

Norwegian University
of Life Sciences

Master's Thesis 2022 30 ECTS

Faculty of Chemistry, Biotechnology and Food Science

Single-cell analysis of ER-positive breast cancer patients treated with letrozole and ribociclib

Karoline Rapp Vander-Elst

Master of science in Chemistry and Biotechnology

Single-Cell analysis of ER-positive Breast Cancer Treated with Letrozole and Ribociclib

Karoline Rapp Vander-Elst

Master of science in Chemistry and Biotechnology

Faculty of Chemistry, Biotechnology and Food Science

Norwegian University of Life Sciences

May 2022

Acknowledgements

This thesis is the final project of my Master of Science degree in Biotechnology and Molecular biology at the department of Chemistry, Biotechnology and Food science at the Norwegian University of Life Sciences. The project was performed in Xavier Tekpli's project group at Ullevål Oslo University Hospital at the department of Medical Genetics, which is a part of prof. Vessela Kristensen's group. The work was carried out under supervision from Xavier Tekpli and Harald Carlsen as internal supervisor and was performed from January 2022 to May 2022.

Firstly, I want to thank my external supervisor Xavier Tekpli for being an excellent mentor through the whole period, by providing scientific knowledge, good feedback, valuable advice and a lot of encouragement. Working on my master's thesis have given me a lot of joy and motivation to continue a scientific career. I also want to thank Marie Fongård for helping me in the lab, by giving good mentoring in excellent lab practice and motivation. Thank you Vessela Kristensen for providing the opportunity to work in your team.

Thank you, Jürgen Geisler, for your support to the group and for your courage in pushing forward clinical trials. There is also a special thanks to all the patients who are willing to participate in the clinical trial and contribute to this research. I want to thank my internal supervisor Harald Carlsen for his good advice and helpful insight in the writing process of the thesis.

Finally, I want to thank my peers and family for support and encouragement through my years of study. Thank you.

Karoline Rapp Vander-Elst

Karoline Rapp Vander-Elst

Abstract

Breast cancer is the most widespread cancer in the world, accounting for 25% of all female cancers. There is a high inter- and intra-tumor heterogeneity in breast cancer which makes it challenging to optimize the treatment for the individual patient. In recent years, the role of immune infiltration in tumor carcinogenesis and pathophysiology has been increasingly recognized. It has therefore become a priority to understand the interactions and cooperation between immune and cancer cells. Despite a thorough attempt to match treatment options with clinicopathological features such as histological classification, grade, stage, biomarkers, molecular subtypes, and intrinsic subtypes, many patients show resistance to treatment. One attempt to overcome treatment resistance is the emergence of combinatorial treatment, meaning treating patients with two drugs at the same time.

CDK4/6 inhibitors are anti-cancer drugs which prohibits cell growth and is shown to have promising results in combination with aromatase inhibitors for breast cancer patients with hormone receptor positive disease. This drug combination is not yet approved in Norway as standard neoadjuvant treatment. The NeoLetRib clinical trial facilitates the access to the combinations of aromatase and CDK4/6 inhibitor to patients. The study also gives the opportunity to investigate potential biomarkers for more personalized treatment, novel predictive biomarkers and assess how the tumor microenvironment changes during treatment.

Single cell analysis is the method we used to capture each cells transcriptome in the tumor microenvironment. We performed scRNA-seq of breast cancer biopsies from patients enrolled in the clinical trial NeoLetRib before the neoadjuvant treatment and after 21 days. This study shows that five cellular subtypes including Tregs, and four monocyte subtypes had a significant proportional change. These cell types have been associated with the promotion of a proinflammatory microenvironment and may be associated with tumor progression.

Table of contents

ACKNOWLEDGEMENTS	I
ABSTRACT	II
ABBREVIATIONS	V
1. INTRODUCTION	1
1.1 BREAST CANCER	1
1.1.1 <i>Epidemiology</i>	1
1.1.2 <i>Anatomy and development of the normal breast</i>	1
1.1.3 <i>Hallmarks of cancer</i>	2
1.1.4 <i>Breast cancer progression</i>	5
1.2 METHODS FOR BREAST CANCER CLASSIFICATION	6
1.2.1 <i>Histological classification</i>	7
1.2.2 <i>Grade</i>	7
1.2.3 <i>Biomarkers</i>	7
1.2.4 <i>Stage</i>	8
1.3 INTRINSIC SUBTYPES	9
1.4 MOLECULAR BREAST CANCER SUBTYPES	12
1.4.1 <i>HR-positive subgroups</i>	12
1.4.2 <i>HR-negative subgroups</i>	13
1.5 TREATMENT	14
1.5.1 <i>Surgery</i>	14
1.5.2 <i>Radiotherapy</i>	14
1.5.3 <i>Chemotherapy</i>	15
1.5.4 <i>Endocrine therapy</i>	16
1.5.5 <i>CDK4/6 inhibitors</i>	18
1.6 IMMUNE CELLS IN THE TME.....	21
1.6.1 <i>Lymphocytes</i>	22
1.6.2 <i>Dendritic cells</i>	24
1.6.3 <i>Macrophages</i>	24
1.6.4 <i>MDSC and neutrophils</i>	25
1.7 IMMUNOEDITING.....	26
1.8 TUMOR MICROENVIRONMENT	29
1.8.1 <i>Fibroblasts</i>	31
1.8.2 <i>Epithelial cells</i>	31
1.8.3 <i>Endothelial cells</i>	32
1.9 SINGLE CELL ANALYSES.....	33

1.9.1 <i>scRNA-seq pipeline</i>	34
2. THESIS AIMS	36
3. MATERIALS AND METHODS	37
3.1 PATIENT INFORMATION	37
3.2 TISSUE PREPARATION	37
3.3 ANALYSATION OF GENE EXPRESSION BY 10X GENOMICS	38
3.4 COMPUTATIONAL ANALYSIS	41
3.5 OWN CONTRIBUTIONS	42
4. RESULTS	43
4.1 CLUSTERING AND ANNOTATION OF ALL CELLS	43
4.2 NK AND T CELL CLUSTER ANNOTATION	45
4.3 MACROPHAGE CLUSTER ANNOTATION	48
4.4 FIBROBLAST CLUSTER ANNOTATION	51
4.5 TREATMENT RESPONSE	53
5 DISCUSSION AND FUTURE PERSPECTIVES	56
5.1 METHODOLOGICAL EVALUATION	56
5.2 CELL TYPES WITH SIGNIFICANT CHANGES IN PROPORTION BETWEEN THE TWO TIME POINTS	57
6 CONCLUSION	60
APPENDIX 1; SOLUTIONS	61
APPENDIX 2; ELBOWPLOT	62
APPENDIX 3; EXAMPLE CODES	63
EXAMPLE CODE FOR SCALING	63
EXAMPLE CODE FOR JACKSTRAW	64
EXAMPLE CODE FOR UMAP, HEATMAP AND BARPLOT	67
APPENDIX 4; DOTPLOT WITH ALL MARKER GENES	69
APPENDIX 5; UMAP AND HEATMAP FOR NK AND T CELLS	70
APPENDIX 6; HEATMAPS FOR FIBROBLAST ANNOTATION	72
APPENDIX 7; SUPPLIES	74
APPENDIX 8; REAGENTS	75
REFERENCES	76

Abbreviations

AI	Aromatase inhibitor	IDC	Invasive ductal carcinoma
ATP	Adenosine triphosphate	IFN	Interferons
BCR	B cell receptor	IL	Interleukin
CAF	Cancer associated fibroblast	ILC	Invasive lobular carcinoma
CDH	Epithelial cadherin	LCIS	Lobular carcinoma <i>in situ</i>
cDNA	Complementary deoxyribonucleic acid	MDSC	Myeloid derived suppressor cells
CSC	Cancer stem cell	MHC	Major histocompatibility complex
CSF	Colony-stimulating factors	mRNA	Messenger RNA
CTL	Cytotoxic T lymphocytes	NCCN	National comprehensive cancer network
DC	Dendritic cell	NGS	Next generation sequencing
DCIS	Ductal carcinoma <i>in situ</i>	NK	Natural killer cell
DNA	Deoxyribonucleic acid	PAM	Prediction analysis of microarray
ECM	Extracellular matrix	PC	Principal component
ER	Estrogen receptor	PCA	Principal component analysis
FACS	Flow-activated cell sorting	PCR	Polymerase chain reaction
FGF	Fibroblast growth factor	PDGF	Platelet derived growth factor
GEM	Gel beads-in-EMulsion	PR	Progesteron receptor
HDI	Human development index		
HER2	Human epidermal receptor 2		
HR	Hormone receptor		

QC	Quality control	TME	Tumor microenvironment
RB	Retinoblastoma protein	TNBC	Triple negative breast cancer
RNA	Ribonucleic acid	TNF	Tumor necrosis factor
ROR	Risk of recurrence	tRNA	Transfer ribonucleic acid
ROS	Reactive oxygen species	tSNE	t-distributed stochastic neighbour embedding
rRNA	Ribosomal ribonucleic acid	TSO	Template switch oligo
SERD	Selective estrogen receptor degraders	UMAP	Uniform manifold approximation and projection
SERM	Selective estrogen receptor modulators	UMI	Unique molecular identifier
TCR	T cell receptor	VEGF	Vascular endothelial growth factor
TDLU	Terminal ducts lobular units		
TGF	Transforming growth factor		
Th	T helper cell		

1. Introduction

1.1 Breast Cancer

1.1.1 Epidemiology

Breast cancer is the most prevalent cancer worldwide accounting for 25% of all female cancers, and the most fatal cancer for women in ninety-two countries (1, 2). Breast cancer can be detected by several ways. In Norway there is a biennial mammography screening for women in the age of 50-69 years old, which has led to increased survival rate related to breast cancer (3). In Norway survival rate over the past five years has increased from 91.5% to 92.1% in addition to an increase of 3.7% in incidence rate compared to the last five-year period (2011-2015). The increase in incidence rate is shown to be limited to women of the age group 60-79 years and may be related to better diagnostic methods (4, 5). 3424 new cases of breast cancer were enumerated in Norway in 2020, which reflects 9,6% of new cancers registered that year. This is a lower number than the previous years and it is assumed that this is due to the national screening program being paused for some months because of the corona virus pandemic (4).

1.1.2 Anatomy and development of the normal breast

The female breast is a complex organ composed of ducts, lobes, fibrous connective tissue, ligaments, and adipose tissue that give support to the breast, this is shown in figure 1a. Between 15-20 clusters of lobes can be found in a breast, each develops terminal duct lobular units (TDLU) during puberty, this is illustrated in figure 1b (6).

The lobes, ducts and lobules consist of an inner layer of luminal epithelial cells, these cells synthesise milk in the lobules. Basal myoepithelial cells surround the luminal cells, have contractive properties, and facilitate milk secretion into the lumen. The basement membrane separates the myoepithelial cells from the stroma as seen in figure 1c (6). During menopause the levels of hormones such as estrogen will change resulting in an involution of the glandular lobes and ducts (7).

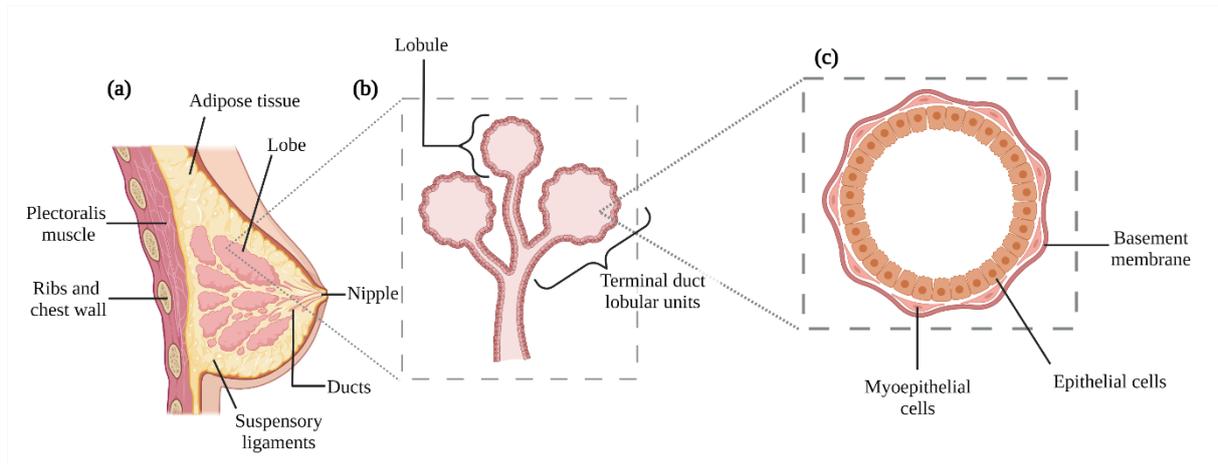


Figure 1. The female breast. (a) Anatomy of a normal female breast. (b) Lobes with TDLU and lobules. (c) The inner layer with luminal epithelial cells is surrounded by myoepithelial cells that are connected to the basement membrane (8).

1.1.3 Hallmarks of cancer

Breast cancer is a genetic disease as both somatic and germline mutations can be the cause of the tumor development and initiation. Despite an increased understanding of the disease the past years, there are still many unanswered questions such as what initiates the primary tumor progression? Which changes occur for the carcinogenic process to start? How does the cancer respond to treatment? A major challenge is that breast cancers are highly inter- and intra-heterogenous and form complex tumor microenvironments (TME) shaped by genetic aberrations, environmental influences, cellular biological context and the characteristics of the patient (9).

The hallmarks of cancer were first presented in year 2000 by Hanahan and Weinberg (10), and have been revised in 2011 (11) and 2022 (12). Cancer is an evolutionary disease as it often requires a plethora of gained characteristics to proliferate and avoid the host immune system. The hallmarks have been extensively described and depicts how a normal cell evolve firstly into a neoplasm, then becoming tumorigenic and eventually malignant. There are now fourteen hallmarks that describe carcinogenesis and tumor progression, which is shown in figure 2.

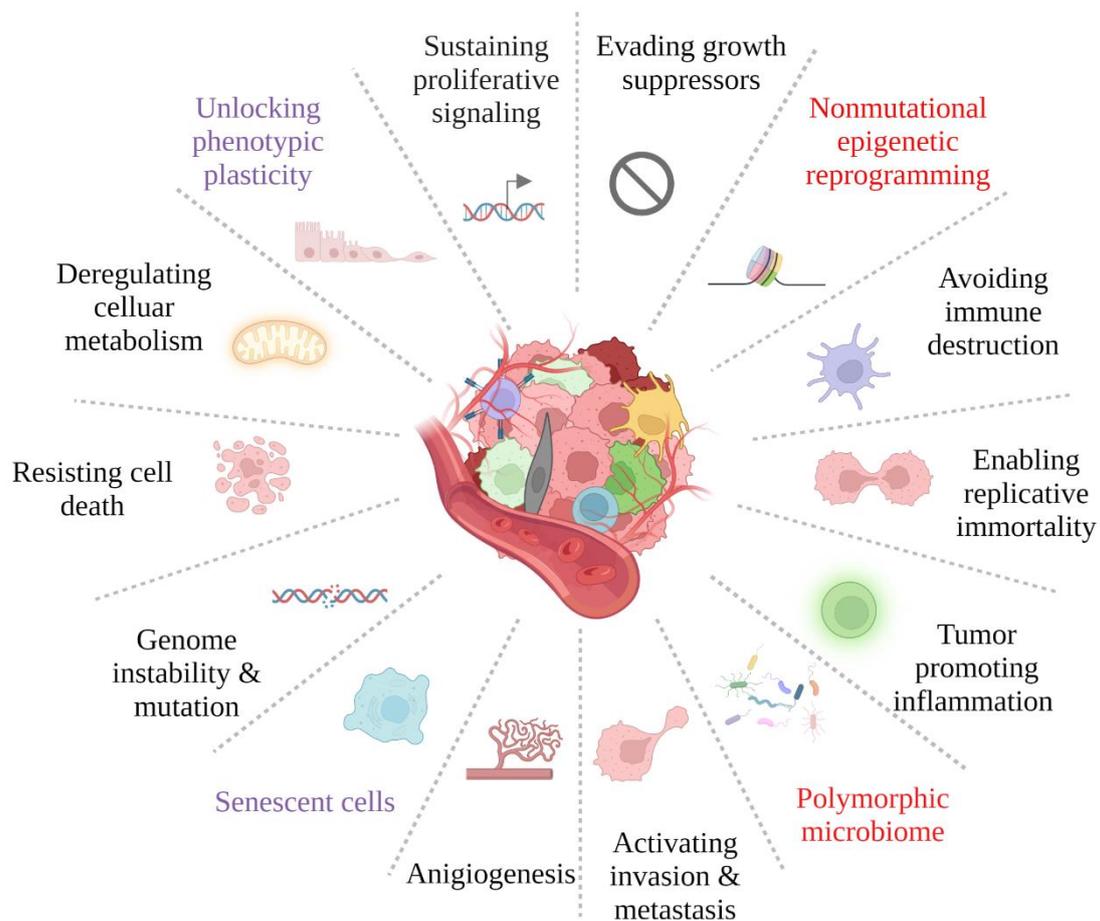


Figure 2, Illustration of the fourteen hallmarks of cancer, the newly proposed enabling characteristics established in 2021 are “nonmutational epigenetic reprogramming” and “polymorphic microbiome”, while the newly presented emerging hallmarks are “unlocking phenotypic plasticity” and “senescent cells” (8).

Sustaining proliferative signalling and **evading growth suppressors** are two hallmarks that are the major driving forces of uncontrolled cell growth, and means in short that a cancer cell is not dependent on external signals in order to divide, and also avoid signals to terminate cell division. A cancer cell can accomplish this in various ways, including producing its growth factors and overexpress receptors which will enhance proliferation. Examples are vascular endothelial growth factor (VEGF), fibroblast growth factor (FGF) and CXCL8 in a combination with the receptor for interleukin (IL) 8 and inactivate tumor suppressor genes such as the retinoblastoma protein (RB1), p53 and transforming growth factor beta (TGF- β). The immune system tolerates the normal cells of the body; however, tumor cells are genetically different from normal cells of the body due to the acquisition of somatic mutations

and can then become a target of the immune system. **Avoiding immune destruction** is therefore a crucial characteristic for cancer cells. How cancer cells avoid the host's immune system is not clearly understood, but it can include expression of proteins that modulates the immune system, acquisition of traits that makes them undetectable and exhaustion of the immune cells. **Enabling replicative immortality** gives the cancer cells the ability to divide without any intrinsic limit. This is in contrast to normal cells with a finite lifetime due to the telomeres being shortened by each cell division eventually leading to senescence. In cancer cells high expression of the enzyme telomerase can keep telomere length above a critical threshold which plays a key role in the proliferation of cancer cells. Cancerous tissues are often referred to as non-healing wounds which sustains inflammation referred to as **tumor-promoting inflammation**. This environment will lead to the recruitment of immune cells such as lymphocytes and macrophages to the TME. Macrophages can release growth factors and cytokines that in turn will enhance cancer cell growth. There are different stages for cancer development, benign lesions are non-cancerous tumors, meaning that they have low proliferation and aggressiveness. Malignant tumors on the other hand are cancerous and have the potential to spread across the body in a process called metastasis. **Activating invasion and metastasis** by cancer cells can be achieved by the breaking of cell-cell adhesion, by suppressing E-cadherin as an example, or degradation of the extracellular matrix (ECM) that usually would keep the neoplasm in place. As cancer cells grow and recruit other cells to the TME the need for oxygen, glucose and nutrients will increase, there will also be a need of removing waste products such as carbon dioxide. One solution to this is to form new blood vessels, through **angiogenesis** which is another hallmark of cancerous cells. **Genome instability and mutation** will participate in the acceleration of the tumor development, this can include one or several genomic changes such as base substitutions, genome rearrangements, insertions and deletions, and chromosome copy-number alterations. Cancer cells can **deregulate cellular energetics** in the TME. One example is that cancer cells will undergo anaerobic respiration whether it has access to oxygen or not, and always transform glucose into lactate. One of the novel hallmarks is **unlocking phenotypic plasticity**, the disturbance of the cellular differentiation can occur in several ways such as dedifferentiation from mature to progenitor cells, trans differentiation to other cell lineages or by blocked differentiation from progenitor cells to the mature state. This can trap the cell in a more proliferative state than the original mature cell state. **Senescent cells** have long been recognized as a protective mechanism against cancer cell progression, but it is observed that senescent cells in the TME can modulate the TME and enable other hallmark capabilities.

One example is therapy induced senescent cancer endothelial cells in the breast that can enhance metastasis and proliferation (12-14) .

There are also two enabling characteristics that can facilitate the development of hallmark abilities, firstly **nonmutational epigenetic reprogramming** which can be induced by many co-opted and corrupted mechanisms which are independent of mutations and genome instability. Secondly **polymorphic microbiomes** are considered to have both a negative and positive effect on several hallmark capabilities in some tumor types. The microbiome is often associated with organs directly in contact with the environment such as the gut, skin, lung, vaginal and oral cancers, but a study of 1526 tumors from breast, brain, bone, and lung among others showed a distinct microbiome mainly inside the cancer cells. It was especially observed a diverse and rich microbiome in breast cancers (15).

1.1.4 Breast cancer progression

Breast cancers are mainly carcinomas, this is cancers that that originates from epithelial cells. Breast cancer carcinogenesis is a comprehensive process which includes many of the hallmarks of cancer such as genetic and epigenetic alterations, and modulations of the TME (12, 16). Invasive breast cancer is the endpoint of the process of carcinogenesis which begins in the TDLUs. Many mechanisms in this process are unknown, one hypothesis is that the cancer cells originate from the TDLUs by progressing through benign breast disease, followed by upregulation of proliferation and finally atypical hyperplasia. Atypical hyperplasia is an accumulation and piling of the overproduced cells and can be either located in the ducts (atypical ductal hyperplasia) or in the lobules (atypical lobular hyperplasia), after this the development is believed to evolve to carcinoma *in situ* which includes ductal carcinoma *in situ* (DCIS) and lobular carcinoma *in situ* (LCIS). The last steps will then be acquisition of invasive and metastatic properties (17-19).

1.2 Methods for breast cancer classification

Breast cancers are divided into different subtypes to provide the best treatment, diagnosis, and prognosis. The subtypes are assigned based on histology, morphology, proteomics, and genomics. As breast cancers are highly heterogeneous these subtypes are of great importance as they attempt to describe the proportions of cell type variation in the TME. This variation is also the cause of the cancers to have different clinical behaviour and biological traits (19). Some of the different breast cancer classifications are shown in figure 3, this also illustrates that the classification is complex and requires many types of analysis to provide the best treatment and prognosis for the patient.

