using Human Tumor Dissociation Kit (Miltenyi). They were transferred to a C Tube (Miltenyi) containing 2.5 ml of a mix of transport medium (for ingredients see above) and a cocktail of digestive enyzmes. They were then placed on a gentleMACS Octo (Miltenyi) and dissociated at 37℃ with a programme of intermittent mechanical shearing ("37℃_h_TDK_3 ") for 1 hour. After digestion the resulting cell suspension was passed through a strainer with 100 μm pore size (Greiner Bio-One), and centrifuged briefly. Then they were resuspended in CryoStor CS10 freezing medium (Stem Cell Technologies) and frozen using a Mr. Frosty Freezing Container (Thermo Fisher) and subsequently biobanked in liquid nitrogen vapour.

5.3 Single-cell library creation and sequencing

Single-cell RNA sequencing (scRNA-seq) was performed using the 10x Genomics Chromium Single Cell 3' v3.1 platform. Samples were retrieved from biobanking and batch-processed. Cells were retrieved from biobanking, then their cell viability was assessed using a hemocytometer. If viability was below 80%, dead and apoptotic cells were depleted using the Dead Cell Removal Kit (Miltenyi) following the manufacturer's instructions. For samples to be multiplexed during the encapsulation step, each individual sample was labelled with a unique cell barcoding antibody (TotalSeq™-B). Cell hashing was performed on the 10x Genomics Chromium Controller. This strategy facilitated the identification and deconvolution of multiplexed samples during downstream analysis. The subsequent steps, including cell encapsulation, reverse transcription, and cDNA amplification, were performed following the 10x Genomics Single Cell 3' v3.1 Library Construction protocol, with the addition of cell hashing reagents. The cDNA libraries and hashing libraries were then prepared separately for each sample or pool of hashed samples, and multiplexed prior to sequencing. The single-cell suspension was loaded onto the 10x Genomics Chromium Controller, aiming for a recovery of 10'000 (non-multiplexed samples) or 30,000 (hashed samples) cells per well. The libraries were sequenced on a Illumina NovaSeq 6000 platform using paired-end sequencing and dual indexing for 100 cycles each.

5.4 scRNAseq demultiplexing

I developed an improved method for demultiplexing pooled single-cell experiments where the individual samples were barcoded with hashing antibodies. By combining a genomic variant-based approach with the hashing approach, I circumnavigated the shortcomings of either. This achieved a higher cell yield and purity, because the method does not have to rely on the successful binding of the hash-

ing antibody to each individual cell. Additionally, I was able to use the levels of hashing antibody to assign genomic cluster to patient sample, which would otherwise require information on the patients' genomic variants and manual labour to match the patients' SNP profiles with the cell clusters.
After undergoing quality control, the raw sequencing reads were mapped to the human reference genome GRCh38, and the quantification of reads per gene per cell was conducted using CellRanger count (10x Genomics, v7.1.0) with the GRCh38 2020 reference genome provided with the CellRanger software (10X). Subsequently, the cells were clustered by genotype (i.e. individual) using Souporcell v2.5 [168], which allows for demultiplexing cells without relying solely on the barcoding tag on each cell, the recovery levels of which can be variable. Briefly, the sequencing reads are re-mapped to a reference genome and candidate variants are called. Then for each cell the alleles are counted, allowing subsequent clustering by genotype and removal of multiplets (i.e. with mixed genotypes). I then inferred the per-cell levels of each barcode using Seurat's HTODemux function. Using these I then matched each cluster with the correct patient, allowing for direct and automated assignment of the correct patient to each cluster.

5.5 scRNAseq data preprocessing, harmonisation and cell type assignment

Raw counts matrices underwent processing through SoupX, calculating an ambient RNA profile from the empty droplets, and correcting the counts for it. Then, for each sample, I removed likely genetically identical (from the same patient) but heterotypic (from different cell types) doublets using scDblFinder v1.12.0 [132] with the default settings. This is done by generation of random artificial doublets, and comparing each measured cell to their profiles. Next, cell-level quality control was performed. I used the addPerCellQCMetrics function from the scuttle pack-

age and used the number of median absolute deviations (nmad) as an automated sample-wise approach to filter out cells that are outliers in terms of any of the following metrics: total number of UMIs per cell and percentage of mitochondrial reads. I then created a Seurat (v4.3.0.1) [169] object, containing the annotated and filtered data for each sample. To address technical variability always present in scRNAseq data, I performed normalisation and variance stabilisation using sctransform v2 from the Seurat package [170]. To harmonise the data and remove batch-related effects, I used the linear embedding model harmony (v1.1.0) [124]. The most variable genes identified by sctransform were used to construct principal components, which were then used as input for graph-based clustering using the FindNeighbours and FindClusters (resolution 0.1) functions of Seurat. The major cell lineages were annotated using established marker genes (*EPCAM* for epithelial cells and cancer cells, *PDGFRA* for fibroblasts, *PECAM1* for endothelial cells *CD3G* for T and NK cells, *MS4A1* for B cells, *IGKC* for plasmablasts, *ITGAX* for myeloid cells and *MS4A2* for mast cells/basophils). Due to their transcriptional similarity, T and NK cells form a single cluster, and where annotated together at this stage.

5.6 Assignment of malignant cells

For each sample, inferCNV was run using T/NK and myeloid cells as the reference group. The resulting CNV scores were integrated into the Seurat objects containing IBC and non-IBC single-cell datasets. CNV metadata were extracted, and a CNV signal vector was computed for each cell as the mean of the squared CNV proportions across all annotated chromosomal regions. Cells were ranked based on their CNV signal, and the top 5% with the highest CNV signal were designated as having a malignant CNA profile. A malignant CNA profile was defined by averaging the CNV scores of these top 5% cells across all relevant chromosomal features.

Subsequently, a correlation coefficient was calculated for each cell by correlating its CNV profile with the derived malignant CNA profile. Cells were classified as malignant if they satisfied the following criteria: (1) CNV signal > 0.02, (2) correlation coefficient with the malignant CNA profile > 0.4, and (3) identified as 'epithelial/cancer cell' within the cell type annotations. Cells that did not meet these thresholds were classified as non-malignant. The integration of CNV signal and correlation metrics into the Seurat objects facilitated robust discrimination between malignant and non-malignant epithelial cells within the single-cell datasets.

```
# Loop through each Seurat object in the list 'IBC_seurats_annot'
for (i in 1:length(IBC_seurats_annot)) {

  # Extract CNV-related metadata
  cnv_metadata <- IBC_seurats_annot[[i]]@meta.data %>% dplyr::select(contains("proportion_scaled_cnv_chr"))

  # Compute CNV signal vector as the mean of squared CNV proportions for each cell
  cna_signal_vec <- apply(cnv_metadata, MARGIN = 1, function(x) { mean(x^2) })

  # Define the threshold for the top 5% of cells with the highest CNA signal
  toppercentile <- quantile(cna_signal_vec, probs = 0.95)

  # Identify malignant CNA profile as the average CNV profile of the top 5% cells
  malignant_cna_profile <- apply(cnv_metadata[which(cna_signal_vec >= toppercentile), ],
                                 MARGIN = 2,
                                 FUN = function(x) { mean(x) })
```

```r

  # Calculate the correlation of each cell's CNV profile to the malignant CNA profile
  cna_corr_vec <- apply(cnv_metadata, MARGIN = 1, function(x) { cor(x, malignant_cna_profile) })

  # Add CNA signal and correlation vectors to the Seurat object
  IBC_seurats_annot[[i]]$CNA_signal <- cna_signal_vec
  IBC_seurats_annot[[i]]$CNA_corr <- cna_corr_vec
}

# Merge all Seurat objects into a single Seurat object
SKref_IBC_balanced_infercnv <- merge(IBC_seurats[[1]], IBC_seurats[2:length(IBC_seurats)])

# Summary of CNA signal and correlation in the merged object
summary(SKref_IBC_balanced_infercnv$CNA_signal)
summary(SKref_IBC_balanced_infercnv$CNA_corr)

# Annotate cells as malignant or non-malignant based on defined criteria
SKref_IBC_balanced_infercnv$malignant <- ifelse(
  SKref_IBC_balanced_infercnv$CNA_signal > 0.02 &
    SKref_IBC_balanced_infercnv$CNA_corr > 0.4 &
    SKref_IBC_balanced_infercnv$cell_type == 'epithelial/cancer cell',
  "malignant",
  "normal"
)
```

Listing 5.1: R code for annotating epithelial cells

5.7 Single-cell data manipulation and visualisation

We utilised the Seurat package (v4.3.0.1) [169] , alongside the SCpubR package (v1.1.2) [171], to effectively manipulate and visualise single-cell data. Seurat, a powerful tool widely employed in single-cell RNA sequencing (scRNA-seq) analysis, provided us with a comprehensive framework for quality control, normalisation, clustering, and dimensional reduction of scRNA-seq datasets. Leveraging Seurat's functionalities, I was able to preprocess and organise our single-cell data efficiently, ensuring accurate downstream analysis. Additionally, I incorporated the SCpubR package, which offers versatile functions for visualising single-cell data, enabling us to generate informative plots and graphs to explore cellular heterogeneity, cluster distributions, and gene expression patterns within our dataset.

5.8 Dataset integration

To integrate our data and published single-cell RNA sequencing datasets, I utilised the Harmony and Seurat packages. Initially, individual Seurat objects were created for each dataset. Each dataset was preprocessed independently using the SCTransform (SCT) method for normalisation and variance stabilisation with the SCTransform function, followed by identification of highly variable features. The datasets were then merged into a single Seurat object using the merge function. I applied Harmony integration to correct for batch effects. Principal component analysis (PCA) was performed on the merged dataset using the RunPCA function. Harmony was then run on the PCA embeddings using the RunHarmony function, specifying the batch variable to harmonise the data. The resulting Harmony embeddings were used for downstream analyses. Subsequently, the Harmony-corrected embeddings were used to construct a shared nearest neighbour (SNN) graph with the FindNeighbors function, and clustering was performed with the FindClusters function. Uniform Manifold Approximation and Projection (UMAP)

was used for visualisation with the RunUMAP function. This integrated dataset facilitated consistent clustering and visualisation across the combined datasets, enabling comprehensive downstream analyses.

5.9 Quantification of relative cell numbers

To analyse and visualise the proportion of subclusters within patient samples, I performed the following steps: Subcluster information was converted to a factor and assigned as cell identities using the seuratobject@meta.data and Idents functions. I created a new metadata column combining IBC/non-IBC and patient information. To retain only samples with more than a specified number of cells, I counted the number of cells per patient and filtered samples accordingly, storing the results in a reduced Seurat object. I then calculated the proportions of cells in each subcluster per patient and annotated these proportions with IBC/non-IBC. Using the ggplot2 (v3.4.3) and ggpubr packages, I visualised the cell type proportions by generating a boxplot for each subcluster, with jittered points representing individual patient samples. The y-axis displayed the cell type proportions as percentages. I also performed t-tests to compare the proportions between IBC and non-IBC groups, displaying the p-values on the plot.