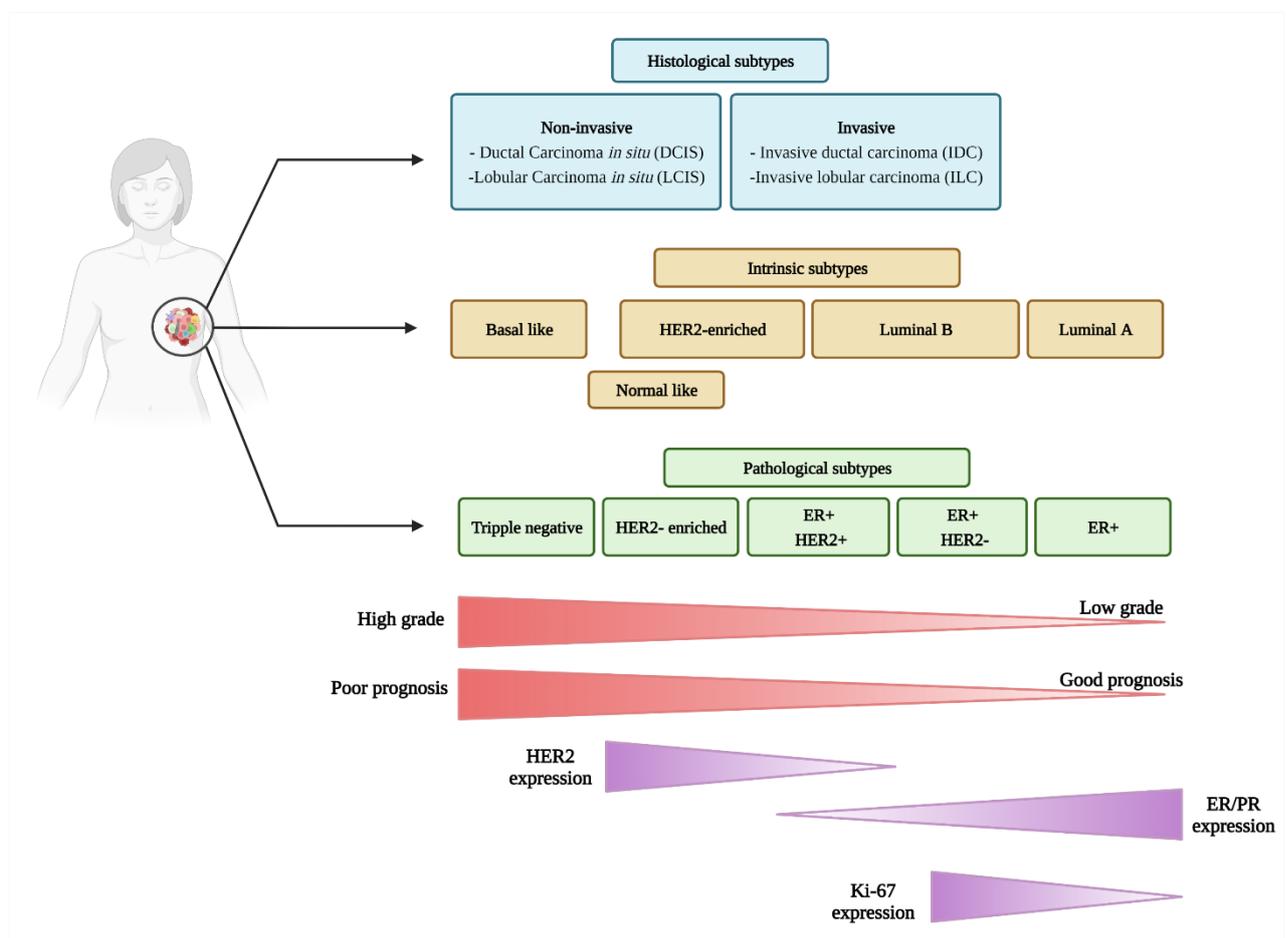


Figure 3. An illustration of breast cancer classifications and subtypes, showing how much a breast cancer can differ in receptor expression, grade, prognosis and location. Breast cancers are divided into histological subtypes. This includes the non-invasive carcinomas DCIS and LCIS, and the invasive carcinomas IDC and ILC. The intrinsic subtypes categorizes the cancers into the basal like, normal like, HER2-enriched, luminal B and luminal A subtypes. The molecular subtypes are based on the pathological profile of the cancers and divides them into the triple-negative, HER2-enriched, Luminal B and Luminal A subtypes. Breast cancers are also categorized by grade which is associated with the aggressiveness of the tumor, HER2-expression, and hormone (ER) expression. Basal like and triple negative disease are associated with a poor prognosis while luminal A have the best prognosis (8).

1.2.1 Histological classification

The histological classification of breast cancer includes two main categories the non-invasive and the invasive carcinomas, consisting of subgroups based on their location in the breast. The non-invasive carcinomas include DCIS, LCIS and Paget's disease, these have not crossed the basement membrane. When cancer cells cross the basement membrane, they are considered invasive, the subtypes of the invasive breast cancers are invasive ductal carcinoma (IDC) which includes tubular, mucinous, medullary, and inflammatory carcinoma, the second group are the invasive lobular carcinoma (ILC) that are often characterized by mutations of epithelial cadherin (CDH1) (18). Ductal carcinoma is the most common type of breast cancer that constitutes 75-80% of all breast cancers (19).

1.2.2 Grade

The invasive carcinomas are also categorized by grade. The grade of the cancer is determined based on three histological attributes: mitotic activity, degree of nuclear atypia and tubule formation. The grade ranges from 1-3, where high grade tumors are more aggressive. The three histological features are given a value from 1-3 and are summarized to give the Bloom-Richardson score. Grade 1 include the tumors with a score between 3-5 and are well differentiated with great homology of the normal breast TDLU, they also have a low mitotic count and tubule formation. Grade 2 breast cancers include tumors with a score between 6-7 and are moderately differentiated. Lastly group 3 which include the breast cancers with a score of 8-9 are poorly differentiated with a high proportion of cellular pleomorphism, high degree of mitoses and no or little visible tubule formation (18, 20).

1.2.3 Biomarkers

To classify neoplasms further molecular analyses are performed to investigate which receptors the tumor overexpresses. Overexpression of the estrogen receptors (ER), progesterone receptors (PR) or human epidermal receptor 2 (HER2) characterize pathological subtypes of breast cancers and are detected by immunostaining of the three proteins. This information is of great importance when it comes to treatment of the patient as hormone receptor (HR)-negative tumors are not or less dependent to hormone levels and tends to grow faster and be more aggressive than HR-positive tumors. The Ki-67 protein has also become an

important proliferation biomarker to the pathological classification. There are also, less systematically used biomarkers such as the genomic biomarkers BRCA1, BRCA2 and PIK3CA, and immunomarkers such as PD-L1 and infiltrating lymphocytes (2).

1.2.4 Stage

Breast cancer stages describe how proliferative the cancer is. It is a scale that ranges from I-IV, where the lowest number indicates lesser spread of the cancer, it is followed by a letter and an early letter indicates lower stage. The scale system takes several factors into consideration the first three are called the TNM system, “T” stands for tumor size, “N” stands for spread to adjacent lymph nodes and “M” is for metastasis to other organs in the body. Other factors taken into consideration for the staging are ER, PR and HER2 receptor statuses and grade (21).

1.3 Intrinsic subtypes

Breast cancers are very heterogenous and high throughput methods such as transcriptome measurements have shown to be of prognostic and predictive value in breast cancer. These methods are only now starting to be routinely performed in some western countries, as they are still costly (22). Technological advancement has allowed to investigate gene expression patterns and by that improve the molecular taxonomy of breast cancers. A study performed by Perou *et al.* applied a cDNA microarray classification technique where they characterised distinct gene expression patterns of the molecular breast cancer subtypes (22). An unsupervised clustering heatmap with 1753 genes involved in breast cancer progression was obtained, which was further reduced to a subset of 496 intrinsic genes with a high variation in expression between cancerous tissues. Clustering based on the expression of these 496 genes resulted in two main branches one with the ER-positive and one with the ER-negative breast cancers (22).

A follow up study by Sørli *et al.* used 456 cDNA clones as the intrinsic set of genes and applied a hierarchal clustering. The cancers were divided into two branches where one group was ER-negative and the other ER-positive breast cancers as in the previously described study. In addition the researcher further managed to observe distinct clusters of 6 molecular subgroups, luminal A, luminal B, luminal C, normal breast-like, HER2-positive and basal like, as shown in figure 4 (23).

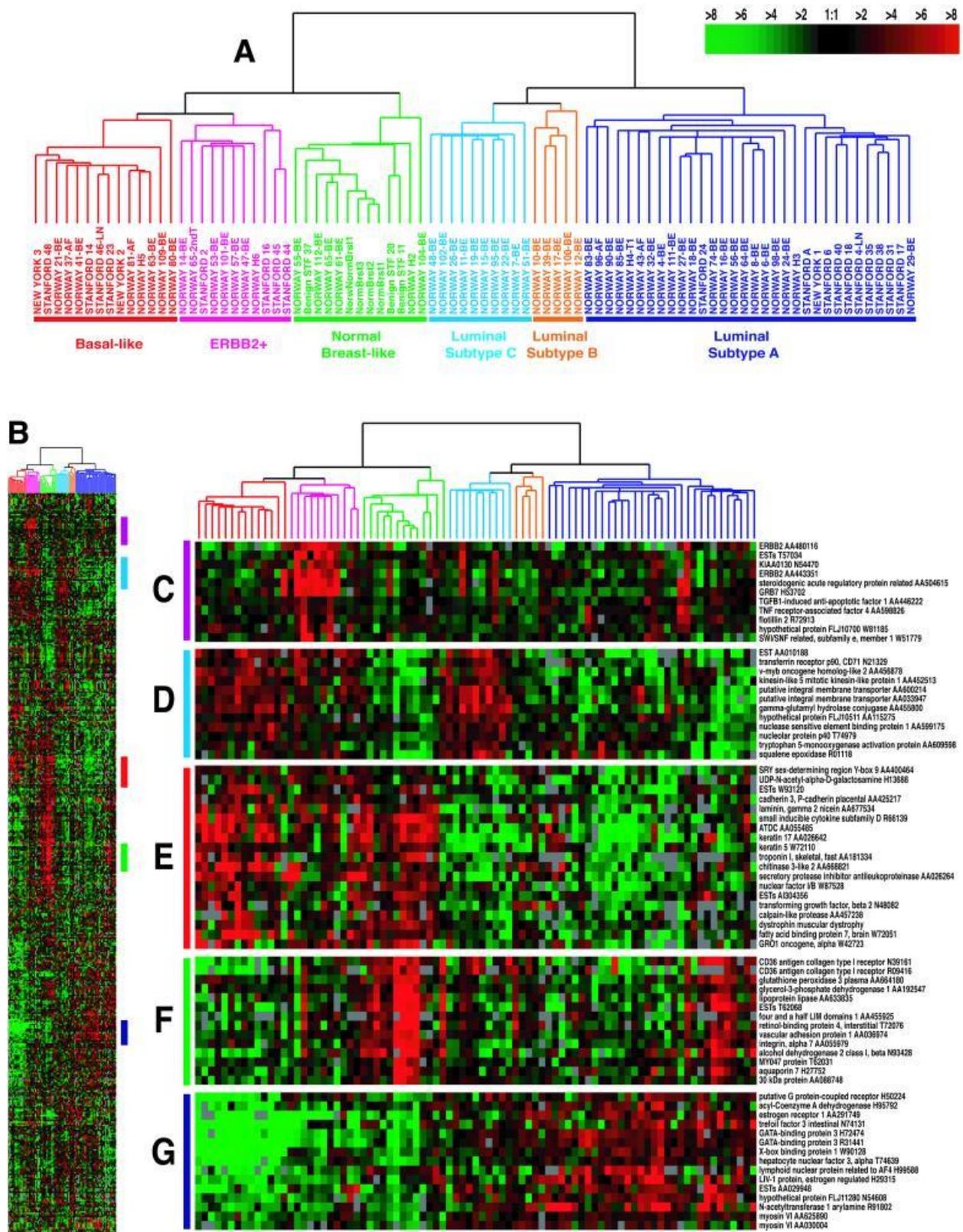


Figure 4. The gene expression pattern by hierarchical clustering of 476 cDNA intrinsic genes. (A) The six molecular subtypes based on difference in gene expression. The luminal A are dark blue, luminal B are yellow, luminal C are light blue, normal breast-like is green, HER2 are pink and TNBC is red. (B) illustrates a scaled down cluster diagram. The bars with different colours on the right of the diagram represent inserts presented in C-G. (C) HER2 amplicon cluster. (D) Novel unknown cluster. Basal epithelial cell-enriched cluster. (F) Normal breast-like cluster. (G) Luminal epithelial gene cluster with ER. The green colour in the diagrams represent down-regulations and the red up-regulation (23).

Parker *et al.* suggested that the above-mentioned molecular subtypes could be determined by using a clustering method called Prediction Analysis of Microarray (PAM) and selected 50 genes of the beforementioned 500 intrinsic genes for this analysis. This method is now known as PAM50 and has in the recent years been introduced in clinical practice. This improvement increased the breast cancer prognosis standard, gave a better prediction, and provided suggestions for treatment options. The PAM50 genes and their association to molecular subtype is illustrated in figure 5 (24). Studies support the application of PAM50 as it shows promising results in improving classification of breast cancers and treatment decisions (19, 25).

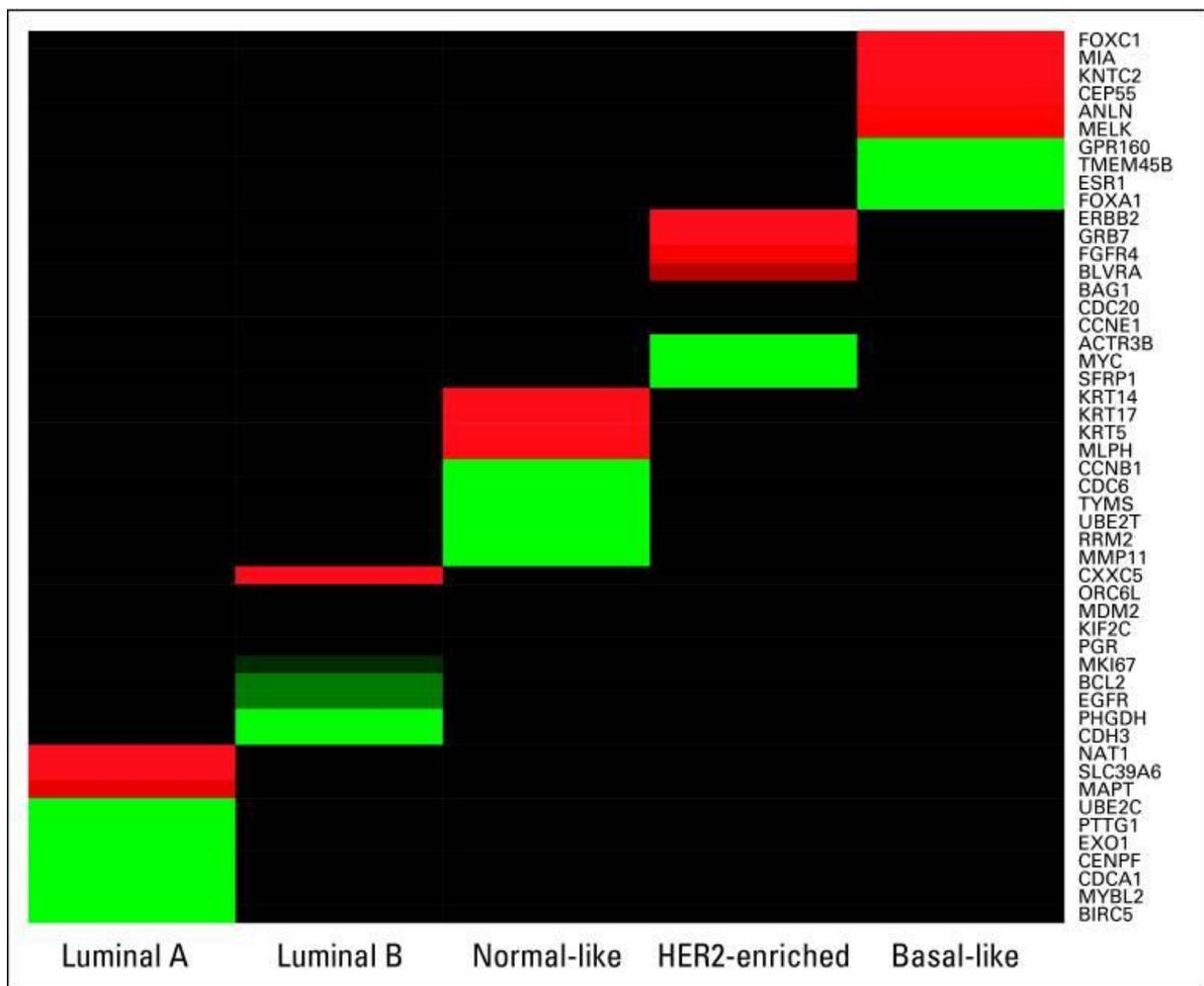


Figure 5. A heatmap of Classification by Nearest Centroids (ClANC). The ClANC algorithm selected 10 genes per class for a total of 50 genes. The genes for the classes are shown as red/green according to their expression in the give class. Black indicates that gene was not selected for the given class (24).

1.4 Molecular breast cancer subtypes

In the clinic immunohistochemical staining are often used to identify biomarkers with strong association with patient prognosis, survival, and treatment response. The routine diagnostic staining of breast cancer usually include ER, PR, the proliferation protein Ki-67, and some cytokeratin's (19).

ER and PR are both transcriptional regulators and gives an indication of the breast cancer being sensitive to endocrine treatment. Cancers with ER and/or PR expression are associated with a good prognosis, while cancers that are ER-negative are associated with a poor prognosis. Another important factor for breast cancer development and progression are the growth factor receptors and ligands that promote proliferation. One of these are HER2 as described earlier. The expression of HER2 is associated with aggressive tumors and poor prognosis. The breast cancer subtypes can therefore be divided into the HR-positive and HR-negative groups and will be elaborated further in the following section (19).

1.4.1 HR-positive subgroups

The HR-positive breast cancers are ER and/or PR positive. They are usually luminal subtypes. They are the most common type of breast cancer and are again divided into two main subgroups. Luminal A is the most prevalent molecular subtype and accounts for 30-70% of all breast cancers. This subtype is usually HER2-negative, ER-positive, PR-positive and has a Ki-67 expression <14%, it has a low grade, low risk of recurrence (ROR) and high survival rate. Luminal A tumors are associated with the best prognosis and commonly respond well to endocrine treatment (2, 26).

Luminal B, the other HR-positive subgroup accounts for approximately 20% of breast cancers. Tumors with this subtype are ER-positive, PR-negative and have a Ki-67 expression >14%, it is seen that this subgroup can have both HER2-positive and HER2-negative expression. The subgroup is associated with a poorer prognosis than luminal A and has a higher proliferation rate as indicated by the higher expression of Ki-67 (2, 26, 27).

The ER status only provide an indication of response to endocrine therapy, it is observed that only ~50% of patients with ER-positive breast cancers respond to the distributed endocrine treatment (26).

1.4.2 HR-negative subgroups

The HR-negative subgroups are associated with poor prognosis and are divided into two main subtypes. Firstly, the HER2-enriched subgroup which accounts for 10-15% of all breast cancers. This subgroup is ER-negative, PR-negative, HER2-positive and has a poorer prognosis than both HR-positive subgroups as well as being a higher-grade cancer. HER2-enriched breast cancer is highly proliferative, 10-15% of these cancers are due to germline mutations in the BRCA1 or BRCA2 genes (2, 19, 26, 27).

The last subgroup is the triple negative breast cancer (TNBC) or basal group which accounts for 15-20% of all breast cancers. With HER2-negative, ER-negative, and PR-negative receptor phenotype, this subtype has the worst prognosis and the highest risk of metastasis. This group is associated with germline mutations in BRCA1 or BRCA2 in 15-20% of the cases and is a high-grade cancer (2, 26, 27).

Other categories of pathological and clinical factors as patients age, tumor size, lymph node status, and lymph vascular invasion are also determined to provide appropriate therapy and prognosis for the patient (2, 19).

1.5 Treatment

The treatment options for breast cancer patients are based on the histological grade, stage, pathological subtype and molecular subtype. Neoadjuvant therapy is a systemic treatment given before local treatment to facilitate surgery and radiation. It is administered through the bloodstream (2, 28). This is the conventional treatment for inoperable and locally advanced breast cancers (29). Adjuvant treatment is given after local surgical or radiational intervention (2). It is observed that some ER-positive breast cancer patients have a high percentage of recurrence, one-half of these recurrences are observed between 10-32 years after primary diagnosis (30). The aim of adjuvant therapy is to provide a combination of treatments that will give the patient the best ease of disease while reducing the ROR and toxicity (29). A better understanding of the treatment response and the TME in breast cancer can provide a better treatment option to reduce the ROR and side effects of treatment. In this chapter a description of the most relevant treatment options for the patient group studied in the thesis are described in more details.

1.5.1 Surgery

The aim of surgical procedures of breast cancer is to attain control over the cancer and get further information about the extent of the cancer spread. There are two main types of surgery for breast cancer patients one is a mastectomy, which is a total removal of the affected breast. The second approach is a lumpectomy, which is a breast conserving surgery and only the primary cancer is removed. During surgery an axillary lymph node dissection can provide information whether the cancer has the potential to spread to the adjacent lymph nodes or metastasis. This is important information regarding further treatment of the breast cancer patient (2, 28).

1.5.2 Radiotherapy

Radiotherapy is high-energy radiance given locally to breast cancer patients. It can be given as a neoadjuvant therapy to reduce the cancer size, as an adjuvant treatment for patients with a breast cancer type associated with a high ROR, or as an attempt to remove undetected cancer cells. The treatment can be applied to the whole breast or limited to only one part of the breast; it can also be directed to the regional lymph nodes (28). Despite radiotherapy being

one of the most efficient cancer treatment some patients can be subject to local recurrence. Tumor size and advanced tumor stage are prompting these relapses even after radiotherapy (31). Radiotherapy can be combined with targeted treatments such as endocrine therapy or anti-HER2 therapy (28).

There are side effects induced by treatment with radiotherapy such as cardiotoxicity, it is therefore of great importance to minimize the exposure of the normal surrounding tissue. Some measures to reduce damage to the heart is the beam angles, partial breast irradiation, intensity modulated radiotherapy and breath-hold techniques (28, 32).

Other factors that affect the tumor outcome after radiotherapy are tumor infiltrating cells which is a part of the formation of the TME (31), immune cells such as dendritic cells, tumor associated macrophages, myeloid-derived suppressor cells and some T-cells are somewhat radioresistant (33).

1.5.3 Chemotherapy

Chemotherapy is a systemic therapy given to breast cancer patients either as a neoadjuvant treatment to shrink the cancer or adjuvantly to patients with high ROR or spread to the lymph nodes to eliminate remaining cancer cells (2, 34). It is mainly administered as a treatment for metastatic and locally advanced breast cancers (28). Chemotherapeutic agents evoke apoptosis in highly proliferating cells by attaching to microtubules, this prevents proteins from assembling properly and inhibits the cell cycle from proceeding (28). The treatment may also break the DNA strands and can promote intercalation of the DNA, which is interference of the DNA double helix leading to distortion of transcription, repair, and replication. The treatment is often given over a period of time to allow the normal cells to repopulate the tumor area (34).

Clinical trials have reported that most aggressive ER-positive breast cancers have a low response to chemotherapy alone and are recommended to be treated with a combination of chemotherapy and endocrine therapy. Chemotherapies also have a lot of side effects such as muscle pain, fatigue, second cancers and sterility (28).