5.10 Ligand-receptor analysis using scRNA-seq data

CellChat To infer the potential intercellular communication network between clusters, I utilised the CellChat R package (v1.1.3). This tool quantitatively measures networks based on the law of mass action, using the average expression values of ligands from each cell group and receptors from each other cell group. I set the ligand-receptor database to CellChatDB.human. The normalised expres-

sion matrix was obtained from the Seurat object of interest, separately for IBC and non-IBC. Then a list of CellChat objects was created using the createCellChat() function. I preprocessed the CellChat objects, identifying overexpressed genes (identifyOverExpressedGenes()) and interactions (identifyOverExpressedInteraction()), inferring cell-cell communication probabilities (computeCommunProb()), filtering communications with fewer than 10 cells (filterCommunication(min.cells = 10)), and computing communication probabilities at the signalling pathway level (computeCommunProbPathway()). The aggregated cell-cell communication (aggregateNet()) network was calculated and saved. Finally, I merged the CellChat objects for IBC and non-IBC, generating visualizations to compare the number and strength of interactions between the cell populations. Differential network analysis and network centrality scores were computed to compare the signalling roles and identify significant changes between datasets. Visualisations were created to compare outgoing and incoming interaction strengths in 2D space and to highlight differential signalling changes associated with specific cell groups

Liana and NicheNet I used a combination of the R package 'liana' (v0.1.13) [140] and 'nichenetr' (v2.0.5) [172] to analyse how ligand-receptor pairs that are predicted to be upregulated can induce specific gene programmes in target cells. In this context, liana was used to rank possible ligand-receptor interactions, while nichenet aims to quantify the correlation to the activation of intracellular signalling and transcriptional targets that are part of the pathway of interest. To that end, I followed the procedure outline by the authors in their tutorial (https://saezlab.github.io/liana/articles/liana_nichenet.html).
Briefly, I created a Seurat object containing only IBC cells, and ran liana using the liana_wrap() function with the 'method' argument set to 'c('natmi', 'connectome', 'logfc', 'sca', 'cellphonedb')'. I then ran the liana_aggregate() function to calculate an overall score taking into account the various prior knowledge databases used in the analysis. For the nichenet part, I calculated a list of background genes, defined

as expressed in at least 5% of cells of the receiver cell population. I then used the ligands from the liana analysis, a list of gene names for the gene set of interest, as well as the prior knowledge ligand-target matrix included in the package to predict the downstream activity for each ligand using the 'predict_ligand_activities()' function. I then used ggplot2 (v3.4.3) to visualise the liana results as a tileplot and the nichent results as a bar plot.

5.11 Gene set enrichment analysis

We used enrichR (v3.2) to assess the enrichment of genes within genesets [173]. To that end, we calculated the marker genes of a cell population of interest using Seurat's FindMarkers() or FindAllMarkers() function, using the "MAST" test for differentially expressed genes. Enrichment of the markers within the genesets of interest was then calculated using the enrich() function of enrichr. It computes p-values for enrichment based on Fisher's exact test.

5.12 Pseudobulk gene expression comparison

To compare the pseudobulk expression of a specific gene per patient between IBC and non-IBC groups using Seurat's AverageExpression function, I followed these steps: I utilised the AverageExpression function to calculate the average expression levels of the specific gene across all cells for each patient within the IBC and non-IBC groups. The resulting data frame provided the average expression values for each patient, grouped by condition. I visualised these average expression values using boxplots to compare the pseudobulk expression between IBC and non-IBC groups. Additional statistical tests, such as t-tests, were conducted to determine the significance of the differences observed between the two groups. This approach enabled a clear comparison of gene expression levels on a pseudobulk basis, facilitating the identification of significant expression differences

between IBC and non-IBC patients.

5.13 Pseudobulk expression heatmap for T/NK cells

To calculate and visualise the cluster averages of gene expression using Seurat and ComplexHeatmap, I followed these steps: First, I used the AverageExpression function on the Seurat object containing T and NK cells, to calculate the average expression levels of genes across different clusters, specified by the metadata column specifying the specific T and NK subsets previously predicted based on a reference dataset. The results were returned as a matrix. The average expression matrix was then normalised between 0 and 1 using a custom function that scales each gene's expression values to this range. Additionally, I standardised the data by calculating z-scores, which involved subtracting the mean and dividing by the standard deviation for each gene.
For visualisation, I utilised the ComplexHeatmap package to generate a heatmap of the normalised average expression data. The heatmap displayed the normalised mean counts of selected genes, with rows representing genes and columns representing clusters. Various parameters were adjusted to enhance the visual representation: row names were displayed on the left, column names on the top, with column names rotated 90 degrees. I used the viridis colour palette for the heatmap, ensuring a visually appealing gradient. The heatmap was not clustered, maintaining the original order of genes and clusters. Additionally, rows were split based on a predefined gene category, providing clear categorisation within the heatmap. This method allowed for an intuitive visual comparison of normalised gene expression across different clusters, highlighting patterns and differences in expression profiles.

5.14 Pathway activity analysis

I used decoupleR [174, 152] to facilitate pathway activity analysis within single-cell RNA sequencing (scRNA-seq) data. Sample identities are defined based on the provided cell type annotations within the Seurat object. Pathway activities are then computed using the decoupleR::run_wmean function, employing a weighted means algorithm. This step retrieves pathway activity information from the scRNA-seq data. Following pathway activity retrieval, three distinct plots were generated: one depicting cell type-specific pathway activities, another grouping pathway activities by IBC and nIBC status, and a third grouping pathway activities by both cell types and IBC status. Each plot includes an average scores heatmap, offering insights into pathway activity variations across different sample subsets. This analysis pipeline provides a comprehensive approach to exploring and visualising pathway activities within scRNA-seq data, enabling insights into cellular functional states and molecular interactions within the studied lineage. I used the do_PathwayActivityPlot from the SCpubr package with the plot_FeaturePlots option set to True, to also obtain feature plots of the pathway activities in UMAP space [171].

5.15 Spatial transcriptomics

The sample collection, sectioning and slide preparation was performed by Dr. Ewgenija Gutjahr at the Institute of Pathology, Dr. Marc Zapatka designed and obtained the gene panel, while the data was recorded by Dr. Jan-Phillip Mallm and the staff of the DKFZ Single-Cell Open Lab. Describing these collaborative efforts, when I write "we", I am referring to me and the mentioned collaborators, as I planned and coordinated these efforts. The data analysis and visualisation was done by me.

Gene panel design We used the Human Breast panel provided by the manufacturer (10X Genomics), containing 280 genes optimised to facilitate cell-type identification. We added another 100 custom genes selected to allow more fine-grained distinguishing of cell types and states (See Supplements 6.1).

Measurement procedure We prepared a Xenium® slide (3000941, 10X Genomics) according to the manufacturer's protocol. To measure multiple samples on the same slide, we prepared a paraffin block containing multiple IBC and control tissue pieces. Care was taken that the size of the tissue array would not exceed the available area for tissue placement on the slide (10.45 mm x 22.45 mm), and that the individual samples would not overlap. We used one section of the block for haematoxylin and eosin staining. We then proceeded to cut a section of 10 μm thickness, floated it in a water bath until it was free of wrinkles and folds, and placed it on the slide, taking care not to obstruct the fiducials indicating the measurement area. After drying at room temperature, the slide was incubated at 42ºC for 3 h, and then stored overnight at room temperature in a desiccator. Imaging was then performed on a fully automated 10X Xenium® Analyzer instrument by the DKFZ single cell open lab core facility. On the Xenium® Analyzer, image acquisition occurs over 15 cycles, encompassing reagent cycling, incubation, fluorescent probe hybridisation, imaging, and probe removal. During each cycle, fluorescently-labelled oligonucleotides bind to amplified barcodes, and the fluorescent intensity in each of the four Xenium® colour channels is measured to identify target genes. Only transcripts with a Quality Value greater than 20 were retained. Z-stacks were captured with a 0.75 μm step size across the entire tissue thickness and subsequently stitched together to create a spatial map of the transcripts across the tissue.

Nuclei segmentation To assign mRNAs to nuclei and facilitate downstream analysis, I utilised DAPI images to detect nuclei, employing the neural network

algorithm provided by 10x Genomics.

Preprocessing I generated a single-cell expression matrix by counting molecules of each gene within the nucleus-bound areas assigned to individual cells by the segmentation algorithm. Cells with fewer than 20 total transcripts within the nucleus were excluded from the analysis. The expression matrix was then log-normalised using the scanpy (v1.10.1) package in combination with squidpy (v1.2.2).

Cell type annotation Data dimensions were reduced using PCA, with the number of components determined via a global maximum likelihood approach based on a translated Poisson mixture model with 20 nearest neighbors, implemented in intrinsicDimension (v1.2.0). Additionally, a Uniform Manifold Approximation and Projection (UMAP) embedding was generated based on the PCA. Cells were then clustered based on the PCA results using Leiden clustering, implemented in scanpy. Clusters were initially manually annotated based on the cell type labels provided with the gene panel (10X Genomics). Then I used tangram (v1.0.4) to project the cell types and states previously identified in single cell data onto spatial data using the tangram.map_cells_to_space function with the mode argument set to "clusters".

Data processing and visualisation I used scanpy(v1.10.1) squidpy (v1.2.2) for data processing and visualisation. I used the scanpy.pl.dotplot function to visualise the expression of marker genes of each cluster. Two-dimensional UMAP embeddings were visualised using the scanpy.pl.umap function, while position of cells in space was plotted by means of the squidpy.pl.spatial_scatter function.
Closeness centrality, average clustering and degree centrality were calculated separately for IBC and non-IBC samples using squid.gr.centrality_scores. The difference of the output matrices was calculated and used as input for the

squidy.pl.centrality_scores function for visualisation.
squidpy.gr.spatial_neighbors and squidpy.gr.nhood_enrichment were used to calculate the neighbourhood enrichment of cell types with other cell types for each sample separately. The difference in means was then calculated and plotted using the seaborn (v0.13.2) with the seaborn..heatmap function. Additionally, the resulting heat map was annotated with the level of significance of the difference in mean values.
Mean shortest distances between one cell of a certain type and another cell of a different type were calculated based on the spatial data using a custom function. This function calculates the mean minimum distance between pairs of specified cell types within samples of a single-cell dataset using spatial coordinates. A list of cell type pairs of interest was defined, each pair including two different cell types for which the spatial distance will be calculated. Unique sample identifiers are then extracted from the Sample column of the adata.obs dataframe. An empty list was initialized to store the results of the distance calculations. The function loops through each sample and each pair of cell types, creating a subset of the data (adata_sample) corresponding to that sample. Within the sample, cells of the two specified types are filtered out, and if either type has no cells in the sample, the function skips to the next pair. The spatial coordinates for the cells of each type are extracted from the obsm['spatial'] attribute. The function computes a pairwise distance matrix between the coordinates of the two cell types using the distance_matrix function from the scipy package (v1.13.0). For each cell in the first type, the minimum distance to any cell in the second type was found, and the mean of these minimum distances was calculated. The mean distance, along with the sample identifier and cell type pair, was appended to the results list. Finally, the results list was converted into a pandas DataFrame for easy manipulation and analysis, and the final results DataFrame was printed, showing the mean minimum distances for each sample and cell type pair. This output summarises the mean minimum spatial distances between the specified pairs of cell types across

different samples in the dataset, providing insights into the spatial relationships and potential interactions between different cell types within the tissue samples.

5.16 Use of Artificial Intelligence for Coding Assistance as well as Thesis Writing, Proofreading, and Typesetting

In the preparation of this thesis, I utilised ChatGPT (versions v3.5, v4, and v4o) for partial assistance in writing code, annotating, and identifying errors. This included R, Python, Bash, and LaTeX code used for data analysis and thesis typesetting. For editorial purposes, ChatGPT (v3.5, v4, and v4o), developed by OpenAI, was also employed to aid in drafting, refining, and editing the text. This assistance helped improve readability, structure, and presentation, including the main body and figure legends, while ensuring the scientific integrity and meaning of the original work remained unchanged. I, as the author, am solely responsible for the scientific content, data interpretation, and final approval of the thesis.