1.5.4 Endocrine therapy

Endocrine or hormonal therapy is mainly offered to patients with ER-positive breast cancer both locally advanced and metastatic (28). The aim of endocrine therapy is to inhibit the action of estrogen as a growth promoting factor of cancer cells. This disruption can be obtained in several ways, one way is to inhibit the synthesis of estrogen done by aromatase inhibitors (AI). Another option is by targeting estrogen receptors using selective estrogen receptor modulators (SERMs) and/or selective estrogen receptor degraders (SERDs). The mechanism of endocrine therapy is illustrated in figure 6.

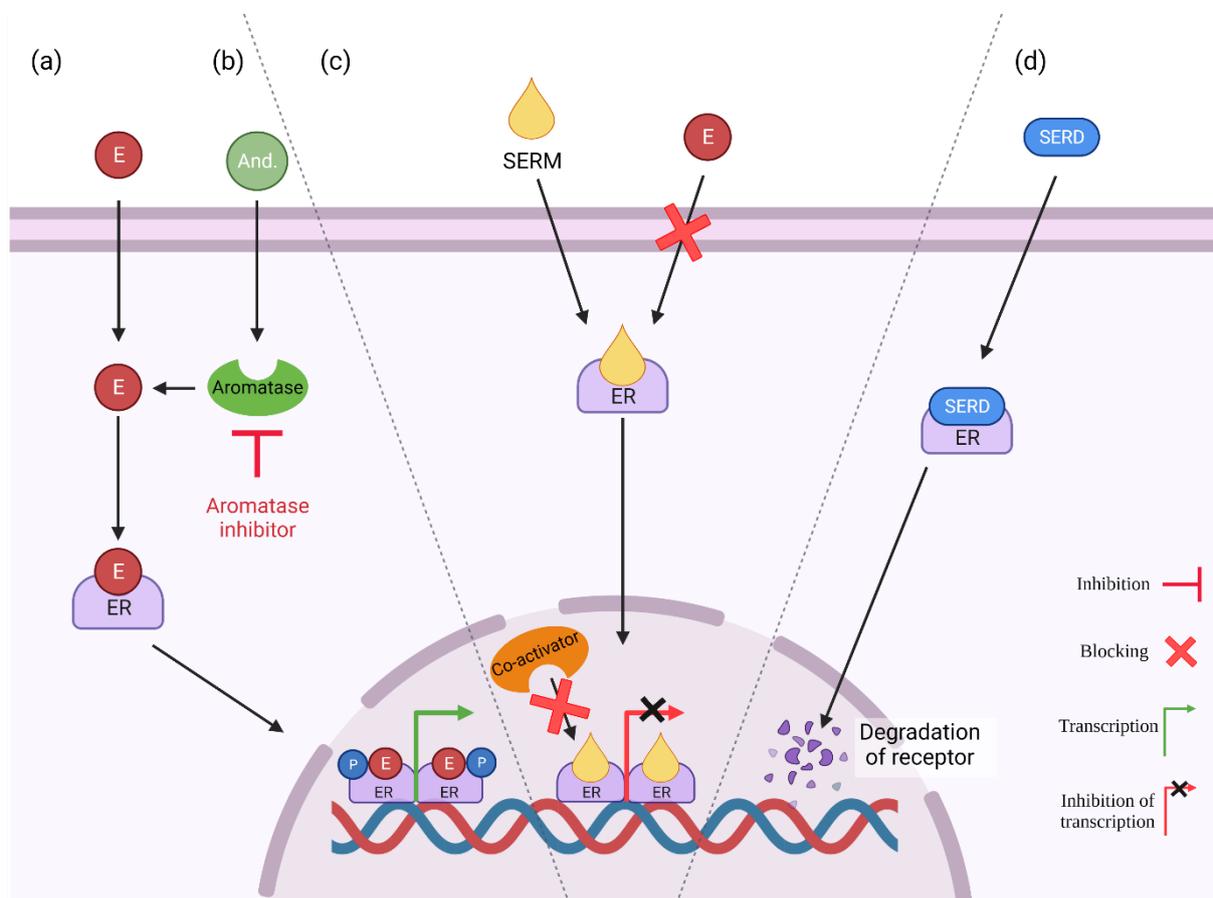


Figure 6. Mechanisms of action for endocrine therapy. (a) Estrogen are small non-polar soluble lipids and can pass the cell membrane by diffusion. When they enter the cell, they attach to the estrogen receptors and cause a dimerization of the ERs. (b) Androgens are another biomolecule that migrates into the cell and are transformed to estrogen by the enzyme aromatase. Aromatase inhibitors will therefore inhibit the production of estrogen in the cells and reduce the overall concentration. This is again limiting the proliferation of the cell. (c) SERMs migrate into the cell and block the ERs from connecting to estrogen. This inhibits the co-activator from initiating transcription in the cell nucleus. (d) SERDs binds to ERs in the cell and induces degradation of ER. Thereby the necessary dimerization will be inhibited. Created with BioRender.com

Aromatase inhibitors (AI)

Aromatase is an enzyme that synthesises estrogen and androgens. Breast cancer proliferation have a positive correlation with high estrogen levels and are therefore associated with a high aromatase expression. AI will block the function of the aromatase enzyme and the production of estrogen as illustrated in figure 6b. There are two classes of AIs which are the non-steroidal and the steroidal. Non-steroidal AIs compete with androgens to bind to the substrate-binding site of the aromatase and prohibits the activation of the aromatase enzyme, letrozole and anastrozole are examples of this therapy. Steroidal AIs includes exemestane and is very similar to androstenedione, which is the aromatase substrate. The binding of the AI will lead to inactivation of the aromatase substrate-binding site.

Postmenopausal women with ER-positive breast cancer have a good response to AI, and many studies show that AI are more efficient than the widely used SERM tamoxifen and other endocrine agents (35, 36). Since AI are reducing the amount of estrogen in the body it can however lead to an increased risk of osteoporosis, cardiovascular diseases, and fatigue. Some breast cancers are resistant to AIs, this can be acquired during the treatment period, or could be present before treatment (36). The resistance can come from an upregulation of the PI3K pathway, ESR1 mutations, epigenetic modifications, or upregulation of the Cyclin dependent kinase 4 and 6 (CDK4/6) pathways as examples. Studies show that the combination of AI and targeted therapy such as CDK4/6 inhibition has better outcomes than treatment with AIs alone (28, 37, 38). A study by Dunbier *et.al* illustrated for the first time a correlation between the amount of infiltrating immune cells and poorer response to AI treatment, as infiltrating immune cells can have both pro- and antiproliferative properties it would be of great interest to find predictive markers regarding AI treatment response (39).

The resistance to AIs can be elucidated better by getting information about the subtypes of cells present in the different breast cancers, as some subgroups can be dependent on other growth signals than estrogen and can be used as predictive markers (36). To increase the understanding of resistance to therapies, it is also important to analyse biopsies of the tumors at different timepoints such as before, during and after treatment too see how the TME evolves under treatment pressure.

SERM & SERD

SERMs are modulators that compete with estrogen for the binding site on estrogen receptors. When SERMs bind to the receptor they will change the conformation of the ligand-binding domain and block the necessary co-factor binding for the gene transcription as illustrated in figure 6c. The most common type of SERMs is tamoxifen (28, 40).

SERDs will entirely block the estrogen signalling pathway by degradation of the estrogen receptor as illustrated in figure 6d. SERDs are often given to counteract the cancer cells that are resistant to SERMs. The NCCN guidelines reported that the combination of fulvestrant, which is the most common SERD and CDK4/6 inhibitors was effective for advanced or metastatic ER-positive breast cancer (28, 40).

1.5.5 CDK4/6 inhibitors

CDK4/6 associates with cyclin D1 and phosphorylates RB which in turn releases the E2F transcription factor. E2F promotes the continuation of the cell cycle from G1 to S phase, this association are therefore a promotor for cell growth, which is one of the hallmarks of cancer (12). An upregulation of CDK4/6 and cyclin D1 are often observed in breast cancer, especially the HR-positive subtypes (41).

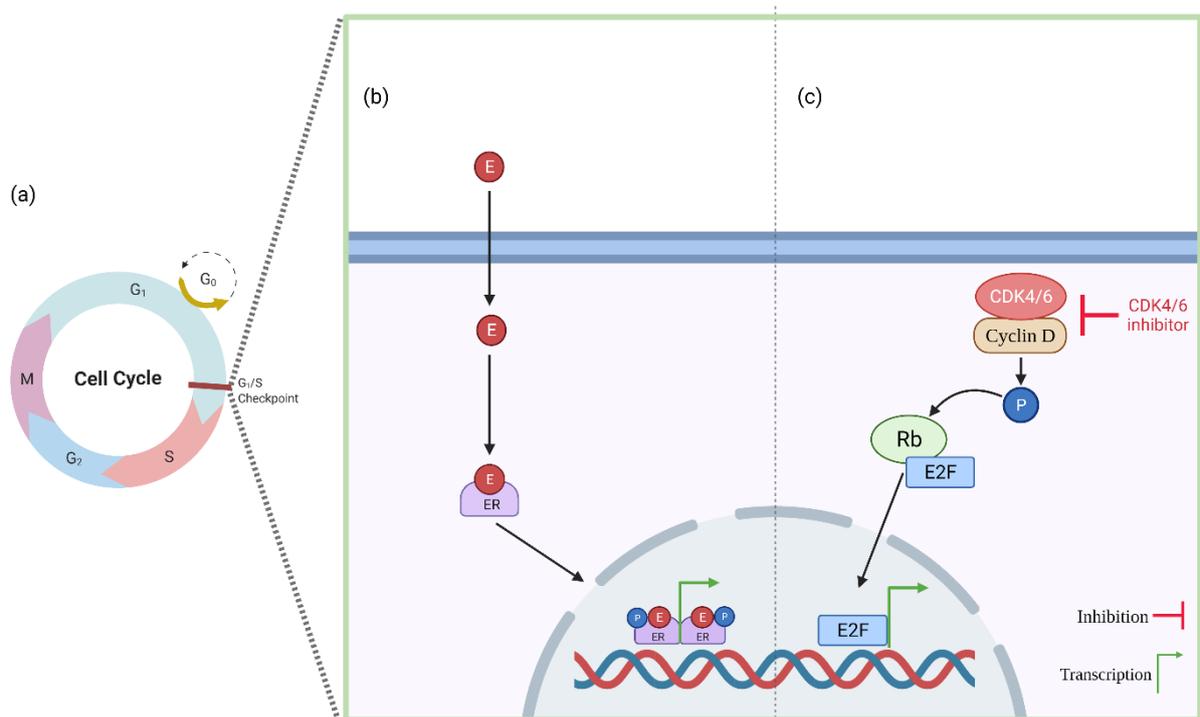


Figure 7. (a) In the cell cycle there are several checkpoints to control normal cell growth. One of these is the G₁-S checkpoint. (b) The estrogen receptor signalling pathway can also activate the cyclin D-CDK4/6 and induce the transition from the G₁ phase to the S phase. This makes the combination of endocrine therapy such as AI and CDK4/6 inhibitors very effective. (c) CDK4/6 inhibitor mechanism of action in the cell. The CDK4/6 inhibitor prevents the transition from G₁ to S phase by binding to the ATP-pocket in CDK4 and CDK6. This leads to a cell cycle arrest by prohibiting phosphorylation of Rb1 and activation of the protein. If the inhibition would not occur the Rb protein would be activated by phosphorylation and release the E2F transcription factor.

CDK4/6 inhibitors are a targeted therapy and mainly used to treat breast cancers patients with HR-positive and HER2-negative cancers but can also be applied to TNBC patients (42). CDK4/6 inhibitors arrest the cancer cells in the G₁ phase of the cell cycle by avoiding activation of the Rb-E2F complex as seen in figure 7. This inhibits the growth and proliferation of the cancer cells. The main three clinically used CDK4/6 inhibitors are palbociclib, ribociclib and abemaciclib (28, 41, 42). Palbociclib and ribociclib have been used in combination with endocrine therapy such as AI due to the ER-receptor signalling pathways ability to activate the cyclin D-CDK4/6-Rb pathway and progression from G₁-S phase, this combination has shown positive effect on the cancer cell growth (38, 42-44).

The CDK4/6 inhibitors have been tested as neoadjuvant treatment in clinical trials and show promising results (45). The combination of AI and CDK4/6 are not a standard treatment in Norway for locally advanced breast cancer patients, and one of the aims of the NeoLetRib clinical trial, studied here, is to make this treatment available for the patients in question. CDK4/6 inhibitors also have effects on the anti-tumor immunity by an increase of tumor antigen presentation, suppression of regulatory T cells (T_{reg}) proliferation and enhanced cytotoxic T-cell clearance of cancer cells (46, 47). Recent studies show that CDK4/6 inhibitors also induce remodelling of the chromatin architecture and activation of transcriptional enhancers that effect apoptosis and luminal differentiation as well as activation of RB1 and AP-1 expression (48).

1.6 Immune cells in the TME

The immune system is a defence system of the body and can be divided into two main groups namely the innate immune system which is our first line of defence and the adaptive immune system. The innate immune system is composed of physical barriers and phagocytic cells such as dendritic cells (DC), neutrophils, natural killer cells (NK), and macrophages. The adaptive immunity is more dynamic and evolves during life through interactions with different cells and microbes. The adaptive immune system is composed of B and T cells which are target specific and selective. A further division of the adaptive immune system is often made, which is the humoral immunity and cell-mediated immunity. B cells produces antibodies that mark targets that needs to be eliminated or neutralized by the immune system this is the humoral immunity (49-51).

Many intratumoral interactions occurs between cancer cells, tumor infiltrating cells and soluble mediators during immunoediting, which will be described in further details below. Some of the cells are tumor suppressors such as CD8⁺ cytotoxic T lymphocytes (CTLs), T helper cells 1 (Th1), DCs, M1 macrophages and N1 neutrophils. Other immune cells have tumor promoting properties, these cells include myeloid derived suppressor cells (MDSCs), regulatory T cells (T_{reg}), M2 macrophages and N2 neutrophils. The amount of the different immune cells influences prognosis and the cancer evolution. It is therefore important to have a good understanding of the quality and quantity of the different immune cells in a tumor. It has been observed a high heterogeneity in immune infiltration between patients with the same breast cancer subtype or between patients with different subtypes. The different immune cells found in breast cancer will here be elaborated further and are illustrated in figure 8 with their most important functions in the breast cancer TME and disease progression (52, 53).

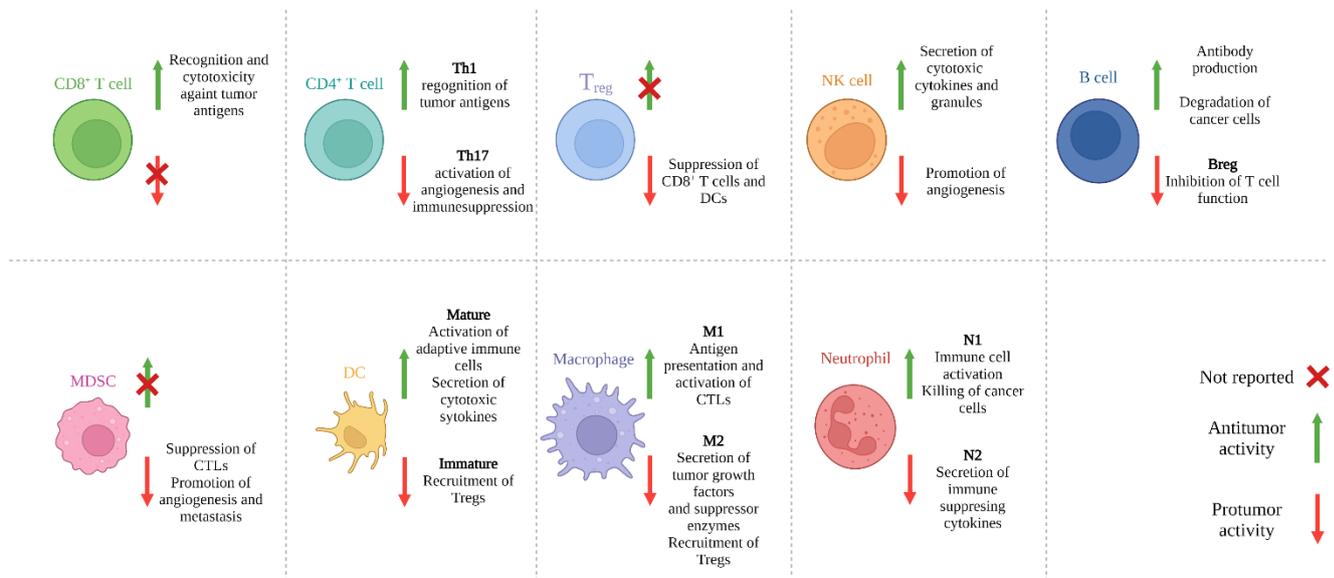


Figure 8. The immune cells found in the TME of breast cancer. CD8⁺ T cells are associated with antitumor activity by recognition of tumor antigens and their cytotoxic activity. CD4⁺ T cells have many subtypes including Th1 which is associated with antitumor activity, by recognising tumor antigens. Th17 are on the other hand associated with protumor activity by inducing angiogenesis and immunosuppression. Tregs are associated with protumor activity by inducing a immunosuppressive environment in the TME. NK cells have both pro- and antitumor activity by secreting cytotoxic cytokines and granzymes as well as promoting of angiogenesis. B cells and their role on breast cancer is not well known but there is observed both pro- and antitumor activity by degradation of cancer cells and antibody production. A subgroup of Bregs is observed to inhibit T cell function and induce tumor progression. MDSCs have protumor activity by suppression of CTLs and promotion of angiogenesis and metastasis. DCs have different activity based on their cellular state. Immature DCs have protumor activity by recruiting Tregs, while mature DCs activates the adaptive immune cells by secretion of cytotoxic cytokines. Macrophages are commonly divided into M1 and M2, M1 have antitumor activity by activation of CTLs and antigen presentation while M2 secretes tumor growth factors and induces the recruitment of Tregs. Neutrophils are commonly divided into N1 and N2, N1 have antitumor activity by activating immune cells and killing of cancer cells. N2 have protumor activity by secreting immune suppressive cytokines.

1.6.1 Lymphocytes

The tumor infiltrating lymphocytes can be both tumor suppressors and promoters. CTLs are associated with a better prognosis and overall survival for breast cancer patients and are essential for the destruction and suppression of cancer cells. CD4⁺ T lymphocytes consists of several subgroups of Ths, namely Th1, Th2 and Th17 as well as Tregs. Of these Th1 is associated with a better overall survival and are found in a lesser abundance in breast cancer patients compared to healthy patients, whereas many of the other CD4⁺ cells are associated with a worse prognosis. Th2 can induce inflammation which is associated with tumor progression. Th17 can have a degree of plasticity and can be associated with both worse prognosis and antitumor immunity (52, 53).

Tregs are immunosuppressive cells and have an important role in providing immune homeostasis (54). It is observed an increased amount of Tregs in the TME in breast cancer and are associated with a lower overall survival rate (55, 56). Tregs are considered as tumor promoting cells in breast cancers as they express granzyme A and B which induce the production of the anti-inflammatory cytokine IL-10, which is highly associated with immune escape. They also produce TGF- β by expressing CTLA-4, PD-1 and PD-L1 (56). Tregs also hinder the activation and proliferation of CTLs, which promotes the tumor progression (57). The relative amount of Tregs compared to CTLs have therefore shown to influence prognosis (52, 53, 57).

B cells are commonly fewer in number than T cells in breast cancer and are mostly associated with a good prognosis. They are often found close to the T cells due to their antigen presenting properties. When these cells mature, they turn into plasma cells which have an anti-tumor humoral effect on the cancer. Bregs have anti-immune abilities by inhibiting the functions of T cells by the production of PD-L1 and the cytokines IL-10 and TGF- β . Some B cells can induce the conversion of T cells to Tregs which have pro-tumor abilities as described above (52, 53, 58).

NKs are lymphocytes and a part of the innate immune system, they recognise abnormal cells, such as cancer cells, that lacks or have an underrepresented amount of major histocompatibility complex (MHC) I on their cell surface. NK cells secrete cytotoxic granules and cytokines that lyses the target cell. There are two main subgroups of NKs the one with highest cytotoxic properties is the CD56^{dim}CD16⁺ this is found in a lower concentration than the other subgroup in breast cancer patients. The other group are the CD^{bright}CD16⁻ which are less cytotoxic but produce more cytokines, this subgroup is reported to have angiogenic properties and can therefore have an importance as prognostic and predictive value (52, 58).

1.6.2 Dendritic cells

DCs belong to the innate immune system but is essential for creating a bridge between the innate and the adaptive immune system. This is mediated by activating CTLs and CD4⁺ T cells through antigen presentation via MHC I and MHC II molecules respectively, leading to priming of T cells. CTLs will then secrete granzymes and perforins leading to apoptosis of the target cell. DCs progress through a maturation process, but evidence is showing that in breast cancer some DCs are kept in the immature state, in which they produce both TGF- β and IL-10, that in turn recruit more Tregs with pro-tumor properties. It is therefore important that DCs mature to acquire anti-tumor properties (52, 59).

1.6.3 Macrophages

Macrophages are cells belonging to the innate immune system and differentiate from monocytes in tissues. They serve several functions such as phagocytosis, antigen presentation, and production of cytokines. As antigens are detected on foreign cells the macrophages can migrate to the target cell and lyse them, followed by the uptake of the antigen of the foreign cell, which the macrophage in turn can present to T cells via MHC I/II molecules (52, 59, 60).

Macrophages are commonly divided into two subgroups based on different phenotypes. The first subgroup is M1, also called the classical activated macrophages. These are activated by the secretion of interferon (IFN) - γ and tumor necrosis factor- α (TNF- α) from Th1 cells. M1 are associated with inflammation and an increased antigen presentation. This is promoted by the production of cytokines such as IL-6, IL-12 and TNF which activates the CTLs and NKs that in turn will lead to apoptosis of cancer cells. As M1 can activate T cells and itself are activated by Th1, they are associated with a good prognosis in cancer (52, 58-60).

The second group are M2, also called the alternative activated macrophages. These are activated by IL-4 and IL-13 secreted by Th2s. M2 are associated with anti-inflammatory responses, phagocytosis, and tissue remodelling. M2 helps the cancer cells to survive by producing tumor growth factors such as TGF- β and platelet-derived growth factor (PDGF),

suppressor enzymes such as IL-12 and chemokines. They also recruit more Tregs which are associated with tumor progression and worse prognosis (52, 58-60).

1.6.4 MDSC and neutrophils

MDSCs are cells with high heterogeneity and are recruited to the TME by cancer cells. In the TME MDSCs have pro-tumor properties and induce angiogenesis, immune evasion, and metastasis. This is done by secretion of the cytokines TGF- β and IL-10 which inhibit the anti-tumor immune cells, they also promote the formation of nitric oxide and reactive oxygen species (ROS) which can inhibit the immune cells from degrading the cancer cells (52).

Another population of cells with very similar morphology are the neutrophils. They are a part of the innate immune cells and are commonly one of the first cells to reach a damaged tissue. At the site of damage, they can perform phagocytosis of cells and eliminate pathogens. The neutrophils can be divided into two subgroups N1 and N2 which have anti- and protumor properties respectively. N1 can activate both the adaptive and innate immune cells. They also produce ROS which may kill cancer cells. N2 on the other hand produce pro-tumorigenic cytokines that can inhibit the communication and activation of T cells and B cells, which will promote the tumor growth. These two subtypes lack good marker genes to distinguish them from each other (33, 52, 58, 59).