5.17 Graphics Software

Adobe Illustrator CS6 and BioRender were used to design and create figures and illustrations and create their final layout.

Chapter 6

Supplementary Material

6.1 Xenium Gene Panel

Genes included in the 10X Breast Cancer Panel: ABCC11 (Breast cancer), ACTA2 (Smooth muscle cells), ACTG2 (Breast myoepithelial cells), ADAM9 (Breast glandular cells), ADGRE5 (Dendritic cells), ADH1B (Adipocytes), ADIPOQ (Adipocytes), AGR3 (Breast glandular cells), AIF1 (Macrophages), AKR1C1 (Smooth muscle cells), AKR1C3 (Endothelial cells), ALDH1A3 (Breast glandular cells), ANGPT2 (Endothelial cells), ANKRD28 (B cells), ANKRD29 (Adipocytes), ANKRD30A (Epithelial cells), APOBEC3A (B cells), APOBEC3B (Breast cancer), APOC1 (Macrophages), AQP1 (Endothelial cells), AQP3 (Breast glandular cells), AR (Breast cancer), AVPR1A (Smooth muscle cells), BACE2 (Breast glandular cells), BANK1 (B cells), BASP1 (Macrophages), C15orf48 (Breast glandular cells), C1QA (Macrophages), C1QC (Macrophages), C2orf42 (B cells), C5orf46 (Breast glandular cells), C6orf132 (Breast myoepithelial cells), CAV1 (Breast cancer), CAVIN2 (Adipocytes), CCDC6 (Breast glandular cells), CCDC80 (Fibroblasts), CCL5 (T cells), CCL8 (Smooth muscle cells), CCND1 (Breast cancer), CCPG1 (B cells), CCR7 (T cells), CD14 (Monocytes), CD163 (Macrophages), CD19 (B cells), CD247 (NK cells), CD27 (B cells), CD274 (T cells), CD3E (T cells), CD3G (T cells), CD4 (T cells), CD68 (Macrophages), CD69 (T cells), CD79A (B cells), CD79B (B cells), CD80 (Macrophages), CD83 (Dendritic cells), CD86 (Macrophages), CD8A (T cells), CD9 (Breast glandular cells), CD93 (Endothelial cells), CDC42EP1 (Breast glandular cells), CDH1 (Breast cancer), CEACAM6 (Breast cancer), CEACAM8 (Neutrophils), CENPF (Breast cancer), CLEC14A (Adipocytes), CLEC9A (Dendritic cells), CLECL1 (B cells), CLIC6 (Breast glandular cells), CPA3 (Mast cells), CRISPLD2 (Smooth muscle cells), CTH (Smooth muscle cells), CTLA4 (T cells), CTSG (Mast cells), CTTN (Breast glandular cells), CX3CR1 (Macrophages), CXCL12 (Breast cancer), CXCL16 (Macrophages), CXCL5 (Macrophages), CXCR4 (T cells), CYTIP (T cells), DAPK3 (Breast glandular cells), DMKN (Breast glandular cells), DNAAF1 (B cells), DNTTIP1 (Breast glandular cells), DPT (Fibroblasts), DSC2 (Breast glandular cells), DSG2 (Breast

glandular cells), DSP (Breast glandular cells), DST (Breast glandular cells), EHD2 (Adipocytes), ELOVL2 (Breast glandular cells), ELTD1 (Endothelial cells), EMCN (Endothelial cells), ENPP2 (Breast glandular cells), EPHB4 (Endothelial cells), EPO (Breast glandular cells), EREG (Breast cancer), ESM1 (Endothelial cells), F13A1 (Endothelial cells), FABP4 (Adipocytes), FADS1 (Breast glandular cells), FCER1A (Dendritic cells), FCER1G (Macrophages), FCGR2A (Breast glandular cells), FCGR3A (Macrophages), FGFR2 (Breast cancer), FPR1 (Macrophages), FZD10 (Breast cancer), FZD2 (Fibroblasts), GATA3 (Breast glandular cells), GFRA1 (Endothelial cells), GGH (Smooth muscle cells), GIMAP5 (T cells), GJB2 (Breast myoepithelial cells), GJB6 (Breast myoepithelial cells), GPR182 (Endothelial cells), GRAP (T cells), HLA-DMA (Dendritic cells), HLA-DMB (Dendritic cells), HLA-DPA1 (Dendritic cells), HLA-DPB1 (Dendritic cells), HLA-DQA1 (Dendritic cells), HLA-DQA2 (Dendritic cells), HLA-DRA (Dendritic cells), HLA-DRB1 (Dendritic cells), ICAM2 (Endothelial cells), ICOSLG (Macrophages), ID3 (Breast cancer), IER3 (Breast cancer), IFI6 (Breast cancer), IL1B (Macrophages), IL6ST (Breast glandular cells), IL7R (T cells), ITGA6 (Breast cancer), ITGB6 (Breast cancer), ITGB8 (Endothelial cells), JAG1 (Breast glandular cells), JAM3 (Endothelial cells), KANK3 (Fibroblasts), KLK5 (Breast cancer), LAMP3 (Dendritic cells), LCN2 (Macrophages), LDHB (Breast glandular cells), LGALS1 (Breast cancer), LGALS3 (Breast cancer), LIF (Fibroblasts), LRP1 (Breast glandular cells), LUM (Endothelial cells), LYVE1 (Endothelial cells), MALL (Breast glandular cells), MAML2 (Breast cancer), MAP3K8 (T cells), MMP14 (Breast glandular cells), MMP2 (Breast glandular cells), MMP9 (Macrophages), SFRP1 (Breast cancer), SFRP4 (Fibroblasts), SH3YL1 (Breast glandular cells), SLAMF1 (B cells), SLAMF7 (B cells), SLC25A37 (Breast glandular cells), SLC5A6 (Breast glandular cells), SMAP2 (T cells), SMS (Breast glandular cells), SNAI1 (Breast cancer), SOX17 (Endothelial cells), SOX18 (Adipocytes), SPIB (Dendritic cells), SQLE (Breast glandular cells), SRPK1 (Breast glandular cells), SSTR2 (Smooth muscle cells), STC1 (Endothelial cells), SVIL (Breast glandular cells), TAC1 (Fibrob-

lasts), TACSTD2 (Breast cancer), TCEAL7 (Fibroblasts), TCF4 (Breast cancer), TCF7 (T cells), TCIM (Breast glandular cells), TCL1A (B cells), TENT5C (T cells), TFAP2A (Adipocytes), THAP2 (Smooth muscle cells), TIFA (Endothelial cells), TIGIT (T cells), TIMP4 (Adipocytes), TMEM147 (Breast glandular cells), TNFRSF17 (Plasma cells), TOMM7 (Breast glandular cells), TOP2A (Breast cancer), TPD52 (Breast glandular cells), TPSAB1 (Mast cells), TRAC (T cells), TRAF4 (Breast myoepithelial cells), TRAPPC3 (Breast glandular cells), TRIB1 (B cells), TUBA4A (T cells), TUBB2B (Breast myoepithelial cells), UCP1 (Adipocytes), USP53 (Breast glandular cells), VOPP1 (B cells), VWF (Endothelial cells), WARS (Endothelial cells), ZEB1 (Breast cancer), ZEB2 (Breast cancer), ZNF562 (Breast glandular cells).

Custom Genes: BATF3, CSPG4, CXCL10, IL4I1, TNC, TREM1, CDH11, DTL, EPS8, FIGN, KRT19, RIMS2, VEGFA, ZFPM2, APOE, CCL2, CCL3, CD74, CPM, HIF1A, HLA-DQA1, HLA-DRA, HSPH1, IDO1, ITGA4, KCNQ3, MSR1, NAMPT, NR4A3, P2RY12, SIGLEC8, SLC16A10, SLC8A1, SNTB1, TFEC, TNF, TPRG1, TREM2, VCAN, AFF3, NAV3, TNFAIP3, CXCL13, IFI44L, ANO2, ARHGAP15, ARHGAP24, CACNB2, CEP128, CLU, COL5A1, DMXL2, ENPEP, FGFR2, FOS, FRMD5, HECW2, IGHGP, ITGA1, NR4A1, NRXN1, RGS1, SGIP1, THBS1, TIMP1, ZBTB16, AL136456.1, AOAH, CD40LG, GZMB, KLRC2, LEF1, NELL2, PDE3B, PTPN13, RUNX2, THEMIS, ITGAE, ARHGAP8, BAIAP2L1, CAP2, ELF3, ESRP1, FAM160A1, GRHL2, RHPN2, SYCP2, PLVAP, CLDN5, CD34, FCN1, DCN, IGLL5, JCHAIN, POU2AF1, F13A1, LYVE1, CD209, KRT17, S100A9.

6.2 List of Abbreviations

ALDH1 Aldehyde Dehydrogenase 1

BCL Base Call Format (used in sequencing)

BATF Basic Leucine Zipper ATF-Like Transcription Factor

CAF Cancer-Associated Fibroblast

CD3 Cluster of Differentiation 3

CD4 Cluster of Differentiation 4

CD8 Cluster of Differentiation 8

CD44 Cluster of Differentiation 44

CXCL C-X-C Motif Chemokine Ligand

CXCR Chemokine Receptor

CXCR4 C-X-C Motif Chemokine Receptor 4

DNA Deoxyribonucleic Acid

EMT Epithelial-to-Mesenchymal Transition

EpCAM Epithelial Cell Adhesion Molecule

ERK Extracellular Signal-Regulated Kinase

FASTQ Sequence Read Format (used in sequencing)

FACS Fluorescence-Activated Cell Sorting

FGF Fibroblast Growth Factor

FOXP3 Forkhead Box P3

FFPE Formalin-Fixed Paraffin-Embedded

H&E Haematoxylin and Eosin

HMGB1 High-Mobility Group Box 1

HSP90B1 Heat Shock Protein 90 Beta Family Member 1

HTO Hashtag Oligonucleotides (used in sample hashing)

IBC Inflammatory Breast Cancer

IFN Interferon

IFNγ Interferon Gamma

IGF Insulin-Like Growth Factor

IGF1 Insulin-like Growth Factor 1

IL-1B Interleukin-1 Beta

IL-1β Interleukin-1 Beta

IL-4 Interleukin-4

IL-4I1 Interleukin 4 Induced 1

JAK-STAT Janus Kinase-Signal Transducer and Activator of Transcription

MAPK Mitogen-Activated Protein Kinase

MKI67 Marker of Proliferation Ki-67

NCAM1 Neural Cell Adhesion Molecule 1

NFKB Nuclear Factor Kappa-Light-Chain-Enhancer of Activated B Cells

NLRP3 NLR Family Pyrin Domain Containing 3

NK Natural Killer (Cells)

PD-1/**PD-L1** Programmed Cell Death Protein 1/Programmed Death-Ligand 1

PDGF Platelet-Derived Growth Factor

PI3K Phosphoinositide 3-Kinase

PCA Principal Component Analysis

QC Quality Control

RAS Rat Sarcoma

RNA Ribonucleic Acid

S4 Illumina NovaSeq 6000 Sequencing Chemistry (used in sequencing platforms)

scDblFinder Single-Cell Doublet Finder

scRNA-seq Single-Cell RNA Sequencing

scVI Single-Cell Variational Inference

SNP Single-Nucleotide Polymorphism

TAM Tumour-Associated Macrophages

TGF-β Transforming Growth Factor Beta

TLR Toll-Like Receptors

UMAP Uniform Manifold Approximation and Projection

UMI Unique Molecular Identifier

VEGF Vascular Endothelial Growth Factor

6.3 List of Figures

1.1 Modifiable lifestyle factors for breast cancer. 5
1.2 The Breast Cancer Microenvironment. 10
1.3 Emerging targeted therapies for breast cancer. 16
1.4 Single cell analysis best practices. 26
1.5 Spatial transcriptomics analysis best practices. 31