1.7 Immunoediting

The immune system has a critical role in cancer evolution. Cells of the immune system may have antitumor or protumor properties this allows the immune system to shape the TME in a process called immunoediting. Immunoediting includes three phases: elimination, equilibrium, and escape, as shown in figure 9 (49-51).

Both the innate and adaptive immune system are active during elimination, and it is thought that this often occurs long before the cancers get clinically visible. Meaning that the patient will be cancer free if this phase is completed. In this phase the cancer cells can release signals such as type I interferons (IFN) which activates DCs and the adaptive antitumor immune response. The tumor cells can also express tumor antigens that are recognized by the adaptive immune system leading to destruction of the cancer cells. This will induce the production of cytokines, which again will trigger the progression to an adaptive immune response, this process is illustrated in figure9 a and b (49-51).

IMMUNOEDITING

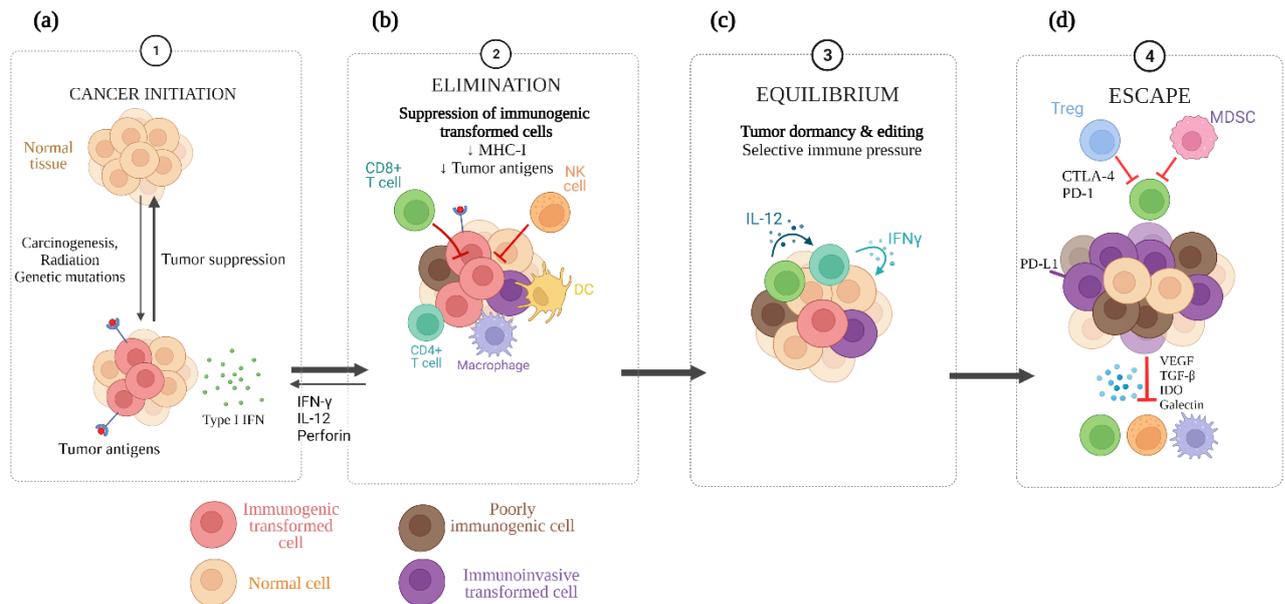


Figure 9: The steps in immunoeediting, a mechanism occurring when a cellular transformation has happened and intrinsic cancer suppressor mechanism have failed, this is illustrated in (a) where normal tissue goes through carcinogenesis where the cancer cells have tumor antigens and can produce IFN signal that will recruit cells from the immune systems. (b) The cells that pass to the first phase of the immunoeediting will therefore be exposed to cells from both the innate and adaptive immune system where the cancer cells are destroyed. (c) Some of the cancer cells can escape this immune destruction enter the equilibrium phase where the tumor is kept in dormancy and editing of the TME can occur. Cells from the adaptive immune system including T-cells, IL-12 and IFN- γ are necessary to keep the tumor in a state of dormancy. (d) Some of the cancer cells evolve due to the high immune selection pressure and develop properties that make them unrecognizable by the immune system, favour an immunosuppressive area within the TME or be insensitive to immune mechanisms. These cancers are the ones that are clinically visible.

If the tumor consists of cancer cells which have survived the elimination process, they enter the second phase, equilibrium illustrated in figure 9c. In this phase cells from the adaptive immune system are the most abundant, which includes T cells, and cytokines derived from T cells such as IL-12, and IFN- γ . The T cells can keep the tumor cells in dormancy and editing of the cancer immunogenicity can occur. Potentially, cancer cells can be kept in dormancy and never evolve to the final phase. When genetically unstable cancer cells are kept in the equilibrium phase over time there will be a lot of immune selection pressure, where the cancer cells can evolve and avoid immune system recognition (49-51).

Some of these cells can pass through to the last phase where they escape the immune surveillance, this is illustrated in figure 9d. This can be cancer cells that are not recognized by the adaptive immune cells due to downregulation of cancer antigens, cells that are insensitive to the effector mechanisms of the immune cells by downregulation of MHC I/II or cells that construct an immunosuppressive environment within the TME. Cancer cells can induce an immunosuppressive state in several ways, one, through production of immunosuppressive cytokines such as galectin, TGF- β , VEGF and indoleamine 2,3-dioxygenase. Another mechanism is the recruitment of Tregs and MDSCs. Tregs are mainly CD4⁺ T cells with CD25 and Foxp3 expression. Such cells will inhibit the tumor specific T lymphocytes by the production of CTLA-4, PD-1 and PD-L1 which induces the production of the immunosuppressive cytokines IL-10 and TGF- β . MDSCs are immature myeloid cells with the ability to inhibit lymphocytes by induction of Tregs, they also decrease the amount of several amino acids important for T cell function (49, 51). After the immunoediting the TME will call for Tregs and macrophages with the ability to suppress other immune cells. Some cancer cells can actively suppress T cells by expressing PD-L1 that binds to PD-1 on CTLs and deactivate them, this is also an example of an immune checkpoint (49-51).

1.8 Tumor microenvironment

Breast cancer cells have dynamic interactions with both resident and infiltrating cells such as immune cells, stromal cells, and secreted factors, that together form the TME illustrated in figure 10. The TME shapes the cancer progression and patient outcome, which includes the response to treatment (61). The tumor stroma consists of non-cellular proteins, fibroblasts, endothelial cells, mesenchymal cells and pericytes. The TME also consists of cancer cells that are the drivers of the disease with mutations described as one of the hallmarks of cancer. The cancer cells show high heterogeneity and differ greatly in their attributes such as their capacity to interact with the immune system, response to therapy and gene expression. The TME has a great molecular and cellular heterogeneity as proven by the high diversity of cell types. The different cells found in the TME may modulate tumor promotion and/or prohibit the cancer growth (62). The TME can consist of several niches such as an immune niche consisting of mostly immune cells that often are found in the periphery of the TME and a hypoxic niche that will have low pH and favour immune escape and are often a characteristic of aggressive cancers (63).

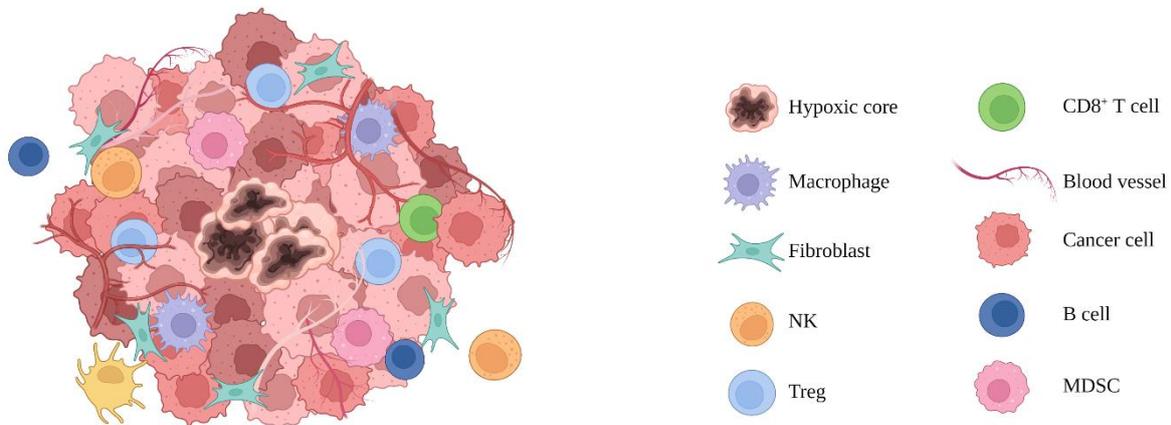


Figure 10. An illustration of some of the various cells composing the TME. Tumors recruit many cells from the immune system including macrophages, NKs, Tregs, DCs, CTLs, B cells and MDSC. They can also recruit the stromal fibroblasts which can obtain a cancer associated phenotype. During the cancer progression there will be formed a hypoxic core due to the high demand of energy in the cancer cells, one way to get energy and nutrients to the TME is by angiogenesis.

The breast cancer TME are often organized into the cellular, soluble, or physical components. The cellular component is again divided into the intratumoral, regional and metastatic categories. The intratumoral compartment includes cancer cells and recruited cells such as lymphocytes, DCs, macrophages, and neutrophils. Tumor-infiltrating lymphocytes that are present in breast cancer are mostly T cells, but there are also found some B cells. There are several subclasses of T cells with different influence on the TME as described earlier (58, 64).

The regional compartment includes stromal cells such as fibroblasts, adipocytes, myoepithelial cells, and endothelial cells. These cells will be described further below.

The metastatic compartment includes distant organs where new TME is forming, lymph nodes, blood, and immune cells. The metastatic compartment will not be described in further detail since we focus on the primary cancer site in this thesis. The soluble components include secreted factors such as enzymes, proteinases, cytokines, IFN, IL, growth factors, TGF- β and VEGF. Physical factors include pH and oxygen levels which can induce hypoxic conditions (58, 64).

The interactions between cells in the tumor microenvironment are caused by factors like chemokines, cytokines, growth factors, matrix remodelling enzymes, inflammatory mediators, exosomes and circulating tumor cells (58, 64). Some examples of factors important for the cell interactions in the TME are colony-stimulating factors (CSF) and cytokines such as CXCL2 and CCL22 which promote the recruitment of the immune suppressive MDSCs to the TME. Another example is TGF- β which have anti-tumor properties in premalignant cells but change to have protumor properties in more advanced cancer cells. TGF- β also suppresses the expression of IL-2 which is necessary for T cell functions, the molecule also recruit Tregs which will help the tumor to proliferate (62). Understanding these interactions is of great importance and a study of all the different specialized cells in the tumor is needed to properly understand the biology of cancers. It would be of great interest to investigate which subtypes of the different recruited cells that are present in the TME. This insight can provide new possible targets for immunotherapy and other treatment options.

1.8.1 Fibroblasts

Cancer-associated fibroblasts (CAFs) are an important component of the TME as they have been associated with invasion, metastasis, proliferation, and immune escape of the cancer cells (65). They are the most prominent stromal cells in the TME and are found in the regional microenvironment. CAFs can secrete tumor growth signals such as VEGF and TGF- β when they are activated. When fibroblasts are recruited to the TME, which consists of damaged cells and tissue, they can become hyper-activated leading to tissue fibrosis. Tissue fibrosis has been observed to be an important driver of breast cancer progression and a crucial part of creating “the wound that never heal” in breast cancer (52, 53, 58, 66).

CAFs can regulate the recruited immune cells by secreting molecules such as chemokines, cytokines, and inflammatory mediators. As they are a component of the stroma, they also have a prominent function of remodelling of the ECM that in turn manage the intratumoral infiltration. The immunosuppressive fibroblasts can collaborate with Tregs and promote tumor growth. They have also been observed to promote differentiation of M2 macrophages and inflammation by secretion of factors such as glycoprotein chitinase-3-like-1 (52, 53, 58, 66).

1.8.2 Epithelial cells

Epithelial cells are the main cells that construct the lobules of the mammary ducts and are commonly described as the luminal epithelial cells and the myoepithelial cells. During the early stage of breast cancer development, myoepithelial cells function as natural tumor suppressors as they make a protective barrier between the luminal epithelial layer and the stroma, as well as inhibiting invasion and angiogenesis. One of the steps in breast cancer progression is therefore an alteration or loss of the normal myoepithelial cells. It is seen that myoepithelial cells in DCIS have an upregulation of genes that promote proliferation, invasion, and angiogenesis such as CXCL12, CXCL14, and matrix metalloproteinases that degrades the basement membrane and facilitates tumor invasion. They also show a lower expression of genes related to normal myoepithelial cell function such as laminin-1 which is important for the basement membrane, oxytocin receptor and proteinase inhibitors that suppresses proliferation of cancer cells (17, 64).

1.8.3 Endothelial cells

Endothelial cells are part of the tumor stroma, these cells have high heterogeneity and are involved in angiogenesis and are building blocks of new blood vessels, thereby supporting the growth and proliferation of the primary tumor. The process of angiogenesis is promoted by growth factors such as VEGF. Tumor endothelial cells also provide inter-signalling pathways and promote metastasis and resistance to drugs, due to the disorganization of the tumor endothelial cells leading to the characteristic leaky vascular system in solid cancers (67).

1.9 Single Cell Analyses

The traditional method for transcriptome analyses is through the lysis and analysis of the RNA obtained from all cells from a tissue sample, this is often called bulk RNA-seq. After bulk RNA-seq the gene expression is therefore the average of all cells in the tissue sample, which is sufficient for some analyses but usually does not allow the description of the tissue heterogeneity at the molecular level and detection of novel cell subtypes (68).

Single-cell RNAseq (scRNA-seq) is a method which allow to analyse the transcriptome of single cells in a tissue and is an improvement of bulk RNA-seq. scRNA-seq was first performed in 2009 by Tang et al. (69). In cancer research this method has the potential to detect novel and rare cell populations and depict the heterogeneity of the TME, by providing a modelling of the transcriptional states in individual cells from tissue samples (61, 70). T lymphocytes are an example of a cell population with high heterogeneity, recently many subtypes of T lymphocytes have been reported in cancers using scRNA-seq (71). The gene expression profile from the single cells gives information of gene regulatory networks, co-expression and regulation, and the cellular states. This gives the opportunity to analyse the transcriptomics of the single cells in a high definition and help to characterise complex tissue heterogeneity and how the cells in the TME differentiate, grow, and metastasise during tumor progression and treatment (68, 72). An early study from a Korean group reports scRNA-seq on 515 cells from 11 patients with different types of breast cancer and found that ER/HER2-positive breast cancers could have a high ER expression and a low HER2 expression, indicating that the patients should receive hormonal treatment to inhibit the cancer growth and not have the HER2 positive cells as a main target (73). Another study performed by Lee et al. found high molecular intra-heterogeneity in cells that were not observed in bulk RNA-seq of a metastatic breast cancer cell line. Cell-specific RNA variants promoted survival of cells capable of evolving a drug-resistant phenotype (74). This points out that many of the cancer cells within a tumor can be killed by targeted therapy, but that some specific subtypes can survive and continue the tumor progression. As scRNA-seq unravels the heterogeneity of the TME it can be a useful method to discover better treatment targets for rare cancer cell types, as intratumoral heterogeneity can cause drug resistance. The characterization and phenotyping of the different cell types found in a tumor is of great importance for future individualized

therapy and provide new information of how the TME responds to current treatment options (72).

1.9.1 scRNA-seq pipeline

There are different protocols for scRNA-seq analysis, and the sensitivity and specificity can vary between them. All scRNA-seq pipelines start with the isolation of single cells from dissociated tissue samples. There are different options for isolation of single intact cells, single nuclei, or a combination. It is important that the isolation of the single cells take place not long after tissue resection as cells will start to die and change when taken out of their systemic context. There are several methods that are compatible with the downstream analysis for the isolation of single cells one of them are flow-activated cell sorting (FACS). In this method up to seventeen markers of fluorescently labelled antibodies can be used to sort and isolate cells based on targeted cell-surface markers. This method is user friendly and cost effective, but requires a large sample volume and is therefore not well suited for samples from needle biopsies and aspirates (75). Another method for the isolation of single cells is a droplet-based technology. 10x Genomics is one of the providers of such technology. After the tissue dissociation the single cells will be embedded in a droplet and brought in presence with a unique bead coated with a unique barcode, which will provide a traceback from the transcripts to the single cells (76).

After the isolation of the single cells the polyadenylated mRNA molecules are captured succeeding lysis of the cells. By targeting the polyadenylated mRNA the most informative RNAs will be kept while the less informative mRNAs such as tRNA and rRNA are left out of the analysis (75). As sequencing requires DNA, a cDNA library is constructed using the reverse transcriptase enzyme. Then an amplification of the cDNA libraries is performed, amplification is commonly done by template-switching on PCR, which favours full-length transcripts and strand specificity, the strand specificity are kept even after fragmentation by selective PCR (selecting 3' ends) (75). After this samples standard library preparation methods are used to obtain samples ready for NGS. Then the results are analysed by the help of computational analysis and bioinformatics. Some of the scRNA-seq methods have full-

length counting while others have counting of only the 3'-end, the full-length counting are shown to have a higher sensitivity (68, 76).

2. Thesis aims

Single-cell RNA sequencing can uncover the phenotype of each cell and, help to refine the molecular subtypes, detect novel and rare cell populations, and depict the heterogeneity of the tumor microenvironment. This paves the way for the investigation of how tumor infiltrating cells and the cancer cells synergises and evolves during treatment, which gives a deep understanding of the patient's response to the given drugs.

The samples in this thesis are from patients enrolled in the NeoLetRib clinical trial, consisting of women with ER-positive, HER2-neagive, locally advanced breast cancer. They are treated with a combination of aromatase and CDK4/6 inhibitors (Letrozole and ribociclib respectively).

Main aims is therefore to use the information from single cell RNA sequencing to assess how the cells within the tumor microenvironment respond and evolve during neoadjuvant treatment. The main question we aim at answering is: Are some specific and specialized immune cells quantities within the tumor changing during treatment? Our work studies the tumor microenvironment of ER-positive breast cancer in a very detailed and systematic manner.

3. Materials and methods

3.1 Patient information

The patients included in the NeoLetRib trial have locally advanced, ER-positive (>50% ER-positive cancer cells), HER2-negative, low proliferative breast cancer. Locally advanced is defined as large primary breast cancers with stage T2 or T3/T4, or N2-3. The patients receive neoadjuvant therapy with a combination of letrozole and ribociclib for at least 6 months before surgery. All patients are postmenopausal women.

Patients receive 2.5 mg letrozole once a day, and 600 mg ribociclib for 21 days, followed by 7 days without ribociclib. A needle-core biopsy is taken before treatment, 21 days after treatment initiation and after 6 months. As this thesis lasts 5 months, the last timepoint will not be analysed.

3.2 Tissue preparation

To prepare single cells suspension from core needle biopsies a dissociation of the tissue is performed. All biopsies were collected at Akershus University Hospital and were dissociated immediately upon arrival in our lab at Ullevål. The tumor was first cut in small pieces on a petri dish placed on ice. The small tissue pieces in the petri dish were collected using 2.5 ml pre-heated dissociation solution (Appendix 1). The solution containing most of the tumor fragments was transferred to a polystyrene tube. A second wash of the petri dish with 2.5 ml pre-heated dissociation mix was performed to collect any remaining tumor pieces. The remaining tumor pieces were transferred to a second polystyrene tube. Both tubes were incubated in a water bath at 37°C for maximum 20 minutes. After 10 minutes the tissue fragments were pipetted carefully to add mechanical pressure and enhance tissue dissociation, the pipette tip was trimmed to remove some of the force.

After the dissociation the samples were filtered through 70 μ M and 40 μ M filters into two clean polystyrene tubes. The two tubes containing the filtered dissociated cells were then centrifuged at 350g for 10 min 4°C. After centrifugation the supernatant was removed without disturbing the pellet. The smallest pellet was then resuspended in 250 μ L of buffer 2

(Appendix 1) and transferred to the tube containing the largest pellet. The first tube was then washed with another 250 μL of buffer 2 before it was transferred to the second tube with the largest pellet. In total, all cells were resuspended in 500 μL of buffer 2. To remove red blood cells 25-50 μL of RBC beads were added according to size and colour of the pellet. The tube without lid was then placed in a magnet for 3 min. After this the cells were transferred to a cold low-bind 1,5 mL Eppendorf tube by inversion.

At this stage cell numbers were evaluated by counting. 12.5 μL of cell suspension which was added to an Eppendorf tube with 37.5 μL PBS and 50 μL of buffer A. Then the sample was vortexed before adding 50 μL buffer B, a second vortex was performed giving a dilution of 1:12 of the alive cells. Another 12.5 μL cell suspensions was added to another Eppendorf tube containing 87.5 μL PBS for the dead cell count with a dilution of 1:8. The cells were counted using a NucleoCounter. After this the samples in the cold low-bind tube were centrifuged at 350g for 10 min at 4°C. The supernatant was removed, and the pellet resuspended in PBS+BSA 0.04% (Appendix 1) for a second counting. The volume of PBS+BSA 0.004% was evaluated after the first counting to target a final concentration of 1000 cells/ μL . To count the alive cells 10 μL cells was added to an Eppendorf tube containing 40 μL PBS and 50 μL buffer A, then the sample was vortexed. 50 μL buffer B was the added and the sample was vortexed again giving a dilution of 1:15. The dead cells were counted by adding 10 μL cells to 90 μL PBS in another Eppendorf tube giving a dilution of 1:10. If the second count gave >1800 cells/ μL the sample was diluted by adding PBS-BSA 0.004% to target 1000 cells/ μL .

3.3 Analysis of gene expression by 10x Genomics

Analyses of the gene expression and the immune repertoire of the single cells was done by using 10x Genomics Chromium Single Cell 5' protocol with reagent kit v2. The method is divided into six steps and illustrated in figure 11; the first step is the generation of Gel beads-in-EMulsion (GEMs) seen in figure 11a. This is done by merging the cells, master mix, 10x Barcoded beads and imbedded them in an oil emulsion performed by the Chromium Next GEM Chip. Where the single cells are introduced to a stream of beads in a limited dilution resulting in 1-9% of the GEMs containing a single cell, this is illustrated in figure 11b. The gel beads inside the GEMs are dissociated and oligonucleotides are formed containing an

Illumina R1 sequence, a 16 nt 10x Barcode, a 10 nt long unique molecular identifier (UMI) and a template switch oligo (TSO) of 13 nt. These are mixed with the master mix, reverse transcription reagents, poly(dT) RT primers, and the cell lysate. After incubation of the GEMs the fully barcoded cDNA are constructed, this process is shown in figure 11c. The 10x Barcode gives a traceback for all cDNAs within one GEM.