3.1 Flow cytometric sorting of primary samples for single-cell library preparation. 39
3.2 Labelling efficiency of cells with hashing antibodies. 41
3.3 Additional cell recovery resulting from combined hashing and SNP-based demultiplexing. 42
3.4 Exemplary output plot of the SoupX function. 44
3.5 Homotypic doublet identification. 45
3.6 Single-Cell Data Quality Control. 46
3.7 Overview of the Wu dataset's properties. 48
3.8 Overview of the Tietscher dataset's properties. 50
3.9 Data harmonisation using Harmony. 54
3.10 Deep Learning-Based Harmonisation. 55
3.11 Celltype Marker Dotplot. 56
3.12 Cell Type Proportions. 58
3.13 Assignment of Malignant Cells. 60
3.14 Overview of the fibroblast subsetting and reclustering process. . . . 62

3.15 Comparison of cell type proportions between IBC and non-IBC samples. 63
3.16 Differential interaction numbers and strengths in IBC and non-IBC. . 64
3.17 Outgoing and incoming interaction strength in IBC and non-IBC . . . 65
3.18 CAF signalling induces an inflammatory response in myeloid cells . 68
3.19 Overview of the Myeloid Subsetting and Reclustering Process 70
3.20 Comparison of cell type proportions between IBC and non-IBC samples. 71
3.21 Properties of IL1B Monocytes 72
3.22 IBC Myeloid Cells Express Key Inflammation Regulators 74
3.23 The Cancer Cell - IL1B Monocyte - Treg Signalling Axis 76
3.24 Overview of the Myeloid Subsetting and Reclustering Process 78
3.25 Heatmap of normalised mean counts for key marker genes across T and NK cell clusters. 80
3.26 Comparison of T and NK cell type proportions between IBC and non-IBC samples. 81
3.27 IBC T Cells Show More Pronounced Exhaustion. 84
3.28 TGFβ Expression in the Microenvironments of IBC and non-IBC . . 85
3.29 TGFβ Is More Highly Expressed in IBC 86
3.30 Subsetting and Cell Type Assignment of Endothelial Cells. 88
3.31 Endothelial Ligand-Receptor Analysis. 90
3.32 Endothelial Cells in IBC Upregulate Proliferation-Associated Pathways 92
3.33 Xenium Slide Assembly and Data Recording 94
3.34 Xenium Data Quality Control 95
3.35 Clustering and Cell Type Assignment 97
3.36 Comparison of Manual and Automated Cell Cype Assignment 99
3.37 Relative neighbourhood enrichment in IBC compared to non-IBC . . 101
3.38 Spatial distribution of epithelial/cancer cells relative to T cells in IBC and non-IBC tissues. 103

4.1 Single-Cell Dataset Size Over Time 109
4.2 Suggested Model of IBC Biology. 111

6.4 List of Tables

1.1 Overview of rare breast cancers and their frequency relative to all breast cancers [75] 18

3.1 Patient IDs, IBC status and molecular subtypes and processing batch of sample collective 36
3.2 Cell numbers during pre-processing for single-cell library preparation. Recovery was based on number of sorted cells, where available, otherwise the initial cell count was used. 37
3.3 Numbers of cells per molecular subtype in the final dataset of IBC and non-IBC. 51

Chapter 7

Bibliography

[1] Hyuna Sung et al. “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries”. In: *CA: A Cancer Journal for Clinicians* 71 (3 May 2021), pp. 209–249. ISSN: 1542-4863. DOI: 10.3322/CAAC.21660.

[2] Nadia Harbeck et al. “Breast cancer”. In: *Nature Reviews Disease Primers* 5 (1 Dec. 2019), pp. 1–31. ISSN: 2056676X. DOI: 10.1038/s41572-019-0111-2.

[3] Ibiayi Dagogo-Jack and Alice T. Shaw. “Tumour heterogeneity and resistance to cancer therapies”. In: *Nature Reviews Clinical Oncology* 15 (2 Feb. 2018), pp. 81–94. ISSN: 17594782. DOI: 10.1038/nrclinonc.2017.166.

[4] Charles M. Perou et al. “Molecular portraits of human breast tumours”. In: *Nature* 406 (6797 Aug. 2000), pp. 747–752. ISSN: 00280836. DOI: 10.1038/35021093.

[5] Agata Szymiczek, Amna Lone, and Mohammad R. Akbari. “Molecular intrinsic versus clinical subtyping in breast cancer: A comprehensive review”. In: *Clinical Genetics* 99 (5 May 2021), pp. 613–637. ISSN: 0009-9163. DOI: 10.1111/cge.13900.

[6] Tracy-Ann Moo et al. “Overview of Breast Cancer Therapy”. In: *PET Clinics* 13 (3 July 2018). doi: 10.1016/j.cpet.2018.02.006, pp. 339–354. ISSN: 1556-8598. DOI: 10.1016/j.cpet.2018.02.006. URL: https://doi.org/10.1016/j.cpet.2018.02.006.

[7] Noam F. Pondé, Dimitrios Zardavas, and Martine Piccart. “Progress in adjuvant systemic therapy for breast cancer”. In: *Nature Reviews Clinical Oncology* 16 (1 Jan. 2019), pp. 27–44. ISSN: 1759-4774. DOI: 10.1038/s41571-018-0089-9.

[8] Fatima Cardoso et al. “European Breast Cancer Conference manifesto on breast centres/units”. In: *European Journal of Cancer* 72 (Feb. 2017), pp. 244–250. ISSN: 09598049. DOI: 10.1016/j.ejca.2016.10.023.

[9] Mojisola Oluwasanu and Olufunmilayo I Olopade. “Global disparities in breast cancer outcomes: new perspectives, widening inequities, unanswered questions”. In: *The Lancet Global Health* 8 (8 Aug. 2020). doi: 10.1016/S2214-109X(20)30307-7, e978–e979. ISSN: 2214-109X. DOI: 10.1016/S2214-109X(20)30307-7. URL: https://doi.org/10.1016/S2214-109X(20)30307-7.

[10] J. Ferlay et al. “Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods”. In: *International Journal of Cancer* 144 (8 Apr. 2019), pp. 1941–1953. ISSN: 10970215. DOI: 10.1002/ijc.31937. URL: https://onlinelibrary.wiley.com/doi/full/10.1002/ijc.31937%20https://onlinelibrary.wiley.com/doi/abs/10.1002/ijc.31937%20https://onlinelibrary.wiley.com/doi/10.1002/ijc.31937.

[11] Beata Smolarz, Anna Zadrożna Nowak, and Hanna Romanowicz. “Breast Cancer—Epidemiology, Classification, Pathogenesis and Treatment (Review of Literature)”. In: *Cancers* 14 (10 May 2022), p. 2569. ISSN: 2072-6694. DOI: 10.3390/cancers14102569.

[12] Martine Bellanger et al. “Are global breast cancer incidence and mortality patterns related to country-specific economic development and prevention strategies?” In: *Journal of global oncology* (2018).

[13] Carol E DeSantis et al. “Breast cancer statistics, 2019”. In: *CA: a cancer journal for clinicians* 69 (6 2019), pp. 438–451. ISSN: 0007-9235.

[14] Ophira Ginsburg et al. “The global burden of women’s cancers: a grand challenge in global health”. In: *The Lancet* 389 (10071 2017), pp. 847–860. ISSN: 0140-6736.

[15] Klim McPherson, CaMa Steel, and J M Dixon. “ABC of breast diseases: breast cancer—epidemiology, risk factors, and genetics”. In: *BMJ: British Medical Journal* 321 (7261 2000), p. 624.

[16] Philippe Autier et al. “Disparities in breast cancer mortality trends between 30 European countries: retrospective trend analysis of WHO mortality database”. In: *Bmj* 341 (2010). ISSN: 0959-8138.

[17] Michael G Marmot et al. “The benefits and harms of breast cancer screening: an independent review”. In: *British journal of cancer* 108 (11 2013), pp. 2205–2240. ISSN: 1532-1827.

[18] Jacques Ferlay et al. “Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012”. In: *International journal of cancer* 136 (5 2015), E359–E386. ISSN: 0020-7136.

[19] Amanda Dibden et al. “Worldwide review and meta-analysis of cohort studies measuring the effect of mammography screening programmes on incidence-based breast cancer mortality”. In: *Cancers* 12 (4 2020), p. 976. ISSN: 2072-6694.

[20] Catherine Duggan et al. “National health system characteristics, breast cancer stage at diagnosis, and breast cancer mortality: a population-based analysis”. In: *The Lancet Oncology* 22 (11 2021), pp. 1632–1642. ISSN: 1470-2045.

[21] Joannie Lortet-Tieulent et al. “Profiling global cancer incidence and mortality by socioeconomic development”. In: *International journal of cancer* 147 (11 2020), pp. 3029–3036. ISSN: 0020-7136.

[22] Carol E. DeSantis et al. “Breast cancer statistics, 2017, racial disparity in mortality by state”. In: *CA: A Cancer Journal for Clinicians* 67 (6 Nov. 2017), pp. 439–448. ISSN: 0007-9235. DOI: 10.3322/caac.21412.

[23] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. “Cancer statistics, 2019”. In: *CA: a cancer journal for clinicians* 69 (1 2019), pp. 7–34. ISSN: 0007-9235.

[24] Mary-Claire King, Joan H Marks, and Jessica B Mandell. “Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2”. In: *Science* 302 (5645 2003), pp. 643–646. ISSN: 0036-8075.

[25] Darcy L. Thull and Victor G. Vogel. “Recognition and Management of Hereditary Breast Cancer Syndromes”. In: *The Oncologist* 9 (1 Feb. 2004), pp. 13–24. ISSN: 1083-7159. DOI: 10.1634/theoncologist.9-1-13.

[26] Clare Turnbull and Nazneen Rahman. “Genetic Predisposition to Breast Cancer: Past, Present, and Future”. In: *Annual Review of Genomics and Human Genetics* 9 (1 Sept. 2008), pp. 321–345. ISSN: 1527-8204. DOI: 10.1146/annurev.genom.9.081307.164339.

[27] Steven K Clinton, Edward L Giovannucci, and Stephen D Hursting. “The world cancer research fund/American institute for cancer research third expert report on diet, nutrition, physical activity, and cancer: impact and future directions”. In: *The Journal of nutrition* 150 (4 2020), pp. 663–671. ISSN: 0022-3166.

[28] Jalal Poorolajal et al. “Factors for the Primary Prevention of Breast Cancer: A Meta-Analysis of Prospective Cohort Studies.” In: *Journal of research in health sciences* 21 (3 July 2021), e00520. ISSN: 2228-7809. DOI: 10.34172/jrhs.2021.57.

[29] Timothy J Key, Pia K Verkasalo, and Emily Banks. “Epidemiology of breast cancer”. In: *The lancet oncology* 2 (3 2001), pp. 133–140. ISSN: 1470-2045.

[30] Robert Baan et al. “Carcinogenicity of alcoholic beverages”. In: *The lancet oncology* 8 (4 2007), pp. 292–293. ISSN: 1470-2045.