Following the GEM incubation, clean-up is the next step where the GEMs are dissociated, pooled, and purified from all primers and biochemical reagents. This was done by the application of silane magnetic beads. Then the full-length cDNA is amplified by template switch on PCR to obtain enough material to construct gene expression libraries for 5' gene expression, T-cell receptors, and B-cell receptors from the same cell. After this sample indexes and Illumina R2 sequence are added by the help of end-repair, A-tailing, adaptor ligation and sample index PCR, this is shown in figure 11d. The libraries were then sequenced on a NovaSeq 6000 Illumina sequencer using 150 bp paired end sequencing. The sequencing was performed by the Norwegian Sequencing Centre.

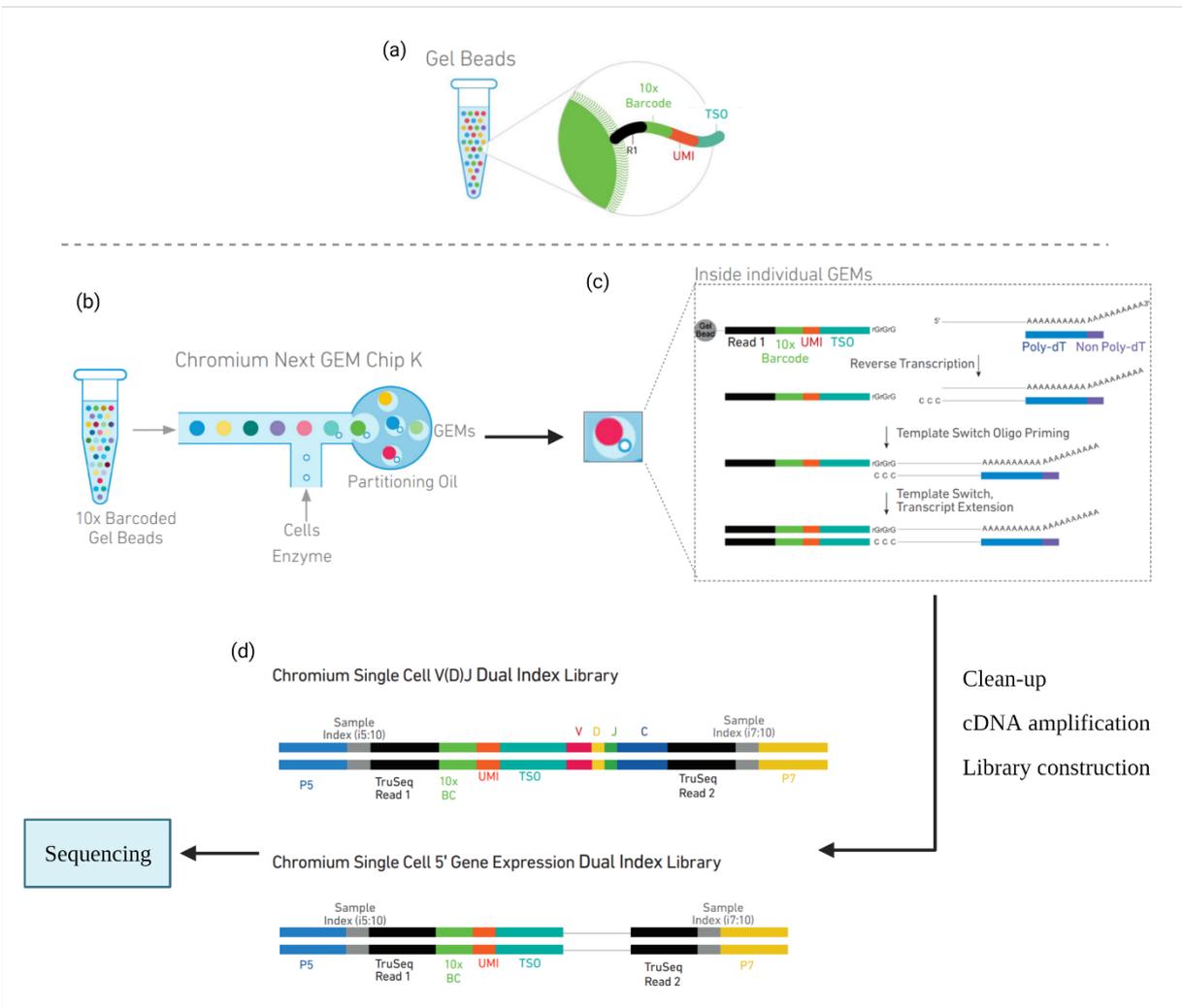


Figure 11. 10x Genomics workflow. (a) The method is based on a droplet technology where Single Cell VDJ 5' gel beads with oligonucleotides consisting of Illumina R1 sequence, 16nt 10x Barcode, 10nt UMI and 13nt TSO. (b) GEMs contain 10x Barcoded Gel Beads and cells, which are imbedded in oil droplets, this is performed on a Chromium Next GEM Chip K. (c) Inside each individual GEM the cell lysate, a master mix containing reverse transcriptase and poly(dT) primers allows reverse transcription in each droplet and template switch priming. Followed by template switch amplification of cDNA by PCR. After this a clean-up of biochemical reagents and primers are executed, and enough cDNA are amplified for library construction. (d) The final dual libraries contain the P5, i5 and P7, i7 sample indexes and Illumina read 2. After library construction the sample is ready for NGS. Figures are modified (77).

3.4 Computational analysis

10X genomics have developed pipelines for analysis of the raw data from the sequencing. The pipeline aligns the reads and map them to the human reference genome hg38 followed by the production of a feature-barcode matrix. The pipeline is called Cell Ranger and was applied by the Norwegian Sequencing Centre. The output from Cell Ranger are count matrices and were then processed further in downstream analyses in this thesis (78).

Seurat is a package designed for analysing scRNA-seq data. Gene expression matrices from Cell Ranger were loaded into R using the `Read10X()` function from the Seurat (v4.0.2) package, this will transform the data into a “Seurat object” which contains both matrixes and metadata for the dataset. Seurat objects are S4 type of objects containing many slots which are filed in downstream analyses. The sparse matrices available as ‘slots’ in the Seurat object contain the genes in rows and cells in columns.

Several bioinformatic quality control steps such as thresholds for the library size, fraction of mitochondrial genes, doublets and empty drops was applied. First, empty droplets and droplets containing two cells were filtered out using the `EmptyDrops` (79) and `scDb1Finder` (80) algorithms respectively. In addition, cell barcodes with < 500 UMIs, <300 expressed genes, >5000 expressed genes or >15% of reads mapping to mitochondrial RNA were filtered out. The gene expression of the remaining good quality cells was normalized.

The normalization was performed with a scale factor of 10 000, this normalized the gene expression value in each cell by the expression seen in the overall expression and gave a more accurate gene expression profile. The scale factor was multiplied with the normalized gene expression and log transformed. This was done by the function `NormalizeData()` included in the Seurat package. After this a scaling and centring of the genes was done by the function `ScaleData()`. Then a reduction of dimensionality was performed by a linear dimensional reduction method called Principal component analysis (PCA), where 35% of the genes with highest variability was included. Thirty-five principal components (PCs) were included in the downstream analysis. This was chosen based on an evaluation of an `JackStrawPlot` where

the distribution of p-values is compared, in this plot the PCs with highest influence on the dataset will have the lowest p-value. Another method used for the evaluation was an `ElbowPlot`, examples of these are found in Appendix 2, where the PCs was ranked by the percentage of variance. In the `ElbowPlot` there will be an “elbow” where the variance flattens out and indicates that the PCs in this area are explaining less variance of the dataset (81). After the reduction of the dimensionality, where one gene was a new dimension, the cells was clustered based on their similarity in transcriptional profile (76).

The clusters were visualized by uniform manifold approximation and projection (UMAP), which applies a non-linear dimensional reduction of the dataset, and placed the clusters with cells similar to each other in a low dimensional space. An example code to go from count matrices to clustering and UMAP plotting is shown in Appendix 3.

3.5 Own contributions

The last five months I have performed all the above-mentioned methods, from tissue preparation to computational analysis, with the exception of the Cell Ranger pipeline which was performed by the Norwegian Sequencing Centre. The scripts used in the computational analysis were mainly written by Xavier Tekpli based on Seurat vignette:

https://satijalab.org/seurat/articles/pbmc3k_tutorial.html, but I have made alterations for the downstream analysis to fit my dataset.

4. Results

4.1 Clustering and annotation of all cells

All cells from the scRNA-seq data were clustered and visualised by the application of the R package Seurat and UMAP, which performs a non-linear dimensionality reduction of the high dimensional dataset. The nine doublets from patient samples comprises a total of 104 728 cells and were separated in 51 clusters shown in figure 12a. The annotation of the clusters obtained was performed by plotting the expression of established marker genes for the main cell types expected to be found in breast tumors (70, 82, 83). Log-normalised marker genes' expression are projected to the UMAP in figure 12b and illustrate how the 51 clusters represent different cell types. Figure 12c shows the average expression and percentage of cells expressing three marker genes for each cell type. A full overview of the marker genes used to annotate the clusters are given in Appendix 4. Following careful inspection of the set of seventy-three marker genes, we obtained the final annotation shown in figure 12d. The number and proportion of each cell type identified is given in table 1.

In conclusion these results show that we were able to identify and annotate eight different main cell types that have been reported previously by performing scRNA-seq. Our results also show that our analytic pipeline of single cell clustering groups cells according to their phenotype while avoiding batch effects.

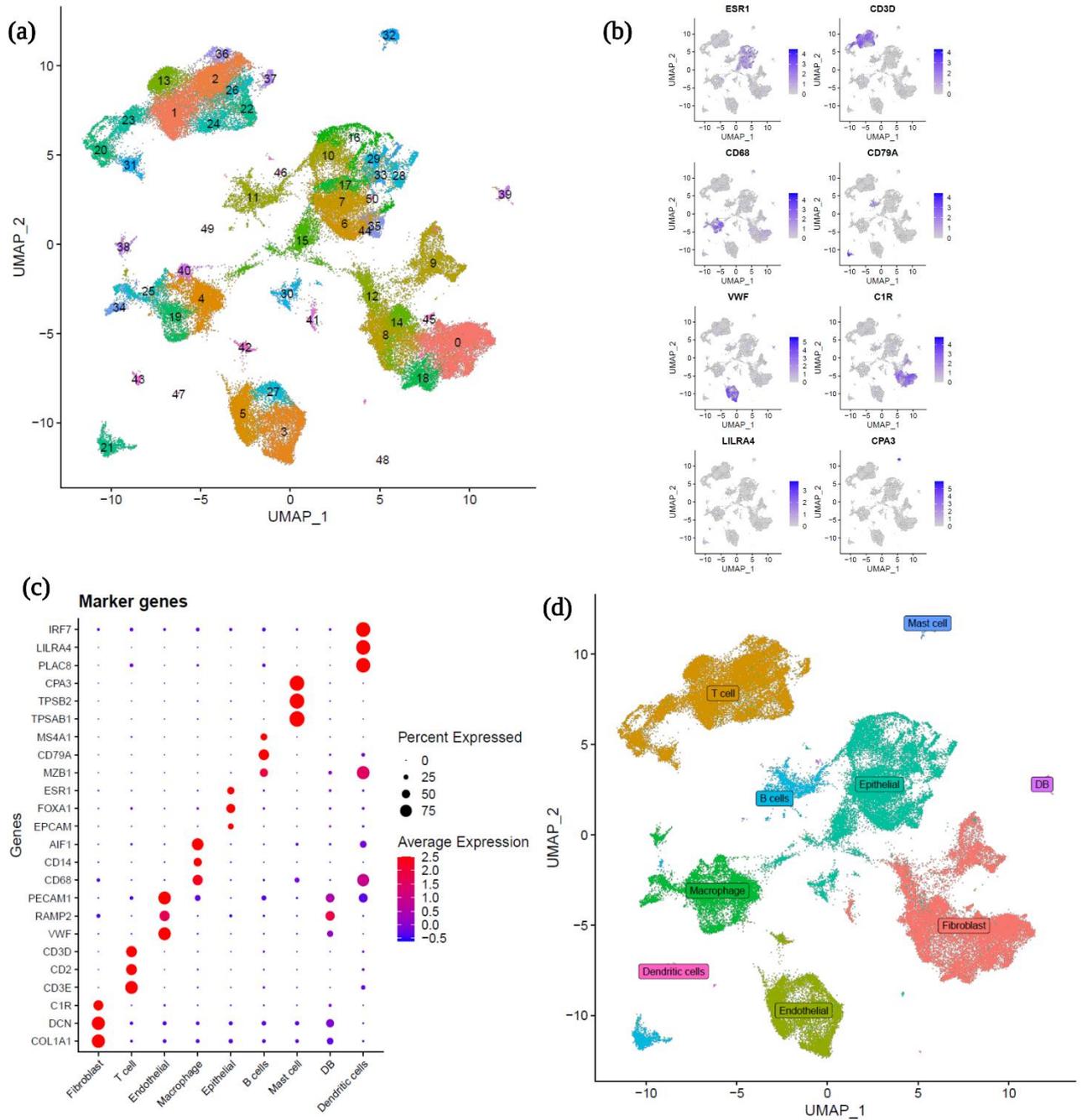


Figure 12. (a) UMAP of 104 728 cells from 9 ER-positive, HER2-negative breast cancer patients before and after treatment. The cells were grouped in 51 clusters. (b), Illustrates log-normalized expression of marker genes for epithelial cells (ESR1), T cells (CD3D), macrophages (CD68), B cells (CD79A), endothelial cells (VWF), fibroblasts (C1R), DCs (LILRA4) and mast cells (CPA3). (c) Dot plot for 3 marker genes, average gene expression, and percentage of expression for the different cell types. (d) Annotated UMAP showing the 9 main cell types identified in breast cancer biopsies.

Table 1 shows the proportions of the different cell types from the dataset. In this thesis the T cells, Macrophages and fibroblasts was selected for further analysis as they are important cells for the tumor progression.

Table 1. A table with the 8 cell types identified by scRNA-seq and a cluster annotated as doublets, with their proportion in the dataset as a percentage.

Cell type	Proportion (%)
B cells	4,60
Dendritic cells	0,27
Doublets (DB)	0,68
Endothelial cells	11,25
Epithelial cells	23,13
Fibroblasts	24,04
Macrophages	9,96
Mast cells	0,80
T cells	25,26

4.2 NK and T cell cluster annotation

The NK and T cells which represented 21 435 cells were clustered independently to identify more specialised and specific T cell types. After clustering we obtained a UMAP with eighteen clusters (Appendix 5). To carefully annotate these eighteen clusters, we used set of genes which have previously allowed to identify T cell types in scRNA-seq data (70, 82). To assess the expression of these marker genes in the clusters, we averaged the expression of each gene in the clusters and plotted them in heatmaps (Appendix 5). Based on the co-expression of marker genes for specific cell types, we identified a total of ten T and NK cell types. Included regulatory T cells (CD4⁺ reg), naïve T cells (CD4⁺ n and CD8⁺ n), exhausted T cells (CD4⁺ ex), effector/memory T cells (CD4⁺ em and CD8⁺ em), tissue-resident memory T cells (CD8⁺ rm), recently activated effector/memory T cells (CD8⁺ emra), resting NK cells (NK rest) and cytotoxic NK cells (NK cyto).

Feature plots with log normalised expression of marker genes for the given phenotypes shown in figure 13a confirms the annotation of the clusters. The T cell types showcased in the feature plots are CD4 effector/memory, regulatory and exhausted cells with respective marker genes being CD40LG, FOXP3 and CXCL13. CD8 phenotypes for tissue-resident memory T cells, effector/memory and recently activated effector/memory are marked by ZNF683, GZMK and GZMH respectively. Marker genes for resting and cytotoxic NK cells are also illustrated with AREG and FCGR3 being the markers of choice respectively. All marker genes used to initially annotate and identify NK and T cell types are shown in the heatmap in figure 13c.

Following careful inspection of marker genes using heatmap, violin plots and UMAP, we settled on the final annotation of T cells and NK cells shown in figure 13b. To visualise how the different patients' T cell composition at baseline (before treatment) may vary, we plotted the proportion of the different T cell types using barplots in figure 13d. The four subtypes with the highest representation are the effector/memory T cells ($CD4^+$ em and $CD8^+$ em), tissue-resident T memory T cells ($CD8^+$ rm) and naïve $CD8^+$ T cells. The T cell composition in the samples varies and confirms a high intra-tumoral heterogeneity of the T cell microenvironment. In conclusion, our results show that the clustering of T cells and NKs allow a detailed annotation of ten subtypes.

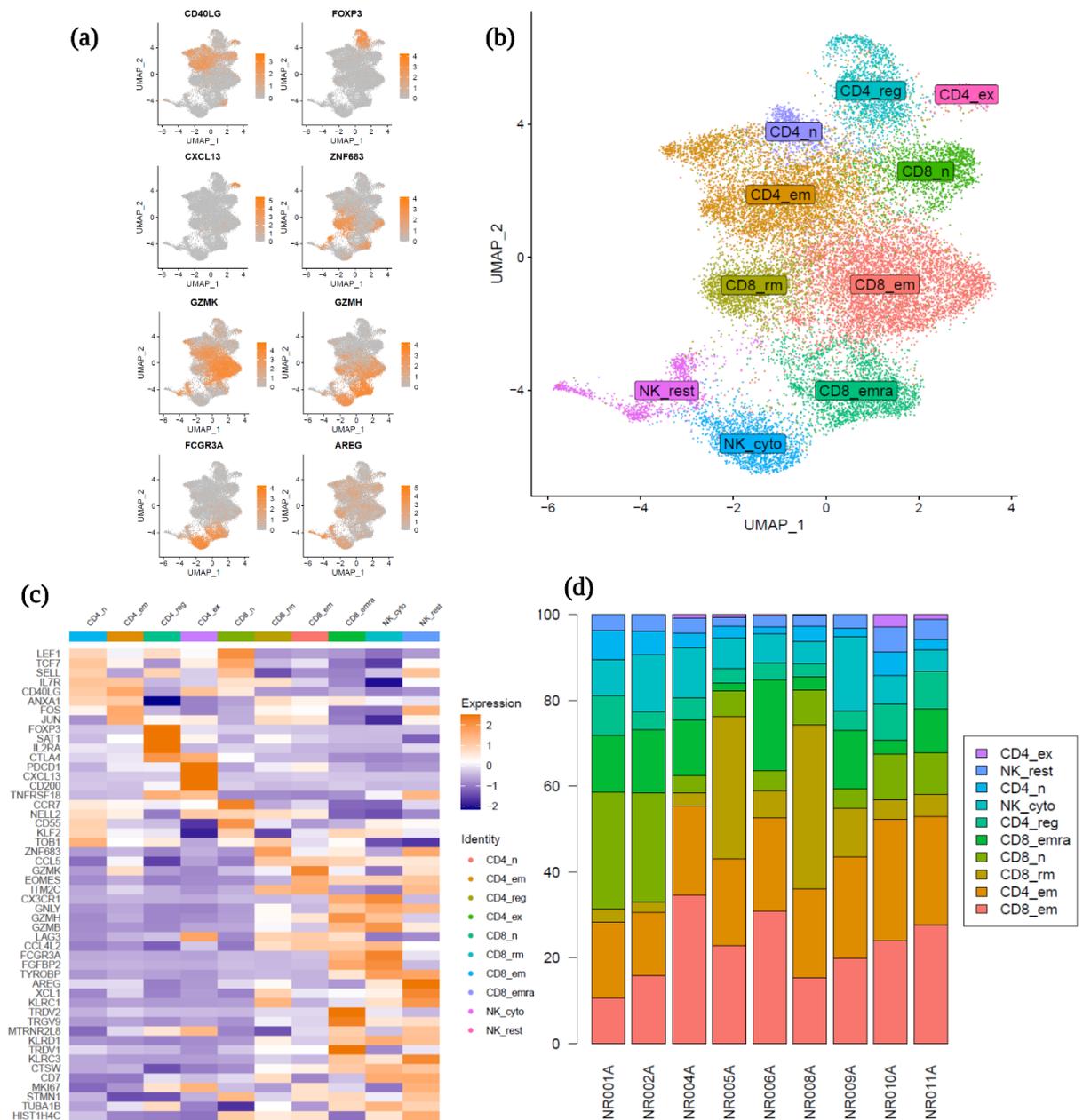


Figure 134. (a) A selection of a few marker genes for specific T and NK cell subtypes is used to illustrate the repartition of CD4 effector/memory, regulatory and exhausted cells as well as resting and cytotoxic NK cells and different subtypes of CD8 cells. (b) The final annotated clustering of 21 435 T cells and NK cells into 10 phenotypes. Two clusters represent cytotoxic and resting NK cells (NK_rest and NK_cyto), regulatory T cells (CD4_reg), naïve T cells (CD4_n and CD8_n), experienced T cells (CD4_ex), effector/memory T cells (CD4_em and CD8_em), tissue-resident memory T cells (CD8_rm) and recently activated effector/memory T cells (CD8_emra). (c) Heatmap showing the expression of the marker genes across the ten phenotypes of NK cells and T cells. (d) Barplot representing the tumor T cell composition of biopsies at baseline, before treatment.

4.3 Macrophage cluster annotation

The first cluster of the macrophages is shown in figure 14a. Annotation of the different clusters obtained was done using marker genes as previously published by Qian et al. (70, 82). The annotated macrophages include eleven types of myeloid cell types based on different gene expression profiles. Three groups were classified as monocytes being classical, non-classical and normal tissue-specific monocytes (Mono_CD14, Mono_CD16 and Mono_normal respectively). The macrophages were divided into five main groups based on the features being early stage (Mf_CCR2 and Mf_CCL2), tumor-associated (Mf_CCL18, Mf_MMP9 and Mf_CX3CR1), perivascular-resident (Mf_LYVE1), redox-related (Mf_MT1G) and hypoxic (Mf_SLC2A1). One cluster had an expression pattern with correlation to both MMP9 and CCR2 phenotypes and was annotated as a mix.

Some of the clusters did not correspond to myeloid cell types including cluster 9, 12 and 15. A differential expression analysis of the genes in each of these clusters was performed to ease their annotation. Cluster 9 had a high expression of CD1A, TACSTD2, HLA-DGB2 and CD207 genes related to DC Langerhans cells, which is an immature DC and associate with tumor promoting features as they stabilize the tumor antigen tolerance and T Cell immunity (84, 85). Looking at the ten genes with highest expression, we observed them to be strongly associated with lymphocytes, indeed, genes such as CD79 which is related to B cells, HLA-DRA and HLA-DPA1 were found overexpressed. Cluster 12 contained very few cells which expressed genes related to lymphocytes (CD3D, CD3E, GZMM, LCK, GSMK and PTPRCAD). By investigating the ten genes with highest expression in cluster 12 it was mostly observed housekeeping genes but also CD74 which is correlated to B cells. Cluster 15 also expressed genes related to lymphocytes such as XCR1, GCSAM, BTLA, CLNK, and IDO1. After inspection, and differential expression analyses cluster 12 and 15 were considered to belong to other cell types than macrophages due to their expression profile and were removed before performing a new clustering. The remaining 33 452 cells are shown in figure 14b with all marker genes used for the annotation. The UMAP of the macrophages and monocytes are shown in figure 14c. The macrophage and monocyte composition in the different baseline samples are shown in figure figure 14d, which showed a high degree of heterogeneity. The most abundant phenotypes are the tumor-associated macrophages

(Mf_CCL18 and Mf_CX3CR1), the redox-related (Mf_MT1G), the perivascular-resident macrophage (Mf_LYVE1) and monocytes (Mono_CD14).