[31] Ruth C Travis and Timothy J Key. “Oestrogen exposure and breast cancer risk”. In: *Breast Cancer Research* 5 (2003), pp. 1–9.

[32] Garnet L Anderson and Marian L Neuhouser. “Obesity and the risk for premenopausal and postmenopausal breast cancer”. In: *Cancer prevention research* 5 (4 2012), pp. 515–521. ISSN: 1940-6207.

[33] Sherene Loi et al. “Obesity and outcomes in premenopausal and postmenopausal breast cancer”. In: *Cancer Epidemiology Biomarkers & Prevention* 14 (7 2005), pp. 1686–1691. ISSN: 1055-9965.

[34] J. Gill. “The Effects of Moderate Alcohol Consumption on Female Hormone Levels and Reproductive Function”. In: *Alcohol and Alcoholism* 35 (5 Sept. 2000), pp. 417–423. ISSN: 14643502. DOI: 10.1093/alcalc/35.5.417.

[35] Evelyn M Monninkhof et al. “Physical activity and breast cancer: a systematic review”. In: *epidemiology* 18 (1 2007), pp. 137–157. ISSN: 1044-3983.

[36] Mary S Wolff et al. “Breast cancer and environmental risk factors: epidemiological and experimental findings”. In: *Annual review of pharmacology and toxicology* 36 (1 1996), pp. 573–596. ISSN: 0362-1642.

[37] Steven A Narod. “Hormone replacement therapy and the risk of breast cancer”. In: *Nature Reviews Clinical Oncology* 8 (11 2011), pp. 669–676. ISSN: 1759-4782. DOI: 10.1038/nrclinonc.2011.110. URL: https://doi.org/10.1038/nrclinonc.2011.110.

[38] Lakhani et al. “WHO Classification of Tumours of the Breast, Fourth Edition”. In: 2012. ISBN: 9789283224334.

[39] Therese Sørlie et al. “Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications”. In: *Proceedings of the National Academy of Sciences* 98 (19 Sept. 2001), pp. 10869–10874. ISSN: 0027-8424. DOI: 10.1073/pnas.191367098.

[40] Soonmyung Paik et al. “A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer”. In: *New England Journal of Medicine* 351 (27 Dec. 2004), pp. 2817–2826. ISSN: 0028-4793. DOI: 10.1056/NEJMoa041588. URL: http://www.nejm.org/doi/abs/10.1056/NEJMoa041588.

[41] Harvey N. Mayrovitz, ed. *Breast Cancer*. Exon Publications, Aug. 2022. ISBN: 9780645332032. DOI: 10.36255/exon-publications-breast-cancer.

[42] Adedayo A Onitilo et al. “Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival”. In: *Clinical medicine & research* 7 (1-2 2009), pp. 4–13. ISSN: 1539-4182.

[43] André Hennigs et al. “Prognosis of breast cancer molecular subtypes in routine clinical care: A large prospective cohort study”. In: *BMC cancer* 16 (2016), pp. 1–9.

[44] Lyndsay Harris et al. “American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer”. In: *Journal of clinical oncology* 25 (33 2007), pp. 5287–5312. ISSN: 0732-183X.

[45] Carsten Denkert et al. “Molecular alterations in triple-negative breast cancer—the road to new treatment strategies”. In: *The Lancet* 389 (10087 2017), pp. 2430–2442. ISSN: 0140-6736.

[46] Joseph A Sparano et al. “Adjuvant chemotherapy guided by a 21-gene expression assay in breast cancer”. In: *New England Journal of Medicine* 379 (2 2018), pp. 111–121. ISSN: 0028-4793.

[47] Laura J Van’t Veer et al. “Gene expression profiling predicts clinical outcome of breast cancer”. In: *nature* 415 (6871 2002), pp. 530–536. ISSN: 0028-0836.

[48] C Pixberg et al. “COGNITION: a prospective precision oncology trial for patients with early breast cancer at high risk following neoadjuvant chemotherapy”. In: *ESMO Open* 7 (6 2022), p. 100637. ISSN: 2059-7029. DOI: https://doi.org/10.1016/j.esmoop.2022.100637. URL: https://www.sciencedirect.com/science/article/pii/S205970292200271X.

[49] Mario Hlevnjak et al. “CATCH: A Prospective Precision Oncology Trial in Metastatic Breast Cancer”. In: *JCO Precision Oncology* (5 Nov. 2021), pp. 676–686. ISSN: 2473-4284. DOI: 10.1200/PO.20.00248.

[50] Andrew E Place, Sung Jin Huh, and Kornelia Polyak. “The microenvironment in breast cancer progression: biology and implications for treatment”. In: *Breast Cancer Research* 13 (6 Dec. 2011), p. 227. ISSN: 1465-542X. DOI: 10.1186/bcr2912.

[51] Gil Friedman et al. “Cancer-associated fibroblast compositions change with breast cancer progression linking the ratio of S100A4+ and PDPN+ CAFs to clinical outcome”. In: *Nature Cancer* (June 2020), pp. 1–17. ISSN: 2662-1347. DOI: 10.1038/s43018-020-0082-y. URL: http://www.nature.com/articles/s43018-020-0082-y.

[52] Dario A. A. Vignali, Lauren W. Collison, and Creg J. Workman. “How regulatory T cells work”. In: *Nature Reviews Immunology* 8 (7 July 2008), pp. 523–532. ISSN: 1474-1733. DOI: 10.1038/nri2343.

[53] Filippo Veglia, Emilio Sanseviero, and Dmitry I. Gabrilovich. “Myeloid-derived suppressor cells in the era of increasing myeloid cell diversity”. In: *Nature Reviews Immunology* 21 (8 Aug. 2021), pp. 485–498. ISSN: 1474-1733. DOI: 10.1038/s41577-020-00490-y.

[54] Roberto Salgado et al. “The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014”. In: *Annals of oncology* 26 (2 2015), pp. 259–271. ISSN: 0923-7534.

[55] Tjoung-Won Park-Simon et al. “Arbeitsgemeinschaft Gynäkologische Onkologie Recommendations for the Diagnosis and Treatment of Patients with Early Breast Cancer: Update 2023”. In: *Breast Care* 18 (4 2023), pp. 289–305. ISSN: 1661-3791. DOI: 10.1159/000531578.

[56] Umberto Veronesi et al. “Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer”. In: *New England Journal of Medicine* 347 (16 2002), pp. 1227–1232. ISSN: 0028-4793.

[57] Bernard Fisher et al. “Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer”. In: *New England Journal of Medicine* 347 (16 2002), pp. 1233–1241. ISSN: 0028-4793.

[58] Krishna B Clough et al. “Improving breast cancer surgery: a classification and quadrant per quadrant atlas for oncoplastic surgery”. In: *Annals of surgical oncology* 17 (2010), pp. 1375–1391. ISSN: 1068-9265.

[59] Jan Haussmann et al. “Recent advances in radiotherapy of breast cancer”. In: *Radiation Oncology* 15 (1 Dec. 2020), p. 71. ISSN: 1748-717X. DOI: 10.1186/s13014-020-01501-x.

[60] Cornelia Liedtke et al. “Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer”. In: *Journal of clinical oncology* 26 (8 2008), pp. 1275–1281. ISSN: 0732-183X.

[61] Harold J Burstein et al. “American Society of Clinical Oncology clinical practice guideline update on the use of chemotherapy sensitivity and resistance assays”. In: *Journal of Clinical Oncology* 29 (24 2011), pp. 3328–3330. ISSN: 0732-183X.

[62] Early Breast Cancer Trialists’ Collaborative Group. “Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-

year survival: an overview of the randomised trials". In: *The Lancet* 365 (9472 2005), pp. 1687–1717. ISSN: 0140-6736.

[63] Dennis J Slamon et al. "Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2". In: *New England journal of medicine* 344 (11 2001), pp. 783–792. ISSN: 0028-4793.

[64] Siddharth Kunte, Jame Abraham, and Alberto J. Montero. "Novel HER2–targeted therapies for HER2–positive metastatic breast cancer". In: *Cancer* 126 (19 Oct. 2020), pp. 4278–4288. ISSN: 0008-543X. DOI: 10.1002/cncr.33102.

[65] Laura Cortesi, Hope S Rugo, and Christian Jackisch. "An overview of PARP inhibitors for the treatment of breast cancer". In: *Targeted oncology* 16 (3 2021), pp. 255–282. ISSN: 1776-2596.

[66] Leisha A. Emens. "Breast Cancer Immunotherapy: Facts and Hopes". In: *Clinical Cancer Research* 24 (3 Feb. 2018), pp. 511–520. ISSN: 1078-0432. DOI: 10.1158/1078-0432.CCR-16-3001.

[67] Peter Schmid et al. "Atezolizumab and nab-paclitaxel in advanced triple-negative breast cancer". In: *New England Journal of Medicine* 379 (22 2018), pp. 2108–2121. ISSN: 0028-4793.

[68] Bernard Asselain et al. "Long-term outcomes for neoadjuvant versus adjuvant chemotherapy in early breast cancer: meta-analysis of individual patient data from ten randomised trials". In: *The Lancet Oncology* 19 (1 Jan. 2018). doi: 10.1016/S1470-2045(17)30777-5, pp. 27–39. ISSN: 1470-2045. DOI: 10.1016/S1470-2045(17)30777-5. URL: https://doi.org/10.1016/S1470-2045(17)30777-5.

[69] Catarina Roma-Rodrigues et al. "Targeting Tumor Microenvironment for Cancer Therapy". In: *International Journal of Molecular Sciences* 20 (4 Feb. 2019), p. 840. ISSN: 1422-0067. DOI: 10.3390/ijms20040840.

[70] Lironne Wein and Sherene Loi. “Mechanisms of resistance of chemotherapy in early-stage triple negative breast cancer (TNBC)”. In: *Breast* 34 (Aug. 2017), S27–S30. ISSN: 15323080. DOI: 10.1016/j.breast.2017.06.023.

[71] Angela DeMichele, Douglas Yee, and Laura Esserman. *Mechanisms of resistance to neoadjuvant chemotherapy in breast cancer*. Dec. 2017. DOI: 10.1056/NEJMcibr1711545. URL: https://pubmed.ncbi.nlm.nih.gov/29211674/.

[72] Pierre Farmer et al. “A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer”. In: *Nature Medicine* 15 (1 2009), pp. 68–74. ISSN: 10788956. DOI: 10.1038/nm.1908.

[73] George S. Karagiannis et al. “Neoadjuvant chemotherapy induces breast cancer metastasis through a TMEM-mediated mechanism”. In: *Science Translational Medicine* 9 (397 2017), pp. 1–16. ISSN: 19466242. DOI: 10.1126/scitranslmed.aan0026.

[74] Haeun Hwangbo et al. “Additivity predicts the efficacy of most approved combination therapies for advanced cancer”. In: *Nature Cancer* 4 (12 Nov. 2023), pp. 1693–1704. ISSN: 2662-1347. DOI: 10.1038/s43018-023-00667-z.

[75] Sarah Jenkins et al. “Rare Breast Cancer Subtypes”. In: *Current oncology reports* 23 (5 May 2021). ISSN: 1534-6269. DOI: 10.1007/S11912-021-01048-4. URL: https://pubmed.ncbi.nlm.nih.gov/33755810/.

[76] Tejaswini P. Reddy et al. “A comprehensive overview of metaplastic breast cancer: clinical features and molecular aberrations”. In: *Breast Cancer Research* 22 (1 Dec. 2020), pp. 1–11. ISSN: 1465542X. DOI: 10.1186/S13058-020-01353-Z/TABLES/2. URL: https://breast-cancer-research.biomedcentral.com/articles/10.1186/s13058-020-01353-z.