In conclusion, our results show that initial clustering of the macrophage subtypes contains some clusters with expression profiles of lymphocytes and DCs, these clusters were removed to give a subset containing only macrophages. The annotated subtypes correspond well to known marker genes for the macrophage subtypes, and they cover the most important subtypes found in breast cancer.

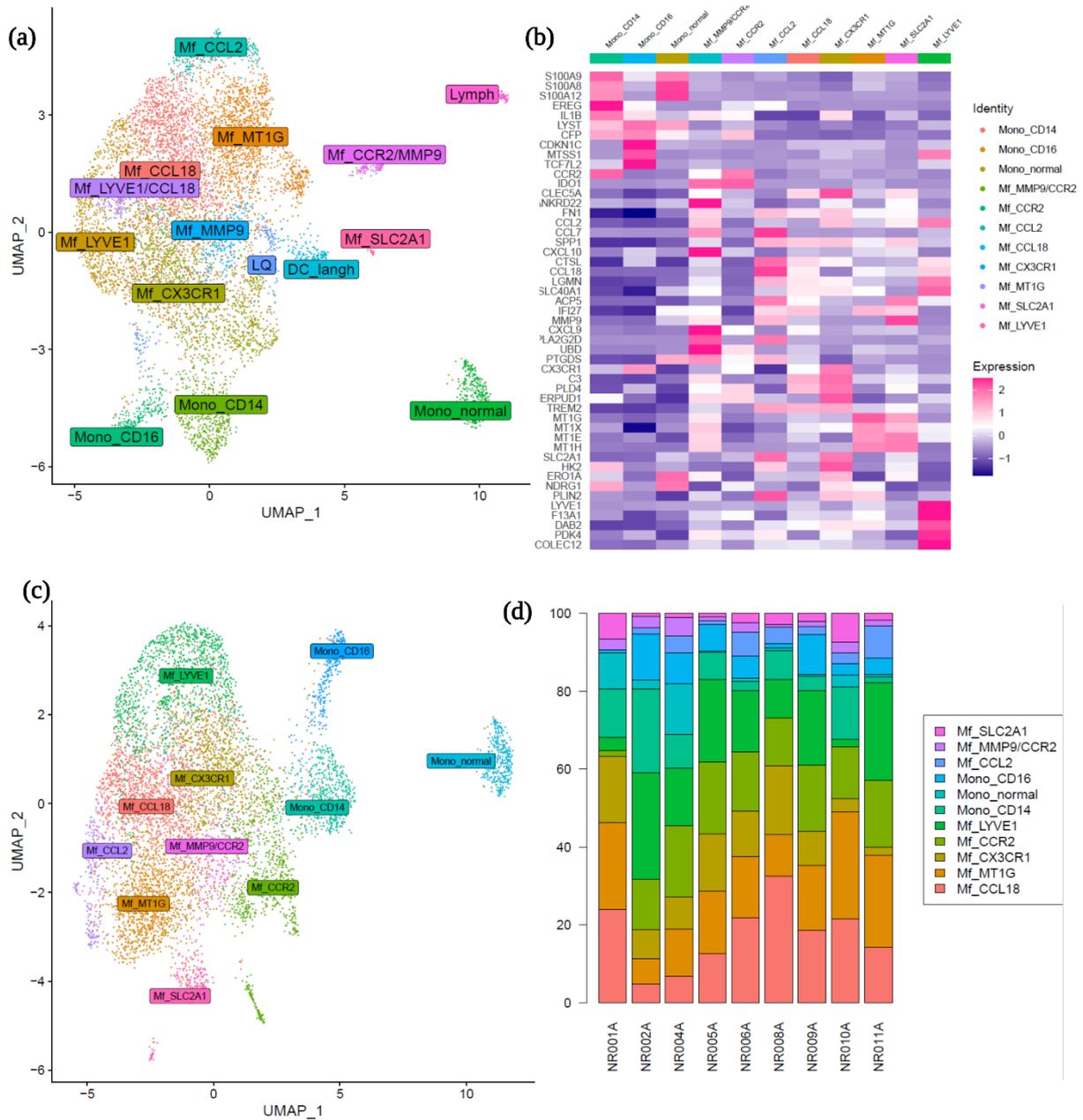


Figure 54. (a) A Annotated cluster of myeloid cells. Three of the clusters represent monocytes with different phenotypes. We also identified eight clusters with different phenotypes of macrophages. It was also observed three clusters annotated as low quality, DC Langhans cells and lymphocytes. (b) Heatmap showing the expression of the marker genes across the eleven cell types of myeloid cell types. (c) The final annotated UMAP with 33 452 cells, after removal of the three clusters annotated as LQ, DC_Langh and Lymph. (d) A barplot representing the tumor myeloid composition of biopsies at baseline, before treatment.

4.4 Fibroblast cluster annotation

The fibroblasts with a total of 23 368 cells were clustered and annotated using the marker genes defined by Wu et al. (83) who identified myoepithelial associated CAFs (myCAF), inflammatory associated CAFs (iCAF), differentiated perivascular-like fibroblast (dPVL) or immature PVLs (iPVL). Further, marker genes established by Qian et al. (70) were screened to provide a deeper annotation of the fibroblast phenotypes. From the myCAF phenotypes cluster myCAF_STAR_NF and myCAF_COMP was identified. From the iCAF phenotypes a cluster with high CFD expression (iCAF_CFD) which is a marker for adipcin led to the identification of adipogenic fibroblast. Figure 15a illustrates some marker genes for the different fibroblast phenotypes, COMP being the marker gene for myCAF_COMP, ITM2A is a marker gene for iCAFs, RERGL is a marker gene for dPVLs and CTSC is a marker gene for iPVLs. The annotation of the different fibroblast phenotypes described above are shown in figure 15b. A heatmap including three marker genes for each fibroblast phenotype is shown in figure 15c. A full heatmap with all marker genes used for the annotation is found in Appendix 6. The fibroblast composition in the different baseline samples is shown in figure 15d. Again, a high heterogeneity in fibroblast composition between the samples was observed at baseline. There were no clear specific fibroblast type more prominent than others.

In conclusion, our results show that the clustering of fibroblasts allows us to subdivide fibroblasts in eight subtypes which fitted well with marker genes previously used.

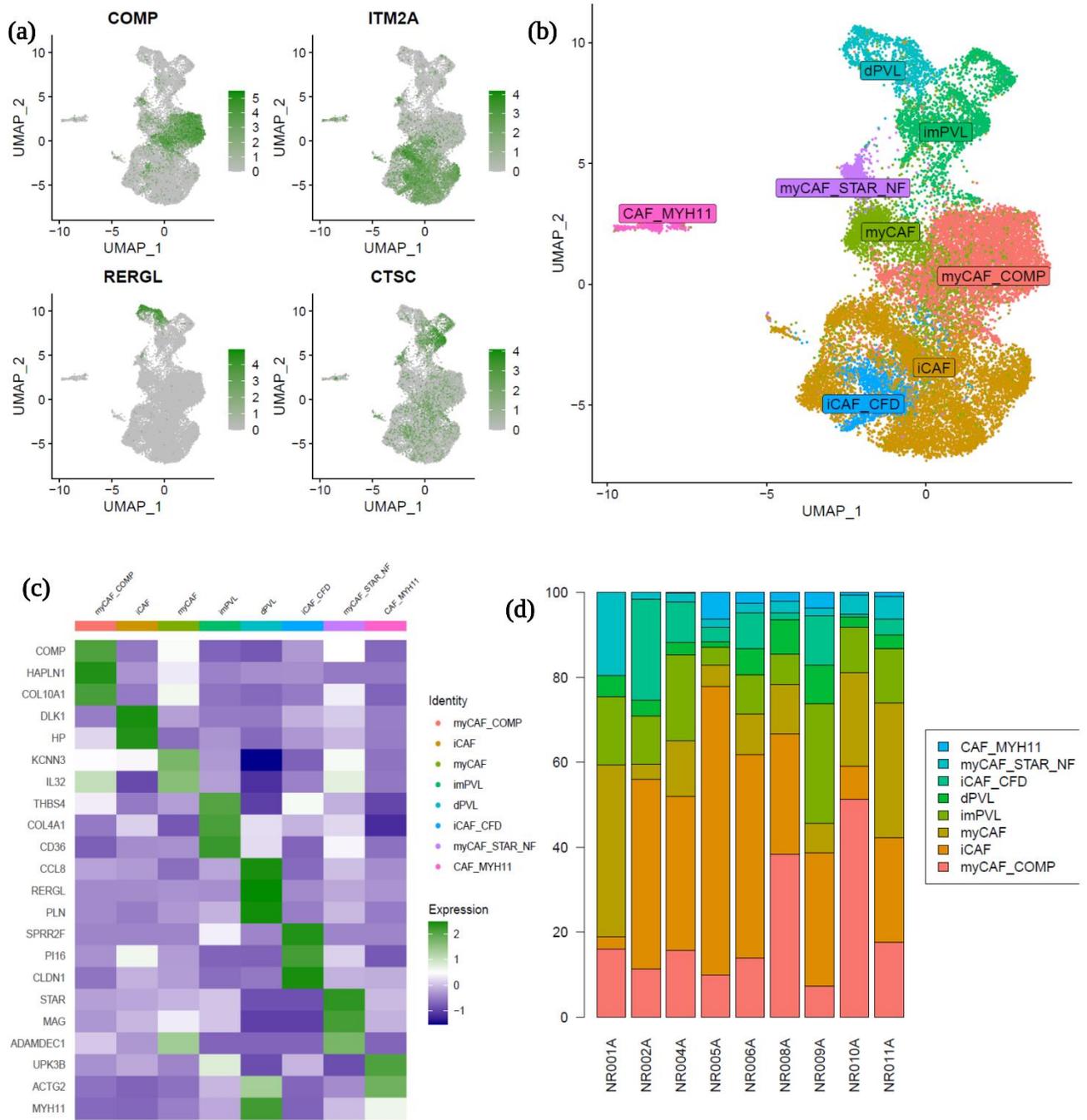


Figure 15. (a) A selection of a few marker genes for specific fibroblast subtypes is used to illustrate the repartition of myCAFs, iCAFs, dPVLs and imPVLs. (b) Annotated UMAP of 23 368 cells from fibroblast phenotypes was divided into three main categories. The first are the PVL phenotypes (dPVL and imPVL), secondly the myCAFs (myCAF_STAR_NF, myCAF and myCAF_COMP), and thirdly the iCAFs (iCAF and iCAF_CFD). A small island with CAF_MYH11 was also annotated. (c) Heatmap showing the expression of three marker genes for each fibroblast phenotype. (d) A barplot representing the tumor fibroblast composition in each patient baseline sample.

4.5 Treatment response

For each T cell, macrophage and fibroblast type identified above, we performed a paired Wilcoxon signed-rank test to assess whether the proportion of a specific type significantly changed between the two timepoints. The statistical test is non-parametric and does not make assumptions about sample size or the population's distribution. The null-hypothesis for this test is that there is no significant change between the paired samples, and the significance level alpha is set to 0,05. A full list of the statistical outcomes for the T cell phenotypes are found in table 2. The medians of each type at the baseline, and at 21 days are also reported. We found one change to be significant for CD4_reg with a p-value of 0,027.

Table 2. Summary statistics for the paired Wilcoxon signed-rank test for the proportion of the different T cell phenotypes. The median from the baseline samples and the second timepoint (after 21 days) are shown as well as the p-value. CD4_reg show significant changes in proportion between the two timepoints with a p-value equal to 0,0273. This is highlighted with bold text.

T cell subtype	Median baseline	Median 21 days	P-value
CD4_reg	4.62	7.88	0.0273
CD8_em	22.8	27.2	0.734
CD4_em	20.8	25.4	0.652
CD8_rm	5.13	7.69	0.57
CD8_n	8.15	6.95	0.359
CD8_emra	12.8	6.36	0.652
NK_cyto	7.26	4.38	0.426
CD4_n	3.33	2.88	0.496
NK_rest	3.60	3.56	1
CD4_ex	0.238	0.859	0.441

A full list of the statistical outcomes for the macrophages are given in table 3, the median at the baseline and after 21 days for the macrophage phenotypes are shown, and the p-value from the paired Wilcoxon signed-rank test. There were four phenotypes from the macrophages with a significant difference in quantity between the two timepoints being Mf_SLC2A1 with a p-value of 0,0078, Mono_CD14 with a p-value of 0,0117, Mf_MT1G with a p-value of 0,0195 and Mf_CCL18 with a p-value of 0,0117.

Table 3. The p-values from the paired Wilcoxon signed-rank test for the proportion of macrophages between the two different time points. Median of the proportions are also shown. The four phenotypes Mf_SLC2A1, Mono_CD14, Mf_MT1G and Mf_CCL18 have a significant difference in quantity between the two timepoints with p-values.

Macrophage subtype	Median baseline	Median 21 days	P-value
Mf_SLC2A1	2.12	0.148	0.00781
Mf_MMP9/CCR2	2.52	2.24	0.91
Mf_CCL2	2.75	1.30	0.91
Mono_CD16	5.59	2.08	0.129
Mono_normal	0.901	0.444	1
Mono_CD14	7.42	2.15	0.0117
Mf_LYVE1	15.9	17.2	0.25
Mf_CCR2	15.1	5.80	0.0547
Mf_CX3CR1	8.90	8.33	0.129
Mf_MT1G	16.0	20.8	0.0195
Mf_CCL18	18.6	26.9	0.0117

A full list of the statistical outcomes for the fibroblast phenotypes are given in table 4.

Including the median at both baseline, after 21 days, and the p-values. None of the fibroblast phenotypes has a significant change of quantity between the timepoints. In conclusion these results show that 5 of the 29 cell types have a significant quantitative change after 21 days of treatment.

Table 4. The p-values from the paired Wilcoxon signed-rank test on the proportion of fibroblasts between the two different time points. The median from baseline and the second timepoint (after 21 days) are shown, as well as the p-value. None of the fibroblasts show a significant change in proportion between the two timepoints.

Fibroblast subtype	Median baseline	Median 21 days	P-value
CAF_MYH11	0.927	0.613	0.129
myCAF_STAR_NF	2.22	2.11	0.164
iCAF_CFD	3.78	4.85	0.82
dPVL	3.79	5.20	0.25
imPVL	11.2	10.8	0.301
myCAF	11.7	13.4	1
myCAF_COMP	15.6	23.3	0.129
iCAF	31.5	19.4	0.652

The five cell subtypes with a significant change are illustrated using boxplots in figure 16. It was observed an increase of CD4_reg in all patients except for one as shown in figure 16a. CD4_reg are associated with an immunosuppressive environment and promotion of tumor progression. There was also observed an increase of the macrophage subtype Mf_MT1G in all patients except for one, this is shown in figure 16b. The macrophage subtype Mf_CCL18 was also upregulated after 21 days of treatment as seen in figure 16c. There is a significant

decrease of the monocyte Mono_CD14 as shown in figure 16d, and the macrophage subtype Mf_SLC2A1 as seen in figure 16e after 21 days of treatment.

In conclusion our results show significant changes of the proportion of five cell subtypes from the microenvironment. There was an upregulation of the cellular subtypes Tregs, Mf_CCL18 and Mf_MT1G which are associated with an immunosuppressive microenvironment. They also show a significant downregulation of the cellular subtypes Mono_CD14 and Mf_SLC2A1.

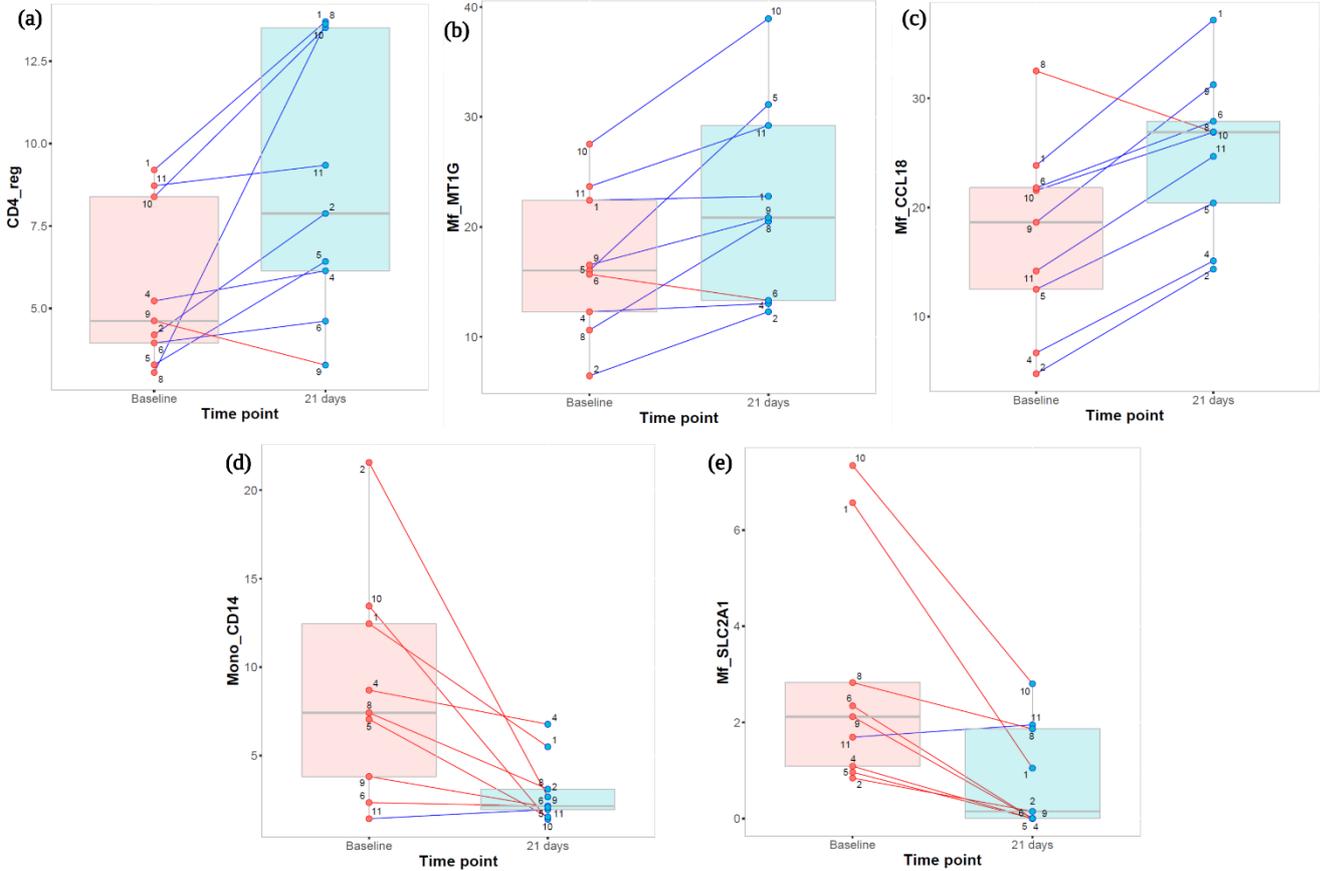


Figure 16. (a) Boxplots representing the proportion of a cell type on the y-axis at two different time points on the x-axis. The line within each box represents the median. Upper and lower edges of each box represent 75th and 25th percentile, respectively. The whiskers represent the lowest datum still within $[1.5 \times (75^{\text{th}} - 25^{\text{th}} \text{ percentile})]$ of the lower quartile, and the highest datum still within $[1.5 \times (75^{\text{th}} \text{ and } 25^{\text{th}} \text{ percentile})]$ of the upper quartile. Each dot represents the proportion for each patient. The blue line across time points show an increase in proportion while the red line shows a decrease in quantity.

5 Discussion and future perspectives

scRNA-seq analyses showed that TME in breast cancer is highly heterogeneous. There is limited understanding on how the cells of the TME evolve and respond to treatment. In this thesis we investigated how the TME of breast cancer patients with ER-positive, HER2-negative lesions changed after 21 days of treatment with ribociclib and letrozole. In the study we found that there was high inter- and intra-heterogeneity and there were five cellular subtypes with a significant proportional change after 21 days of treatment.

5.1 Methodological evaluation

scRNA-seq provides high-resolution analyses of the single cell transcriptomics in complex tissues. It is an invaluable method to characterise cells in tumors and can give a better understanding of the disease development under treatment pressure. In this thesis the annotation of the cells was performed using first clustering with UMAP, then a manual inspection of the clusters obtained, and allocation of cell types to the clusters based on marker genes previously described (70, 82, 83). Marker genes can be expressed in more than one cell type which make the annotation very complex. Selecting good and unique markers for cell types and cellular subtypes is difficult and may lead to inaccurate annotations. In addition, manual annotations are time consuming and requires the analyser to have good knowledge about the cell types, marker genes, and functional genes. To overcome these problems several computational methods have been developed such as **SingleR** and **EcoTyper**. **SingleR** compares the single cell gene expression to reference genomes from pure cell types and thereby assign the cellular identity of the cells (86). **EcoTyper** is a machine learning algorithm and applies statistical methods and deconvolution with the machine learning method **CIBERSORTx** (87). This method can annotate cell states and multicellular microenvironments, as well as detecting novel cellular subtypes (88). The errors which easily occurs when the annotation is performed manually might be reduced by applying computational methods. Furthermore, the combination of manual and computational based annotation may allow better cell type annotation in the future.

When clustering the macrophages, a cluster was identified to be DC Langerhans cells (LCs). This cluster was not considered to belong to the macrophage cell type and was removed. After

more research it is seen that LCs have gene expression profiles similar to both macrophages and DCs (89). These cells have a specific function in the immune system by activation of several immune responses including activation of CTLs, Th17 and the humoral immunity (85). Since they express genes from both DCs and macrophages they may cluster with DCs or macrophages and removing them will lead to a loss of information from LCs, which play a role in the immune response. We may consider keeping them in the macrophage cluster in the future or cluster them with the other DCs.

5.2 Cell types with significant changes in proportion between the two time points

The five cellular subtypes with a significant change were CD4_regs, Mf_MT1G, Mf_CCL18, Mono_CD14 and Mf_SLC2A1. CD4_regs were upregulated in all patients except for one after 21 days of treatment, Tregs are associated with an immunosuppressive environment and promotion of tumor progression. It is reported that patients receiving ribociclib have a decrease of Tregs after 6 months of treatment, which is the opposite of what is seen after 21 days (90). It would be interesting to investigate the proliferation rate of Tregs. Since ribociclib is a CDK4/6 inhibitor and the Tregs have an increase, they might have a low proliferation rate and not be affected that much by ribociclib.

Others performing bulk RNA seq to patients receiving the same treatment have observed the same trend as seen in this thesis, where the tumor seems to be immune cold after 21 days of treatment (personal communication at a conference). Surprisingly when communicating these results to the oncologists recruiting the patients in the clinical trial, it appeared that the patients analysed are actually responding well to treatment with nearly no tumor observed after 6 months of treatment. Further, investigation is needed to understand the biological significance of Tregs. To analyse a tissue sample between 21 days of treatment and 6 months would be of great interest to get a proper understanding of what happens to the cells during the last phases of the neoadjuvant treatment.