[77] Emily R. McMullen, Nicholas A. Zoumberos, and Celina G. Kleer. “Metaplastic Breast Carcinoma: Update on Histopathology and Molecular Alterations”. In: *Archives of Pathology and Laboratory Medicine* 143 (12 Dec. 2019), pp. 1492–1496. ISSN: 0003-9985. DOI: 10.5858/ARPA.2019-0396-RA. URL: https://meridian.allenpress.com/aplm/article/143/12/1492/433835/Metaplastic-Breast-Carcinoma-Update-on.

[78] Rebecca A Nelson et al. “Survival Outcomes of Metaplastic Breast Cancer Patients: Results from a US Population-based Analysis”. In: *Oncol* 22 (2015), pp. 24–31. DOI: 10.1245/s10434-014-3890-4.

[79] N. Tray, J. Taff, and S. Adams. “Therapeutic landscape of metaplastic breast cancer”. In: *Cancer treatment reviews* 79 (Sept. 2019). ISSN: 1532-1967. DOI: 10.1016/J.CTRV.2019.08.004. URL: https://pubmed.ncbi.nlm.nih.gov/31491663/.

[80] Alessandro Inno et al. “Neuroendocrine Carcinoma of the Breast: Current Evidence and Future Perspectives.” In: *The oncologist* 21 (1 Jan. 2016), pp. 28–32. ISSN: 1549-490X. DOI: 10.1634/theoncologist.2015-0309.

[81] Julia Y. Tsang and Gary M. Tse. “Breast cancer with neuroendocrine differentiation: an update based on the latest WHO classification”. In: *Modern Pathology* 34 (6 June 2021), pp. 1062–1073. ISSN: 08933952. DOI: 10.1038/s41379-021-00736-7.

[82] Azzurra Irelli et al. “Neuroendocrine Cancer of the Breast: A Rare Entity.” In: *Journal of clinical medicine* 9 (5 May 2020). ISSN: 2077-0383. DOI: 10.3390/jcm9051452.

[83] Marion Lavigne et al. “Comprehensive clinical and molecular analyses of neuroendocrine carcinomas of the breast.” In: *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 31 (1 Jan. 2018), pp. 68–82. ISSN: 1530-0285. DOI: 10.1038/modpathol.2017.107.

[84] Burton J Lee. “Inflammatory carcinoma of the breast”. In: *Surg Gynecol Obstet* 39 (1924), pp. 580–595.

[85] S Chang et al. “Inflammatory breast carcinoma incidence and survival: the surveillance, epidemiology, and end results program of the National Cancer Institute, 1975-1992.” In: *Cancer* 82 (12 June 1998), pp. 2366–72. ISSN: 0008-543X.

[86] Kenneth W Hance et al. “Trends in inflammatory breast carcinoma incidence and survival: the surveillance, epidemiology, and end results program at the National Cancer Institute.” In: *Journal of the National Cancer Institute* 97 (13 July 2005), pp. 966–75. ISSN: 1460-2105. DOI: 10.1093/jnci/dji172.

[87] Hannah Grace Abraham et al. “Incidence and survival of inflammatory breast cancer between 1973 and 2015 in the SEER database”. In: *Breast Cancer Research and Treatment* 185 (1 2021), pp. 229–238. ISSN: 1573-7217. DOI: 10.1007/s10549-020-05938-2. URL: https://doi.org/10.1007/s10549-020-05938-2.

[88] CD Haagensen. “Inflammatory carcinoma”. In: *Diseases of the Breast* (1971).

[89] Arjun Menta et al. “Inflammatory breast cancer: what to know about this unique, aggressive breast cancer”. In: *Surgical Clinics* 98 (4 2018), pp. 787–800. ISSN: 0039-6109.

[90] I A Jaiyesimi, A U Buzdar, and G Hortobagyi. “Inflammatory breast cancer: a review”. In: *J Clin Oncol* 10 (1992), pp. 1014–1024.

[91] H Masuda et al. “Long-term treatment efficacy in primary inflammatory breast cancer by hormonal receptor-and HER2-defined subtypes”. In: *Annals of oncology* 25 (2 2014), pp. 384–391. ISSN: 0923-7534.

[92] D Dano et al. “Metastatic inflammatory breast cancer: survival outcomes and prognostic factors in the national, multicentric, and real-life French cohort (ESME).” In: *ESMO open* 6 (4 Aug. 2021), p. 100220. ISSN: 2059-7029. DOI: 10.1016/j.esmoop.2021.100220.

[93] Jennifer A. Schlichting et al. “Inflammatory and non-inflammatory breast cancer survival by socioeconomic position in the Surveillance, Epidemiology, and End Results database, 1990–2008”. In: *Breast Cancer Research and Treatment* 134 (3 Aug. 2012), pp. 1257–1268. ISSN: 0167-6806. DOI: 10.1007/s10549-012-2133-2.

[94] Tamer M Fouad et al. “Inflammatory breast cancer: a proposed conceptual shift in the UICC–AJCC TNM staging system”. In: *The Lancet Oncology* 18 (4 2017), e228–e232. ISSN: 1470-2045.

[95] Grantley W Taylor and Adolph Meltzer. ““Inflammatory carcinoma” of the breast”. In: *The American Journal of Cancer* 33 (1 1938), pp. 33–49. ISSN: 0099-7374.

[96] G. Cserni, E. Charafe-Jauffret, and P.J. van Diest. “Inflammatory breast cancer: The pathologists’ perspective”. In: *European Journal of Surgical Oncology* 44 (8 Aug. 2018), pp. 1128–1134. ISSN: 07487983. DOI: 10.1016/j.ejso.2018.04.001.

[97] Erminia Manfrin et al. “Comparison between invasive breast cancer with extensive peritumoral vascular invasion and inflammatory breast carcinoma: a clinicopathologic study of 161 cases”. In: *American Journal of Clinical Pathology* 142 (3 2014), pp. 299–306. ISSN: 1943-7722.

[98] Sepi Mahooti et al. “Breast carcinomatous tumoral emboli can result from encircling lymphovasculogenesis rather than lymphovascular invasion”. In: *Oncotarget* 1 (2 2010), p. 131.

[99] Thomas Bryant. *The diseases of the breast*. Cassell, 1887.

[100] Archibald Leitch. “Peau d’orange in acute mammary carcinoma: Its cause and diagnostic value.” In: *The Lancet* 174 (4490 1909), pp. 861–863. ISSN: 0140-6736.

[101] Heather L Lehman et al. “Modeling and characterization of inflammatory breast cancer emboli grown in vitro”. In: *International journal of cancer* 132 (10 2013), pp. 2283–2294. ISSN: 0020-7136.

[102] Jay Arora et al. “Inflammatory breast cancer tumor emboli express high levels of anti-apoptotic proteins: use of a quantitative high content and high-throughput 3D IBC spheroid assay to identify targeting strategies”. In: *Oncotarget* 8 (16 2017), p. 25848.

[103] Robert H Hester, Gabriel N Hortobagyi, and Bora Lim. “Inflammatory breast cancer: early recognition and diagnosis is critical”. In: *American journal of obstetrics and gynecology* 225 (4 2021), pp. 392–396. ISSN: 0002-9378.

[104] Steven J Van Laere et al. “Uncovering the molecular secrets of inflammatory breast cancer biology: an integrated analysis of three distinct affymetrix gene expression datasets”. In: *Clinical cancer research* 19 (17 2013), pp. 4685–4696. ISSN: 1078-0432.

[105] Ilse Van der Auwera et al. “Integrated miRNA and mRNA expression profiling of the inflammatory breast cancer subtype”. In: *British journal of cancer* 103 (4 2010), pp. 532–541. ISSN: 1532-1827.

[106] Wendy A Woodward et al. “Genomic and expression analysis of microdissected inflammatory breast cancer”. In: *Breast cancer research and treatment* 138 (2013), pp. 761–772. ISSN: 0167-6806.

[107] F Bertucci et al. “Gene expression profiles of inflammatory breast cancer: correlation with response to neoadjuvant chemotherapy and metastasis-free survival”. In: *Annals of Oncology* 25 (2 2014), pp. 358–365. ISSN: 0923-7534.

[108] Emmanuelle Charafe-Jauffret et al. “Aldehyde dehydrogenase 1–Positive cancer stem cells mediate metastasis and poor clinical outcome in inflammatory breast cancer”. In: *Clinical cancer research* 16 (1 2010), pp. 45–55. ISSN: 1078-0432.

[109] Laura E Stevens et al. “JAK-STAT Signaling in Inflammatory Breast Cancer Enables Chemotherapy-Resistant Cell States.” In: *Cancer research* 83 (2 Jan. 2023), pp. 264–284. ISSN: 1538-7445. DOI: 10.1158/0008-5472.CAN-22-0423.

[110] Bora Lim et al. “Inflammatory breast cancer biology: the tumour microenvironment is key”. In: *Nature Reviews Cancer* 18 (8 Aug. 2018), pp. 485–499. ISSN: 14741768. DOI: 10.1038/s41568-018-0010-y. URL: https://pubmed.ncbi.nlm.nih.gov/29703913/.

[111] Steven Van Laere et al. “Is there a role for mammary stem cells in inflammatory breast carcinoma? A review of evidence from cell line, animal model, and human tissue sample experiments”. In: *Cancer* 116 (S11 2010), pp. 2794–2805. ISSN: 0008-543X.

[112] Xiaotong Li et al. “Whole-genome sequencing of phenotypically distinct inflammatory breast cancers reveals similar genomic alterations to non-inflammatory breast cancers.” In: *Genome medicine* 13 (1 Apr. 2021), p. 70. ISSN: 1756-994X. DOI: 10.1186/s13073-021-00879-x.

[113] Jeffrey S Ross et al. “Comprehensive genomic profiling of inflammatory breast cancer cases reveals a high frequency of clinically relevant genomic alterations”. In: *Breast cancer research and treatment* 154 (2015), pp. 155–162. ISSN: 0167-6806.

[114] Christopher A Hamm et al. “Genomic and immunological tumor profiling identifies targetable pathways and extensive CD8+/PDL1+ immune infiltration in inflammatory breast cancer tumors”. In: *Molecular cancer therapeutics* 15 (7 2016), pp. 1746–1756. ISSN: 1535-7163.

[115] François Bertucci et al. “PDL1 expression in inflammatory breast cancer is frequent and predicts for the pathological response to chemotherapy”. In: *Oncotarget* 6 (15 2015), p. 13506.

[116] Paul H Levine et al. “Evaluation of lymphangiogenic factors, vascular endothelial growth factor D and E-cadherin in distinguishing inflammatory from locally advanced breast cancer”. In: *Clinical breast cancer* 12 (4 2012), pp. 232–239. ISSN: 1526-8209.

[117] Sudeshna Goswami and Neelam Sharma-Walia. “Osteoprotegerin secreted by inflammatory and invasive breast cancer cells induces aneuploidy, cell proliferation and angiogenesis”. In: *BMC cancer* 15 (1 2015), pp. 1–21. ISSN: 1471-2407.

[118] Komal Jhaveri et al. “Hyperactivated mTOR and JAK2/STAT3 pathways: molecular drivers and potential therapeutic targets of inflammatory and invasive ductal breast cancers after neoadjuvant chemotherapy”. In: *Clinical breast cancer* 16 (2 2016), pp. 113–122. ISSN: 1526-8209.

[119] Christoph Ziegenhain et al. “Comparative Analysis of Single-Cell RNA Sequencing Methods.” In: *Molecular cell* 65 (4 Feb. 2017), 631–643.e4. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2017.01.023.

[120] Efthymia Papalexi and Rahul Satija. “Single-cell RNA sequencing to explore immune cell heterogeneity”. In: *Nature Reviews Immunology* 18 (1 Jan. 2018), pp. 35–45. ISSN: 1474-1733. DOI: 10.1038/nri.2017.76.

[121] Rapolas Zilionis et al. “Single-cell barcoding and sequencing using droplet microfluidics”. In: *Nature protocols* 12 (1 2017), pp. 44–73. ISSN: 1754-2189.

[122] Marlon Stoeckius et al. “Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics”. In: *Genome biology* 19 (2018), pp. 1–12.

[123] Lukas Heumos et al. “Best practices for single-cell analysis across modalities.” In: *Nature reviews. Genetics* 24 (8 Aug. 2023), pp. 550–572. ISSN: 1471-0064. DOI: 10.1038/s41576-023-00586-w.

[124] Ilya Korsunsky et al. “Fast, sensitive and accurate integration of single-cell data with Harmony”. In: *Nature Methods* 16 (12 Dec. 2019), pp. 1289–1296. ISSN: 1548-7091. DOI: 10.1038/s41592-019-0619-0.

[125] Andrew Butler et al. “Integrating single-cell transcriptomic data across different conditions, technologies, and species”. In: *Nature biotechnology* 36 (5 2018), pp. 411–420. ISSN: 1087-0156.

[126] Emelie Berglund et al. “Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity”. In: *Nature communications* 9 (1 2018), p. 2419. ISSN: 2041-1723.

[127] Alma Andersson et al. “Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography”. In: *Communications biology* 3 (1 2020), p. 565. ISSN: 2399-3642.

[128] Reuben Moncada et al. “Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas”. In: *Nature Biotechnology* 38 (3 Mar. 2020), pp. 333–342. ISSN: 1087-0156. DOI: 10.1038/s41587-019-0392-8.

[129] Leeat Keren et al. “MIBI-TOF: A multiplexed imaging platform relates cellular phenotypes and tissue structure”. In: *Science Advances* 5 (10 Oct. 2019). ISSN: 2375-2548. DOI: 10.1126/sciadv.aax5851.

[130] *Saturn3*. Accessed: 2024-09-27. URL: https://www.saturn3.org.

[131] Matthew D Young and Sam Behjati. “SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data”. In: *GigaScience* 9 (12 Dec. 2020). ISSN: 2047-217X. DOI: 10.1093/gigascience/giaa151.

[132] Pierre-Luc Germain et al. “Doublet identification in single-cell sequencing data using scDblFinder.” In: *F1000Research* 10 (2021), p. 979. ISSN: 2046-1402. DOI: 10.12688/f1000research.73600.2.

[133] Sunny Z. Wu et al. “A single-cell and spatially resolved atlas of human breast cancers”. In: *Nature Genetics 2021 53:9* 53 (9 Sept. 2021), pp. 1334–1347. ISSN: 1546-1718. DOI: 10.1038/s41588-021-00911-1. URL: https://www.nature.com/articles/s41588-021-00911-1.

[134] Sandra Tietscher et al. “A comprehensive single-cell map of T cell exhaustion-associated immune environments in human breast cancer”. In: *Nature Communications* 14 (1 2023), p. 98. ISSN: 2041-1723. DOI: 10.1038/s41467-022-35238-w. URL: https://doi.org/10.1038/s41467-022-35238-w.

[135] Romain Lopez et al. “Deep generative modeling for single-cell transcriptomics”. In: *Nature Methods* 15 (12 Dec. 2018), pp. 1053–1058. ISSN: 1548-7091. DOI: 10.1038/s41592-018-0229-2.

[136] Luuk Harbers et al. “Somatic Copy Number Alterations in Human Cancers: An Analysis of Publicly Available Data From The Cancer Genome Atlas”. In: *Frontiers in Oncology* 11 (July 2021). ISSN: 2234943X. DOI: 10.3389/fonc.2021.700568.

[137] *inferCNV of the Trinity CTAT Project. https://github.com/broadinstitute/inferCNV*. Accessed: 2024-09-27.

[138] Lena Cords et al. “Cancer-associated fibroblast classification in single-cell and spatial proteomics data”. In: *Nature Communications* 14 (1 2023), p. 4294. ISSN: 2041-1723. DOI: 10.1038/s41467-023-39762-1. URL: https://doi.org/10.1038/s41467-023-39762-1.

[139] Karin E. de Visser and Johanna A. Joyce. “The evolving tumor microenvironment: From cancer initiation to metastatic outgrowth”. In: *Cancer Cell*

41 (3 Mar. 2023), pp. 374–403. ISSN: 15356108. DOI: 10.1016/j.ccell.2023.02.016.

[140] Daniel Dimitrov et al. “Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data”. In: *Nature Communications* 13 (1 June 2022), p. 3224. ISSN: 2041-1723. DOI: 10.1038/s41467-022-30755-0.

[141] Arthur Liberzon et al. “The Molecular Signatures Database Hallmark Gene Set Collection”. In: *Cell Systems* 1 (6 2015), pp. 417–425. ISSN: 2405-4712. DOI: https://doi.org/10.1016/j.cels.2015.12.004. URL: https://www.sciencedirect.com/science/article/pii/S2405471215002185.

[142] Arthur Liberzon et al. “Molecular signatures database (MSigDB) 3.0”. In: *Bioinformatics* 27 (12 June 2011), pp. 1739–1740. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btr260. URL: https://doi.org/10.1093/bioinformatics/btr260.

[143] Wenxuan Ren et al. “HMGB1 and Toll-like receptors: potential therapeutic targets in autoimmune diseases”. In: *Molecular Medicine* 29 (1 Sept. 2023), p. 117. ISSN: 1528-3658. DOI: 10.1186/s10020-023-00717-3.

[144] Kevin Mulder et al. “Cross-tissue single-cell landscape of human monocytes and macrophages in health and disease”. In: *Immunity* 54 (8 2021), 1883–1900.e5. ISSN: 1074-7613. DOI: https://doi.org/10.1016/j.immuni.2021.07.007. URL: https://www.sciencedirect.com/science/article/pii/S1074761321002934.

[145] Eva Hadadi et al. “Differential IL-1β secretion by monocyte subsets is regulated by Hsp27 through modulating mRNA stability”. In: *Scientific Reports* 6 (1 Dec. 2016), p. 39035. ISSN: 2045-2322. DOI: 10.1038/srep39035.

[146] Ting Liu et al. “NF-kappaB signaling in inflammation”. In: *Signal Transduction and Targeted Therapy* 2 (1 July 2017), p. 17023. ISSN: 2059-3635. DOI: 10.1038/sigtrans.2017.23.

[147] WanJun Chen and Peter ten Dijke. “Immunoregulation by members of the TGFβ superfamily”. In: *Nature Reviews Immunology* 16 (12 2016), pp. 723–740. ISSN: 1474-1741. DOI: 10.1038/nri.2016.112. URL: https://doi.org/10.1038/nri.2016.112.

[148] Ye Hu et al. “INHBA (+) cancer-associated fibroblasts generate an immunosuppressive tumor microenvironment in ovarian cancer”. In: *npj Precision Oncology* 8 (1 2024), p. 35. ISSN: 2397-768X.

[149] Aaron J. Wilk et al. “Comparative analysis of cell–cell communication at single-cell resolution”. In: *Nature Biotechnology* (May 2023). ISSN: 1087-0156. DOI: 10.1038/s41587-023-01782-z.

[150] Douglas Hanahan and Robert A Weinberg. “The Hallmarks of Cancer”. In: *Cell* 100 (1 Jan. 2000). doi: 10.1016/S0092-8674(00)81683-9, pp. 57–70. ISSN: 0092-8674. DOI: 10.1016/S0092-8674(00)81683-9. URL: https://doi.org/10.1016/S0092-8674(00)81683-9.

[151] Vincent Geldhof et al. “Single cell atlas identifies lipid-processing and immunomodulatory endothelial cells in healthy and malignant breast”. In: *Nature Communications* 13 (1 2022), p. 5511. ISSN: 2041-1723. DOI: 10.1038/s41467-022-33052-y. URL: https://doi.org/10.1038/s41467-022-33052-y.

[152] Pau Badia-i-Mompel et al. “decoupleR: ensemble of computational methods to infer biological activities from omics data”. In: *Bioinformatics Advances* 2 (1 Jan. 2022). ISSN: 2635-0041. DOI: 10.1093/bioadv/vbac016.

[153] Michael Schubert et al. “Perturbation-response genes reveal signaling footprints in cancer gene expression”. In: *Nature Communications* 9 (1 Jan. 2018), p. 20. ISSN: 2041-1723. DOI: 10.1038/s41467-017-02391-6.

[154] Shujing Wang et al. “Targeting M2-like tumor-associated macrophages is a potential therapeutic approach to overcome antitumor drug resistance”. In: *npj Precision Oncology* 8 (1 2024), p. 31. ISSN: 2397-768X. DOI: 10.

1038/s41698-024-00522-z. URL: https://doi.org/10.1038/s41698-024-00522-z.

[155] Katy Vandereyken et al. “Methods and applications for single-cell and spatial multi-omics.” In: *Nature reviews. Genetics* 24 (8 Aug. 2023), pp. 494–515. ISSN: 1471-0064. DOI: 10.1038/s41576-023-00580-2.

[156] Alev Baysoy et al. “The technological landscape and applications of single-cell multi-omics.” In: *Nature reviews. Molecular cell biology* 24 (10 Oct. 2023), pp. 695–713. ISSN: 1471-0080. DOI: 10.1038/s41580-023-00615-w.

[157] Xianwen Ren et al. “Insights Gained from Single-Cell Analysis of Immune Cells in the Tumor Microenvironment.” In: *Annual review of immunology* 39 (Apr. 2021), pp. 583–609. ISSN: 1545-3278. DOI: 10.1146/annurev-immunol-110519-071134.

[158] Sophia K Longo et al. “Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics.” In: *Nature reviews. Genetics* 22 (10 Oct. 2021), pp. 627–644. ISSN: 1471-0064. DOI: 10.1038/s41576-021-00370-8.

[159] Valentine Svensson, Eduardo da Veiga Beltrame, and Lior Pachter. “A curated database reveals trends in single-cell transcriptomics”. In: *Database* 2020 (Nov. 2020). ISSN: 1758-0463. DOI: 10.1093/database/baaa073.

[160] Dakai Yang et al. “Cancer-associated fibroblasts: from basic science to anticancer therapy”. In: *Experimental & Molecular Medicine* 55 (7 July 2023), pp. 1322–1332. ISSN: 2092-6413. DOI: 10.1038/s12276-023-01013-0.

[161] Xinglong Geng et al. “Cancer-Associated Fibroblast (CAF) Heterogeneity and Targeting Therapy of CAFs in Pancreatic Cancer”. In: *Frontiers in Cell and Developmental Biology* 9 (July 2021). ISSN: 2296-634X. DOI: 10.3389/fcell.2021.655152.

[162] Emanuela Brunetto et al. “The IL-1/IL-1 receptor axis and tumor cell released inflammasome adaptor ASC are key regulators of TSLP secretion by cancer associated fibroblasts in pancreatic cancer”. In: *Journal for ImmunoTherapy of Cancer* 7 (1 Dec. 2019), p. 45. ISSN: 2051-1426. DOI: 10.1186/s40425-019-0521-4.