The macrophage subtype Mf_MT1G was upregulated in all samples except for one after 21 days of treatment. MT1s are metallothionein's, meaning they are metal-binding proteins, induced by different stimuli such as hormones, cytokines, and growth factors (91). MT1 binds to zinc which can withdraw zinc from the transcription factor protein p53. This will lead to

inactivation and spatial remodelling of p53 and promote cancer cell proliferation. The MT1 proteins are also seen to have a positive correlation with the proliferative marker Ki-67 and high grade breast cancers (92). While in breast cancer MT1G seems to be protumorigenic a study of the effect of MT1G on thyroid cancer by Fu et al. demonstrated a correlation between MT1G and tumor suppressor properties. This was done through suppression of cell growth and invasiveness by modulating the PI3K/Akt pathway, and promotion of cell cycle arrest and apoptosis by inhibition of the Rb/E2F signalling pathway (93). It has also been reported that the concentration of MT1s is most abundant at the G1/S phase transition in the cell cycle, at this stage the cells prepare for DNA synthesis, which show the effect MT1s can have on tumor cell proliferation (94). Therefore, it would be of great interest to investigate where in the cell cycle the cancer cells are after 21 days of treatment, and if ribociclib is sufficient to inhibit the cell cycle and uncontrolled growth. This analysis can be performed on the scRNA-seq dataset obtained in this thesis, by finding the gene names related to the different phases of the cell cycle from a database, and apply the `CellCycleScoring()` function which is a part of the Seurat package (95).

There was also a significant upregulation of the macrophage subtype `Mf_CCL18`. CCL18 is a cytokine produced by macrophages and is associated with promotion of metastasis of breast cancer, and the induction of epithelial-mesenchymal transition of the cancer cells (96). CCL18 is produced mainly by macrophages of the so called M2 phenotype which is the most dominant macrophage in breast cancer (97). It has been reported that CCL18 expressing macrophages recruit circulating naïve CD4⁺ T cells to the TME, and that the T cell receptor profile of the naïve CD4⁺ T cells and Tregs in the TME of breast cancer have a significant overlap. They suggest that most Tregs in the TME of breast cancer originates from naïve CD4⁺ T cells and are recruited as differentiated Tregs (98). Applying a trajectory analysis to investigate how the cells in the TME differentiate can be done by the toolkit `Monocle` (99). This can give a useful insight to get a better understanding of the changes observed in this thesis. The increase of Tregs and CCL18 could be correlated, however patient 9 and patient 8 will go against this.

There was a significant decrease of the monocyte Mono_CD14, this is a classical monocyte with expression of CD14 and S100A8/9, they also express CCLR2 a cytokine with pro-migratory properties (100). The decrease of monocytes is positive for the breast cancer progression and might be due to no new immune cells entering the TME. Indeed, it is assumed that most of the monocytes when reaching tissue will differentiate into macrophages, therefore a decrease in monocyte proportion could be due to few or no new circulating immune cells reaching the tumor site. Which makes the proportion of these cells decrease in the TME. No new immune cells entering the TME also confirms an immunosuppressive environment where immune cells are dead, removed, senescent or exhausted.

The macrophage phenotype Mf_SLC2A1 was observed to have a significant decrease after 21 days of treatment. SLC2A1 encodes glucose transporter 1 (GLUT1) which is found in malignant cells. GLUT1 provides energy to cancer cells by aerobic glycolysis, inhibition of GLUT1 is therefore associated with anti-proliferative effects on cancer (101). This might be an indication of a reduced metabolism in the cancer cells after 21 days of treatment. If the cancer cells lose access to energy, will they die and stop proliferate. Making them more manageable to be removed by surgery.

Why there is an upregulation and downregulation of the abovementioned cellular subtypes could be investigated further, some of the cells can have a correlation in their proportional change during treatment. There seem to be connections between several of them and it would be helpful to make a heatmap with the correlation between the different cellular subtypes.

scRNA-seq provides a better understanding of the TME as it gives information regarding all the cells in the TME but as the cells are dissociated the spatial context is lost, the dissociation of the tissue can also induce stress to the cell's and change their expression profile in some degree. Investigation of the interactions between the cells and their location within the TME can provide a better understanding of the treatment response, tumor evolution and how specific cells interact with each other. By performing spatial transcriptomics more information about what and how the cellular subtypes are recruited to the TME, and genes that are important for cellular interactions can be obtained (102, 103).

6 Conclusion

We observed an immunosuppressive environment in the TME after 21 days of treatment. This is contrary to what has been observed by others after 6 months of treatment. It remains to be established how the changes we observe in the immune TME relate to patients' response. To assess whether the immunosuppressive TME persist across treatment, it would therefore be of great interest to perform scRNA-seq on another timepoint between 21 days and 6 months of treatment. It will also provide a better insight to the changes in the TME and the response to treatment with the analysis of the last timepoint, after 6 months. Spatial transcriptomics from the different timepoints will also give a better understanding of the cellular interactions, recruitment, and position of the cellular subtypes in the tumor, and would be an interesting method to perform on the samples in the future. There are several bioinformatic analyses that can be performed on the obtained dataset from this thesis which can provide a deeper insight into the changes observed and changes in the tumor ecosystems over time. Trajectory inference analyses and cell cycle scoring are interesting parts to better understand the effects of the combination of aromatase and CDK4/6 inhibitors. An important part of the work of this thesis was identifying specific and specialized immune cell types, it would also be of great interest to compare the manual annotation performed here with a machine-driven annotation of both cells and clusters, to see if we could validate and/or improve our annotations.

Appendix 1; Solutions

Dissociation solution:

Reagents	Volume per tissue sample
DMEM	4.4 mL
Collagenase P	50 mg/mL
DNase I	1 mg/mL

STEM Buffer 2:

Reagents	Volume	Final concentration
PBS	48.4 mL	
FBS/FCS	1 mL	2%
0.5M EDTA (pH 8)	600 μ L	

PBS+BSA 0.04%:

Reagents	Volume	Final concentration
PBS	19.84 mL	
BSA (50mg/mL stock)	160 μ L	0.04%

Appendix 2; Elbowplot

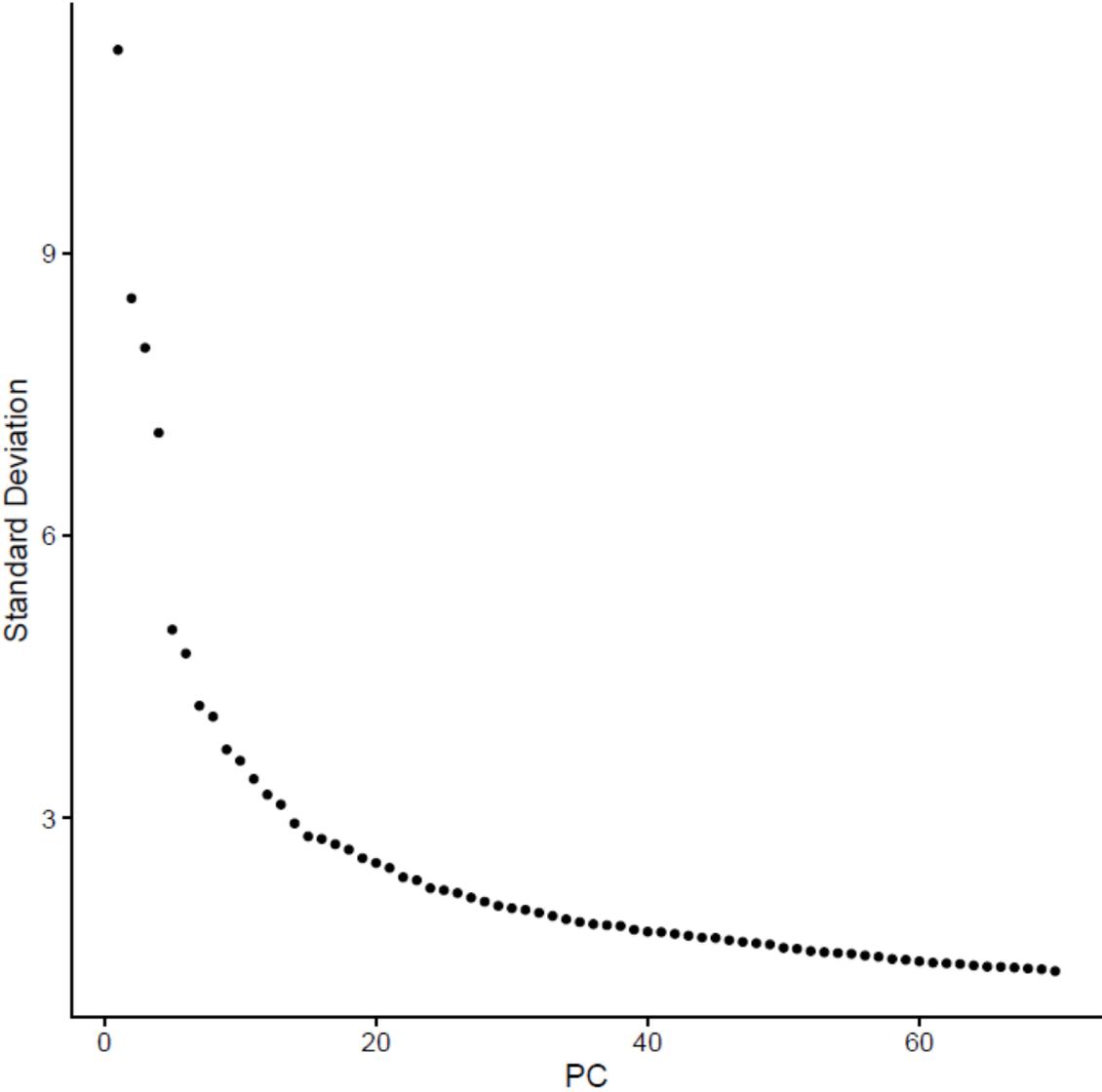


Figure 6. Elbowplot used to decide how many PCs to use for further analysis.

Appendix 3; Example codes

Example code for Scaling

```
setwd ("/cluster/projects/n9751k/Karoline/NR")

library(Seurat)
library(plyr)
library(dplyr)
library(future)
plan("multicore", workers=10)
options(future.globals.maxSize = 100*1024^3)

load("Merged_NeoLetRib_algos_08_03_22_full.RData")

#####
#### Subset Merged1 for 9 doublets ####
#####

Merged_subset <- subset(Merged1, orig.ident %in% c(
  "NR001A", "NR001B",
  "NR002A", "NR002B",
  "NR004A", "NR004B",
  "NR005A", "NR005B",
  "NR006A", "NR006B",
  "NR008A", "NR008B",
  "NR009A", "NR009B",
  "NR010A", "NR010B",
  "NR011A", "NR011B"))

#####
#### Normalization ####
#####

Merged <- NormalizeData(Merged_subset, normalization.method =
"LogNormalize", scale.factor = 10000)

#####
#### Scale and regress out variables ####
#####
lapply(Merged@meta.data, class)
all.genes <- rownames(Merged)
Scaled <- ScaleData(Merged,
                    features = all.genes,
                    vars.to.regress = c("percent.mt", "nCount_RNA",
"nFeature_RNA"), split.by="orig.ident")
load ('common_genes_long.RData')
```

```
#####  
##### Save as new RData #####  
#####
```

```
library (gdata)  
keep(Scaled, common, sure=TRUE)  
save.image("Scaled_all.RData")
```

Example code for Jackstraw

```
setwd ("/cluster/projects/nn9751k/Karoline/NR")
```

```
library(Seurat)  
library(plyr)  
library(dplyr)
```

```
load ("Scaled_all.RData")
```

```
#load ("common_NR.RData")  
#load ("common_genes_long.RData")
```

```
variable <- FindVariableFeatures(Scaled, selection.method = "vst",  
nfeatures = round(dim (Scaled)[1]*35/100))  
variable_feature <- VariableFeatures(variable)  
feature <- unique (variable_feature)  
print(length (variable_feature))  
print (length (feature))  
print (dim (Scaled))
```

```
#####  
#####  
##### Inspecting the PCA / metagene significance to set the number of  
component used to run UMAP #####  
#####  
#####
```

```
#Scaled <- RunPCA(Scaled, features = common, npcs = 50, ndims.print = 1:5,  
nfeatures.print = 30, weight.by.var=F)  
Scaled <- RunPCA(Scaled, features = feature, npcs = 90, ndims.print = 1:5,  
nfeatures.print = 30, weight.by.var=T)  
#Scaled <- RunPCA(Scaled, features = feature, npcs = 60, ndims.print = 1:5,  
nfeatures.print = 30, weight.by.var=T)
```

```

Scaled <- JackStraw(Scaled, num.replicate = 100, dims = 80)
Scaled <- ScoreJackStraw(Scaled, dims = 1:80)

## show the genes driving the different PC
pdf(file="NR_Scaled_inspect_PCs.pdf", paper="a4", onefile=TRUE)
VizDimLoadings(Scaled, dims = 1:6, reduction = "pca", nfeatures=15)
VizDimLoadings(Scaled, dims = 7:12, reduction = "pca", nfeatures=15)
DimHeatmap(Scaled, dims = 1:20, cells = 1000, balanced = TRUE)
DimHeatmap(Scaled, dims = 21:40, cells = 1000, balanced = TRUE)
DimHeatmap(Scaled, dims = 41:60, cells = 1000, balanced = TRUE)
DimHeatmap(Scaled, dims = 61:80, cells = 1000, balanced = TRUE)
JackStrawPlot(Scaled, dims = 1:80)
ElbowPlot(Scaled, ndims = 80, reduction = "pca")
dev.off()

#####
##### Loop UMAP through 15 to 35 components #####
#####

pdf(file="NR_Scaled_loop_NPC.pdf", paper="a4", onefile=TRUE)
for (i in c(80:90)){
  NPC=i
  Scaled <- RunUMAP(Scaled, dims = 1:NPC, verbose = FALSE)
  Scaled <- FindNeighbors(Scaled, dims = 1:NPC, verbose = FALSE)
  Scaled <- FindClusters(Scaled, verbose = FALSE)
  plot(DimPlot(Scaled, label = F, group.by='orig.ident'))
  print (i)
}
dev.off()

#####
#####
##### Choose the Number of principal component and the resolution for
clusters #####
#####
#####
NPC=70
Scaled <- FindNeighbors(Scaled, reduction = "pca", dims = 1:NPC, k.param =
30, compute.SNN = TRUE, nn.method = "rann", nn.eps = 0)
Scaled <- FindClusters(Scaled, resolution = c(0.8,1,1.2))
Idents(Scaled)

resol=0.8
pdf(file="NR_Scaled_NPC70_loop_resol.pdf", paper="a4", onefile=TRUE)
for (i in c(1:2)){
  Seurat <- SetIdent(object = Scaled, value = paste('RNA_snn_res', resol,
sep='.'))
  Seurat <- RunUMAP(Seurat, dims = 1:NPC)
  plot(UMAPPlot(Seurat, reduction = "umap", pt.size=0.25,label=T))
  print (resol)
  resol= resol + 0.2
}
dev.off()

```

```

resol = 'RNA_snn_res.1.2'
Scaled <- SetIdent(object = Scaled, value = resol)
Scaled <- RunUMAP(Scaled, dims = 1:NPC)
#plot(UMAPPlot(Scaled, reduction = "umap", pt.size=0.25,label=T))

library (gdata)
keep(Scaled, common, sure=TRUE)
save.image("Scaled_NR_all.RData")

#####
##### PDF files #####
#####

pdf("clusters_orig_ident.pdf")
DimPlot(Scaled, label = F, group.by='orig.ident')
dev.off()

pdf("clusters.pdf")
DimPlot(Scaled, label = TRUE, group.by='RNA_snn_res.1.2')
dev.off()
print (table (Scaled@meta.data$RNA_snn_res.1.2))

pdf("feature Tcells.pdf")
#Tcell
FeaturePlot(Scaled, features = c("CD3D", "CD3E", "IL7R"), order = T)
#cytotoxic Tcell
FeaturePlot(Scaled, features = c("CD8A", "GZMK", "GZMA"), order = T)
#Treg
FeaturePlot(Scaled, features = c("FOXP3", "TIGIT", "KLRC1", "CTLA4"), order = T)
dev.off()

pdf("feature Macrophages.pdf")
#Macrophages
FeaturePlot(Scaled, features = c("CD68", "LYZ", "CD14", "AIF1"), order = T)
#Macrophages-proinf
FeaturePlot(Scaled, features = c("TNF", "CCL4", "IL1B"), order = T)
dev.off()

pdf("feature Bcells.pdf")
#Plasma cell
FeaturePlot(Scaled, features = c("IGHG1", "CD38", "MZB1", "IGHA2"), order = T)
#Bcell
FeaturePlot(Scaled, features = c("CD79A", "CD19", "MS4A1", "CD83"), order = T)
dev.off()

pdf("feature Fibroblast.pdf")
#Fibroblast)

```

```

FeaturePlot(Scaled, features = c("COL1A1", "C1R", "SERPINE1", "MYL9"), order =
T)
#Fibroblast-MYH11
FeaturePlot(Scaled, features = c("MYH11", "RGS5", "NOTCH3", "TFPI"), order =
T)
dev.off()

pdf("feature Cancer.pdf")
#Cancer cell
FeaturePlot(Scaled, features = c("EPCAM", "FOXA1", "ESR1", "S100A1"), order =
T)
FeaturePlot(Scaled, features = c("ERBB2", "MKI67", "CDK4", "HES4"), order =
T)
dev.off()

pdf("feature Mast- dendritic.pdf")
#Mast cell
FeaturePlot(Scaled, features = c("CPA3", "TPSAB1", "TPSB2", "GATA2"), order =
T)
#Dendritic cell
FeaturePlot(Scaled, features = c("LILRA4", "PLAC8", "IRF7"), order = T)
dev.off()

pdf("feature Endothelial.pdf")
#Endothelial cells
FeaturePlot(Scaled, features = c("CA4", "CD36", "VWF", "ACKR1"), order = T)
dev.off()

```

Example code for UMAP, Heatmap and Barplot

```

setwd ("/cluster/projects/nn9751k/Karoline/NR/T_cells")
load("Final_Annotated_T-cell_clusters.RData")

pdf('Final_Annotated_T_cells.pdf')
DimPlot(Scaled, reduction = "umap", label = TRUE, pt.size = 0.2,
label.size=5, label.box= T) + NoLegend()
dev.off()

Idents(Scaled) <- "main.ids"
cluster.averages <- AverageExpression(Scaled, return.seurat=TRUE)

T_cell_marker_genes <- c('LEF1', 'TCF7', 'SELL', 'IL7R', 'CD40LG', 'ANXA1',
'FOS', 'JUN', 'FOXP3', 'SAT1', 'IL2RA', 'CTLA4', 'PDCD1', 'CXCL13',
'CD200', 'TNFRSF18',
'CCR7', 'NELL2', 'CD55', 'KLF2', 'TOB1',
'ZNF683', 'CCL5', 'GZMK', 'EOMES', 'ITM2C', 'CX3CR1', 'GNLY', 'GZMH',
'GZMB', 'LAG3', 'CCL4L2', 'FCGR3A', 'FGFBP2', 'TYROBP',

```

```

'AREG', 'XCL1', 'KLRC1', 'TRDV2', 'TRGV9',
'MTRNR2L8', 'KLRD1', 'TRDV1', 'KLRC3', 'CTSW', 'CD7', 'MKI67', 'STMN1',
'TUBA1B', 'HIST1H4C')

my_levels <- c('CD4_n', 'CD4_em', 'CD4_reg', 'CD4_ex', 'CD8_n', 'CD8_rm',
'CD8_em', 'CD8_emra', 'NK_cyto', 'NK_rest')
levels(cluster.averages) <- my_levels

pdf('Heatmap_T_cells.pdf')
par(mar = c(2,2,5,2))
DoHeatmap(cluster.averages, features = T_cell_marker_genes, draw.lines = F,
size = 2.5,
group.colors = c('#00B4F0', '#DE8C00', '#00C08B', '#C77CFF',
'#7CAE00', '#B79F00', '#F8766D', '#00BA38', '#00BFC4', '#619CFF')) +
scale_fill_gradientn(colors = c("darkblue", "white", "darkorange2"))
theme(axis.text.x=element_text(size=2), axis.text.y=element_text(size=5))
dev.off()

ggplotColours <- function(n = 6, h = c(0, 360) + 15){
if ((diff(h) %% 360) < 1) h[2] <- h[2] - 360/n
hcl(h = (seq(h[1], h[2], length = n)), c = 100, l = 65)
}

color_list <- ggplotColours(n=12)
print(color_list)

#####Barplot#####

dta <- as.data.frame.matrix(prop.table(table (Scaled@meta.data$orig.ident,
Scaled@meta.data$main.ids),1)*100)
dta_bas <- dta [substr (rownames (dta),6,6)=="A", ]

pdf ('test_barplot.pdf')
par(mar = c(5, 5, 5, 8))
barplot(t(dta_bas), col = color_list, legend.text = TRUE , args.legend =
list(x = "right", inset = c(- 0.3, 0)), las=2)
dev.off()

```

Appendix 4; Dotplot with all marker genes

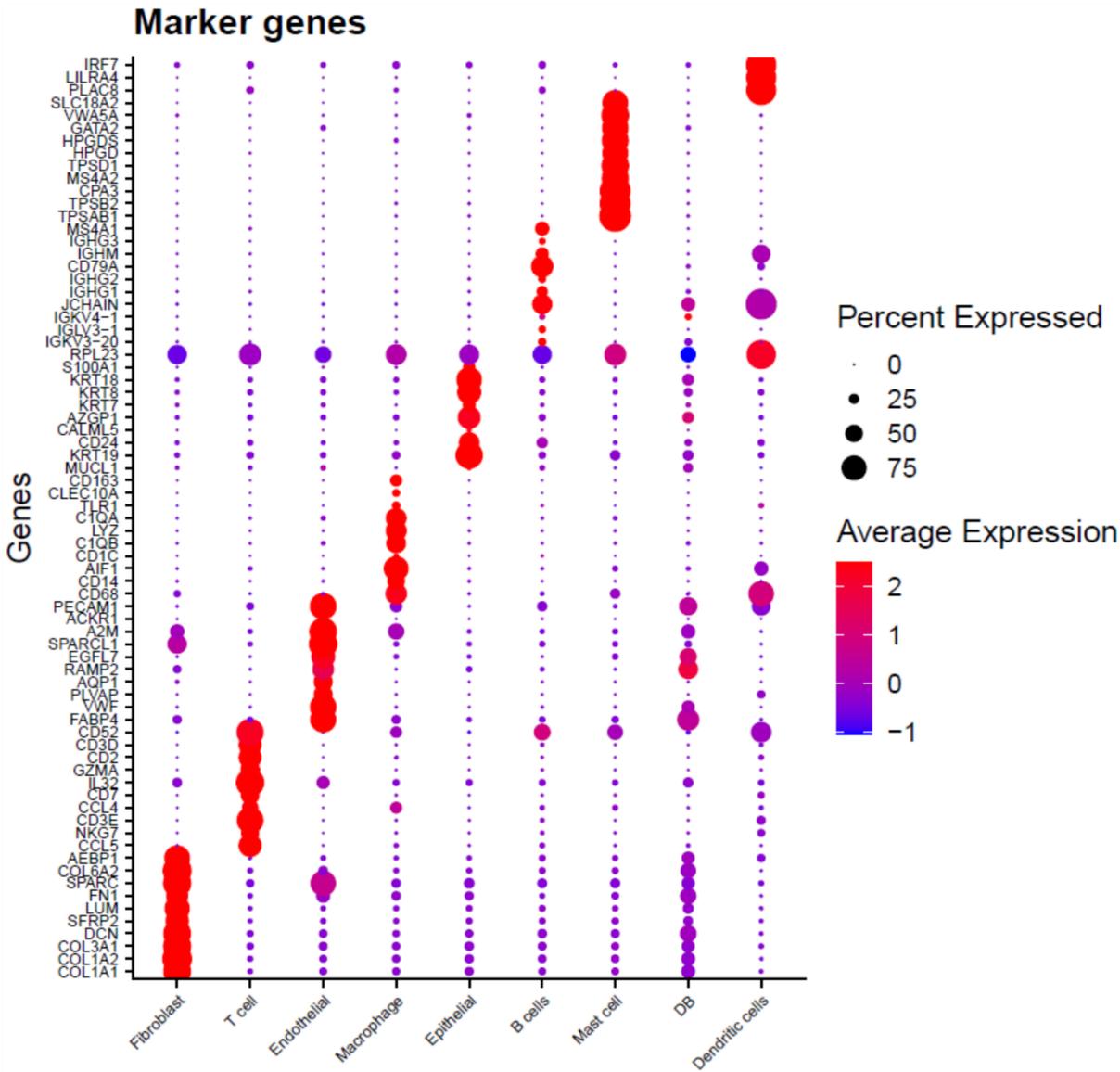


Figure 7. Dotplot with all marker genes for the cell types used to annotate the clusters.