[163] Bharti Garg et al. “NFkappaB in Pancreatic Stellate Cells Reduces Infiltration of Tumors by Cytotoxic T Cells and Killing of Cancer Cells, via Up-regulation of CXCL12”. In: *Gastroenterology* 155 (3 Sept. 2018), 880–891.e8. ISSN: 00165085. DOI: 10.1053/j.gastro.2018.05.051.

[164] Larry D. Sasich and Sana Rikabi Sukkari. “The US FDAs withdrawal of the breast cancer indication for Avastin (bevacizumab)”. In: *Saudi Pharmaceutical Journal* 20 (4 Oct. 2012), pp. 381–385. ISSN: 13190164. DOI: 10.1016/j.jsps.2011.12.001.

[165] EMA. *EMA Avzivi (bevacizumab) authorisation*. Accessed: 2024-09-27. URL: https://www.ema.europa.eu/en/medicines/human/EPAR/avzivi#overview.

[166] Zhanyan Gao et al. “T-cell exhaustion in immune-mediated inflammatory diseases: New implications for immunotherapy.” In: *Frontiers in immunology* 13 (2022), p. 977394. ISSN: 1664-3224. DOI: 10.3389/fimmu.2022.977394.

[167] Erin C Connolly, Julia Freimuth, and Rosemary J Akhurst. “Complexities of TGF-beta targeted cancer therapy.” In: *International journal of biological sciences* 8 (7 2012), pp. 964–78. ISSN: 1449-2288. DOI: 10.7150/ijbs.4564.

[168] Haynes Heaton et al. “Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes”. In: *Nature Methods* 17 (6 June 2020), pp. 615–620. ISSN: 1548-7091. DOI: 10.1038/s41592-020-0820-1.

[169] Yuhan Hao et al. “Integrated analysis of multimodal single-cell data”. In: *Cell* 184 (13 June 2021), 3573–3587.e29. ISSN: 00928674. DOI: 10.1016/j.cell.2021.04.048.

[170] Saket Choudhary and Rahul Satija. “Comparison and evaluation of statistical error models for scRNA-seq”. In: *Genome Biology* 23 (1 Jan. 2022), p. 27. ISSN: 1474-760X. DOI: 10.1186/s13059-021-02584-9.

[171] Enrique Blanco-Carmona. “Generating publication ready visualizations for Single Cell transcriptomics using SCpubr”. In: *bioRxiv* (Jan. 2022), p. 2022.02.28.482303. DOI: 10.1101/2022.02.28.482303. URL: http://biorxiv.org/content/early/2022/03/01/2022.02.28.482303.abstract.

[172] Robin Browaeys, Wouter Saelens, and Yvan Saeys. “NicheNet: modeling intercellular communication by linking ligands to target genes”. In: *Nature Methods* 17 (2 Feb. 2020), pp. 159–162. ISSN: 1548-7091. DOI: 10.1038/s41592-019-0667-5.

[173] Edward Y Chen et al. “Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool”. In: *BMC Bioinformatics* 14 (1 Dec. 2013), p. 128. ISSN: 1471-2105. DOI: 10.1186/1471-2105-14-128.

[174] Dénes Türei et al. “Integrated intra- and intercellular signaling knowledge for multicellular omics analysis”. In: *Molecular Systems Biology* 17 (3 Mar. 2021). ISSN: 1744-4292. DOI: 10.15252/msb.20209923.

Chapter 8

Contributions

The following people contributed to the work presented in this dissertation:
The thesis work was conducted under the formal supervision and in the division of Prof. Dr. Andreas Trumpp. He acquired funding, enabling the experiments.
Dr. Martin Sprick was my supervisor and mentor. Together, we conceived the project and he constantly provided invaluable feedback and guidance as it progressed.
Tasneem Cheytan contributed to the brainstorming and initial conception of the project, as well as the first experiments. Specifically, we identified suitable samples, pitched the project to our supervisors and performed the initial flow cytometric sort together.
Dr. med. Ewgenija Gutjahr provided invaluable support at the institute of pathology. She identified archived FFPE tissue with IBC, as well as matching controls. She created the tissue microarray used for spatial transcriptomics and prepared the special 10x Xenium slide with a section of that tissue block. Additionally, she performed H&E staining on a neighbouring section, and annotated the resulting scan.
Dr. Yunhee Jeong provided her computational and coding advice and expertise.
The Saturn3 consortium provided their kind feedback during meetings. In addition, Dr. Marc Zapatka designed and provided the 10x Xenium gene panel.
The DKFZ single-cell open lab, especially Dr. Jan-Philipp Mallm, Denise Keitel and Dr. Pooja Sant, contributed their expertise and assistance to the creation of single-cell libraries. They also performed the 10x Xenium measurement.
The DKFZ flow cytometry core facility provided assistance during all cell sorting procedures, especially Tobias Rubner, Florian Blum and Klaus Hexel.
The DKFZ Genomics and Proteomics Core Facility performed next-generation sequencing of all samples analysed as part of this thesis.
The DKFZ Omics and Data Core Facility (ODCF) provided the computing resources and technical support.
The HI-STEM and METICS groups contributed their diverse expertise when pro-

viding feedback during and after group meetings and seminars.
Dr. Khwab Sanghvi proofread this thesis and provided valuable feedback.

Chapter 9

Acknowledgements

Firstly, I would like to express me sincere gratitude to my supervisor, **Dr. Martin Sprick**. You gave me the opportunity to join your METICS group, and supported me from the very start. I really appreciate all the time we spent in your office discussing science, concerning my projects and beyond. Sometimes our meetings took several hours; still most of the time I walked out of them feeling more motivated to work and do some science than before. We also went together to my first ever real scientific meeting, the integrate-TN meeting in Marburg, and I have very fond memories of the two of us, discussing science for the entire train rides, and chatting over a drink at the hotel lobby. Funnily, they were so completely unprepared that someone actually ordered drinks, that they even struggled to prepare two G&Ts and had to run to fetch ice somewhere, much to our amusement.
But even through the rockier bits of the PhD journey, you were always supportive. When my first project to culture organoids turned out to be a failure, you said something along the lines of 'Don't worry, we will make something else work and in the end you will have a nice PhD thesis'. I was honestly not 100% sure I believed it at the time, but it turned out to be true, and here I am, 3.5 years and many meetings in your office later, having a PhD thesis I am actually proud of. You are a great scientist, and I am always so impressed by the breadth and depth of your knowledge. Thank you for making me the scientist I am today.

I would like to thank **Prof. Dr. Andreas Trumpp** for letting me join HI-STEM, and providing the environment making research at this level possible. Your ability to foster social interaction, but also an environment that produces science on this level, is truly admirable. I was also always impressed by your ability to keep track and be up to speed with everyone's projects, considering the sheer scale of HI-STEM. I always highly appreciated your feedback on my work, which went beyond just technicalities, instead making me see the bigger picture and providing true directional guidance. I also really appreciated the chances you gave me, by trusting me to present my work and represent HI-STEM at the biggest international cancer conferences there are. Thank you for making my PhD journey possible!

I would like to thank **Prof. Dr. Peter Lichter** for your role as my TAC member, second referee and defence committee member. Thank you for your many valuable contributions!

I want to thank **Prof. Dr. Wai Leong Tam** for your meaningful contributions as my TAC committee member and your continued support throughout the years.

I would like to sincerely thank **Prof. Dr. Ingird Lohmann** and **Prof. Dr. Stefan Wiemann** for agreeing to be members of my defence committee.

I am also thankful I got to meet so many amazing people and make so many lovely friends during this time in Heidelberg and at the dkfz.

Christina, you are such a great person, always so chill, and I just love spending time with you, it is such a joy. I am so glad we decided to go bouldering together during my first week at dkfz. Seeing you on Wednesdays just became such an important constant in my life, it just would not have been the same without. We summitted various mountains together (including the one in Erlangen), climbed rocks at the quarry, the Riesenstein, and basically everywhere we could find them, even sometimes in skiing boots.

Aga, you are the most altruistic, kind, supportive and loving person I know. With you around, it never takes long before I am crying tears of laughter. Even if life sometimes throws you off the carousel, I know you are the kind of person who wouldn't love but be there to help me back up.

Yunhee, right after we met I once told you I hope we can become good friends, but I had no idea how important you would become to me. Thank you so much for always being there to help me out when I was again playing at being a bioinformatician and I found myself between a rock (trying to install a package) and a hard place (odcf).

Khwab, you are so much fun to spend time with, discuss fashion, go on parties, and dance the night away. Your directness is unmatched, so thank you for never holding back with your opinion. And of course thanks for proofreading this thesis.

Anna, you are such a kind soul, and I love spending time with you. From 'repairing' other people's bikes, to going on conferences together, having dinner, drinks or going to the pool, it is always a great time.

Simay, you are so lovely and kind. It's so much fun spending time with you, attending your wedding in Turkiye was one of the best things of the PhD time.

Cansu, you are amazing. You are so kind and lovely and have such a nice humour. I truly admire your resilience to keep a smile on your face even in the toughest of times.

Flavio, even though you have been away for a while, it is always so nice to see you. I love your dry humour and kind nature. I also admire your botanical knowledge, which prevented us from eating a cake decorated with an amount of poisonous blossoms suitable to kill a grown horse.

Sven, you are a great friend and a one of a kind personality. I am immensely grateful for all the good times, parties and deep conversations we've had, and especially for our trip through Central Asia, which I will never forget.

Enrique, Nicola, Vivi, Yassin, Bea, Sonia, Giuglia, Jonas, Sandro, Daniela, I love our group dynamic, and appreciate each and every one of you so much. My PhD would not have been the same without you!

My PhD journey would of course also not have been the same without my fantastic lab mates.

Tim, you are awesome. I love sharing beds with you, even when there's a window to the bathroom. I am so glad we got to go on so many fun work trips together and became friends!

Shub, you are a phenomenon. I am very thankful and lucky to be your friend. Your talent for organising parties and entertain the crowds is unmatched. I am also especially curious and excited to see what our professional future has in store for us together!

Vera, you are one of a kind and have grown very dear to me. You are I think the only person with whom I can have tough humour banter, insulting one another all the time, but with all the love.

Jonas, Maija, Ines, Manu, Jen, Karo, Jenny, Tasneem, Frank, it is really the people who make the lab, and I couldn't wish for better friends around.

I would also like to thank my friends, whom I was fortunate enough to meet during previous chapters of my life, but who made me who I am and continued to support me throughout the years. **Yannik, Kilian, Lucio, Raul, Nico, Marc, Jonah and Lukas**, you are true friends and I love you.

Nicole, you accompanied me for the largest part of this PhD chapter. And even though our lives took different paths now, I still truly appreciate the time we had together and the support you gave me throughout the years.

Paulina, ich bin sehr glücklich, so eine wunderbare Person wie Dich gefunden und in meinem Leben zu haben. Du bist unglaublich toll und ich freue mich schon sehr auf unser nächstes gemeinsames Kapitel!

Ganz besonders möchte ich den wichtigsten Menschen in meinem Leben danken, meinen Eltern **Tini** und **Ingo** und meinem Bruder **Tom**. Ihr seid mein felsenfester Rückhalt, habt mich immer bedingungslos unterstützt und zu der Person gemacht, die ich bin. Ich bin so dankbar das Glück zu haben, Teil einer so liebenden Familie zu sein. Ich freue mich über jede Gelegenheit mit Euch zu sprechen und bin sehr froh, dass Ihr immer und zu jeder Zeit für mich da seid. Ihr seid die beste Familie, die man sich wünschen kann und ich liebe Euch von ganzem Herzen.