Appendix 5; UMAP and heatmap for NK and T cells

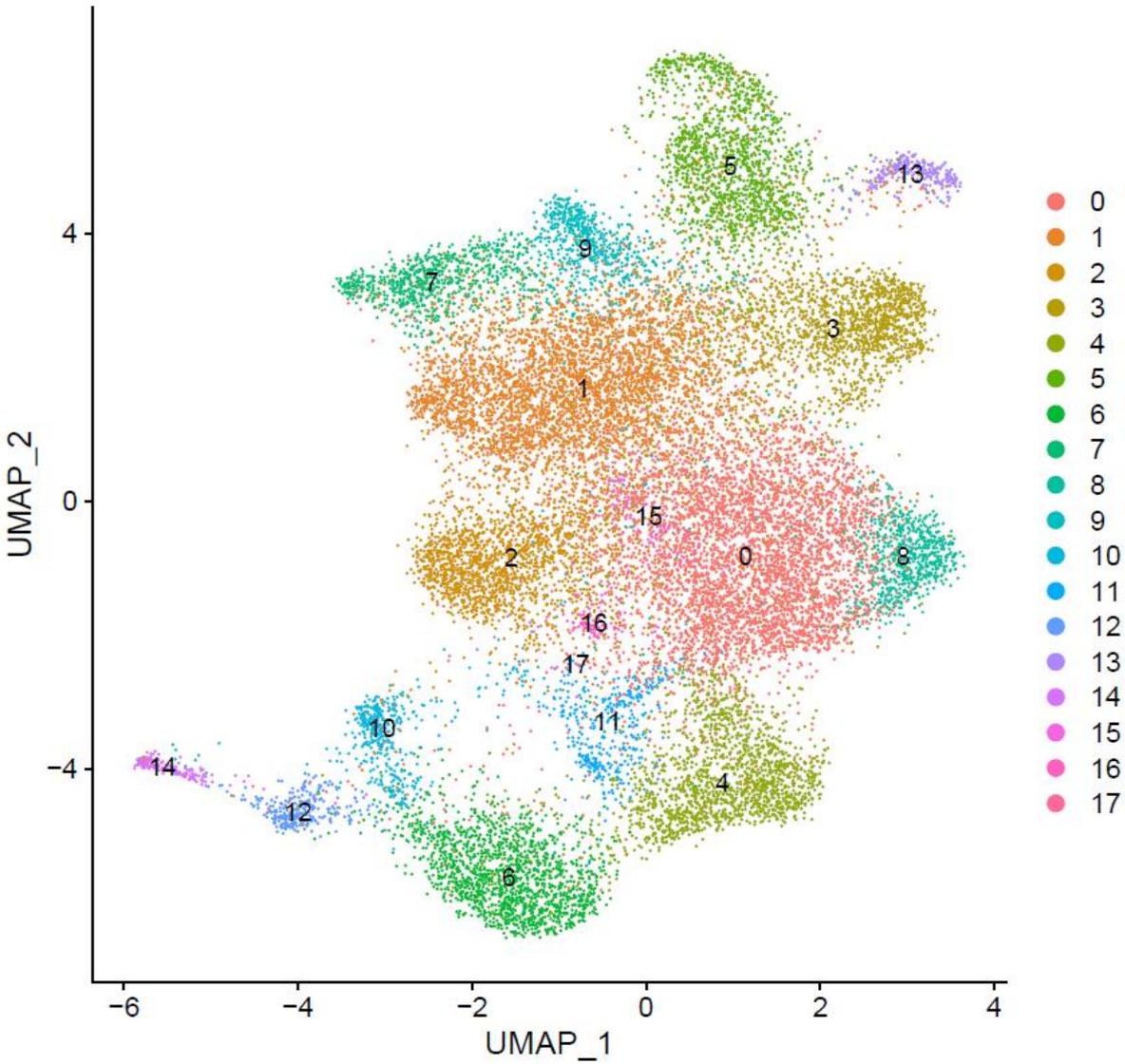


Figure 8. UMAP of the NK and T cells after clustering. There was 18 clusters for this clustering.

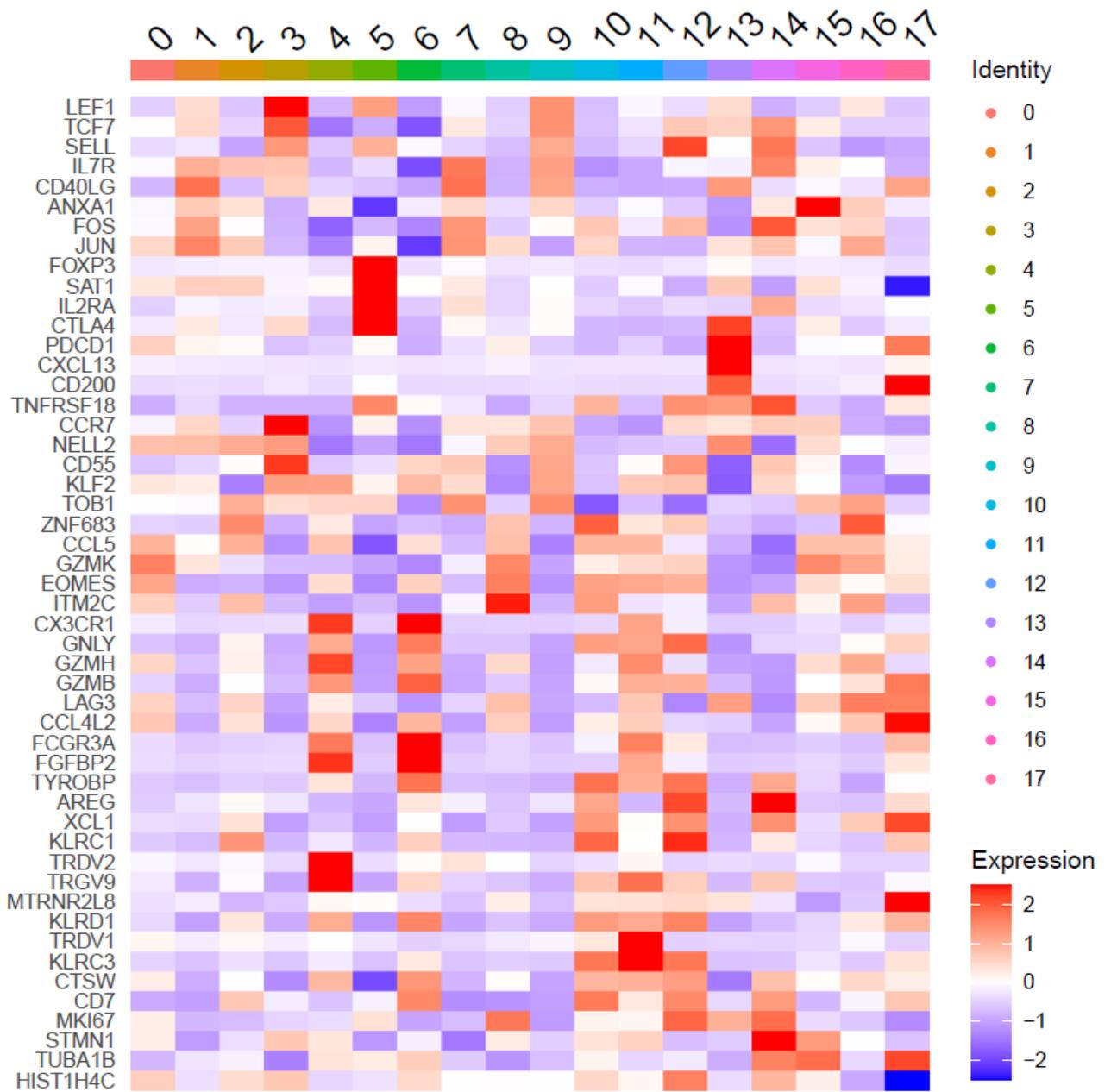


Figure 9. Heatmap with all marker genes (70, 82) used to annotate the clusters for NK and T cells.

Appendix 6; Heatmaps for Fibroblast annotation

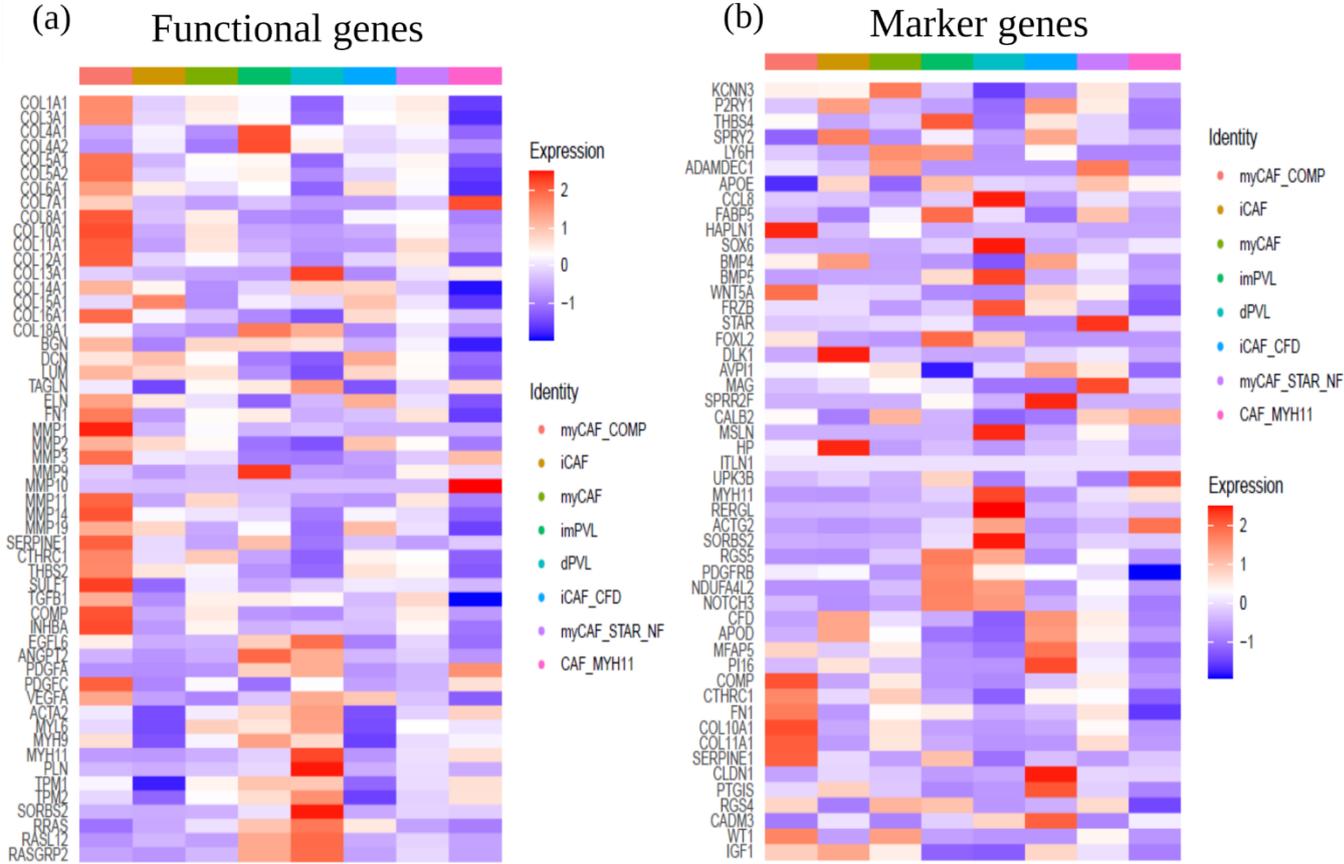


Figure 10. (a) Heatmap with all functional genes (70) used to annotate the fibroblast clusters. (b) Heatmap with all marker genes (70) used to annotate the fibroblast clusters.

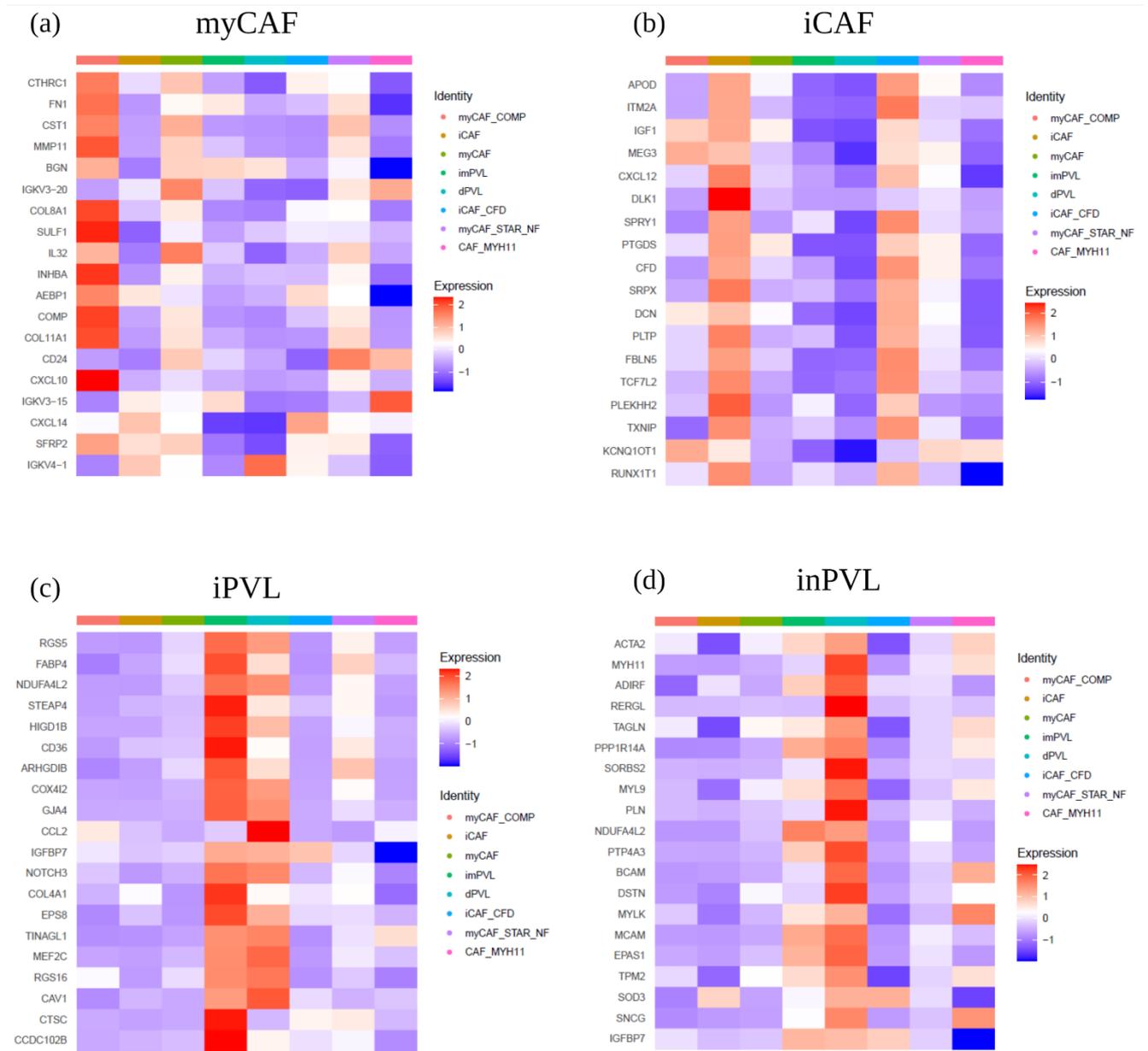


Figure 11. (a) Heatmap with marker genes (83) used to identify myCAFs. (b) Heatmap with marker genes used to annotate iCAFs. (c) Heatmap with marker genes used to annotate iPVLs. (d) Heatmap with marker genes to annotate inPVLs.

Appendix 7; Supplies

Supplies	Product name	Company	Catalogue number
1.5 ml Eppendorf tube			
1.5 ml low-bind tube			
15 ml falcon tube			
40 μ M filter	EASYstrainer	Greiner Bio-One	542040
70 μ M filter	EASYstrainer	Greiner Bio-One	542070
NecleoCounter	NucleoCounter [®] NC-100 ⁻		N/A
Centrifuge			N/A
Magnet	EasySep [™] Magnet	Stemcell technologies	N/A
Polystyrene tube	Falcon [™] Round-Bottom Polystyrene Test Tubes	fisher scientific	10100151

Appendix 8; Reagents

Reagent	Product name	Company	Catalogue number
BSA	UltraPure™ Bovine Serum Albumin	Sigma	AM2616
Collagenase P	Collagenase P	Roche	11213857001
DNase I	Deoxyribonuclease I from bovine pancreas	Merck	DN25-10MG
DMEM	Dulbecco's Modified Eagle's Medium/Nutrient Mixture F-12 Ham	Merck	D-6434-500ML
EDTA (0.5M), pH8	EDTA (0.5 M), pH 8.0, RNase-free	Thermo fisher	AM9260G
PBS	Dulbecco's Phosphate Buffered Saline	Merck	D8537-500ML
RBC	Red Blood Cell Lysing Buffer Hybri-Max™	Sigma	R7757-100M
10x Single Cell kit	10x Chromium Single Cell 2' kit	10x Genomics	N/A
70-100% Alcohol			

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. 2021. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians* 71:209-249.
2. Loibl S, Poortmans P, Morrow M, Denkert C, Curigliano G. 2021. Breast cancer. *Lancet* 397:1750-1769.
3. Dibden A OJ, Duffy S.W, Gabe R. 2020. Worldwide Review and Meta-Analysis of Cohort Studies Measuring the Effect of Mammography Screening Programmes on Incidence-Based Breast Cancer Mortality. *Cancer (Basel)* 12:976.
4. Norway CRo. 2021. Cancer in Norway 2020 - Cancer incidence, mortality, survival and prevalence in Norway. Cancer Registry of Norway, Oslo.
5. Sebuødegård S BE, Hofvind S. 2020. Breast Cancer Mortality After Implementation of Organized Population-Based Breast Cancer Screening in Norway. *J Natl Cancer Inst* 112:839-846.
6. Javed A, Lteif A. 2013. Development of the human breast. *Semin Plast Surg* 27:5-12.
7. McNally S, Stein T. 2017. Overview of Mammary Gland Development: A Comparison of Mouse and Human. *Methods Mol Biol* 1501:1-17.
8. BioRender. <https://app.biorender.com/>. Accessed
9. Børresen-Dale AL, Sørli T, Kristensen VN. 2010. On the molecular biology of breast cancer. *Mol Oncol* 4:171-3.
10. Hanahan D, Weinberg RA. 2000. The hallmarks of cancer. *Cell* 100:57-70.
11. Hanahan D, Weinberg RA. 2011. Hallmarks of cancer: the next generation. *Cell* 144:646-74.
12. Hanahan D. 2022. Hallmarks of Cancer: New Dimensions. *Cancer Discov* 12:31-46.
13. Hwang HJ, Lee YR, Kang D, Lee HC, Seo HR, Ryu JK, Kim YN, Ko YG, Park HJ, Lee JS. 2020. Endothelial cells under therapy-induced senescence secrete CXCL11, which increases aggressiveness of breast cancer cells. *Cancer Lett* 490:100-110.
14. Wang D, Xiao F, Feng Z, Li M, Kong L, Huang L, Wei Yg, Li H, Liu F, Zhang H, Zhang W. 2020. Sunitinib facilitates metastatic breast cancer spreading by inducing endothelial cell senescence. *Breast Cancer Research* 22:103.
15. Nejman D, Livyatan I, Fuks G, Gavert N, Zwang Y, Geller LT, Rotter-Maskowitz A, Weiser R, Mallel G, Gigi E, Meltser A, Douglas GM, Kamer I, Gopalakrishnan V, Dadosh T, Levin-Zaidman S, Avnet S, Atlan T, Cooper ZA, Arora R, Cogdill AP, Khan MAW, Ologun G, Bussi Y, Weinberger A, Lotan-Pompan M, Golani O, Perry G, Rokah M, Bahar-Shany K, Rozeman EA, Blank CU, Ronai A, Shaoul R, Amit A, Dorfman T, Kremer R, Cohen ZR, Harnof S, Siegal T, Yehuda-Shnaidman E, Gal-Yam EN, Shapira H, Baldini N, Langille MGI, Ben-Nun A, Kaufman B, Nissan A, Golan T, Dadiani M, et al. 2020. The human tumor microbiome is composed of tumor type-specific intracellular bacteria. *Science* 368:973-980.
16. Byler S, Goldgar S, Heerboth S, Leary M, Housman G, Moulton K, Sarkar S. 2014. Genetic and epigenetic aspects of breast cancer progression and therapy. *Anticancer Res* 34:1071-7.
17. Cichon MA, Degnim AC, Visscher DW, Radisky DC. 2010. Microenvironmental influences that drive progression from benign breast disease to invasive breast cancer. *J Mammary Gland Biol Neoplasia* 15:389-97.
18. Yoder BJ, Wilkinson EJ, Massoll NA. 2007. Molecular and morphologic distinctions between infiltrating ductal and lobular carcinoma of the breast. *Breast J* 13:172-9.

19. Mueller C, Haymond A, Davis JB, Williams A, Espina V. 2018. Protein biomarkers for subtyping breast cancer and implications for future research. *Expert Rev Proteomics* 15:131-152.
20. Rakha EA, Reis-Filho JS, Baehner F, Dabbs DJ, Decker T, Eusebi V, Fox SB, Ichihara S, Jacquemier J, Lakhani SR, Palacios J, Richardson AL, Schnitt SJ, Schmitt FC, Tan PH, Tse GM, Badve S, Ellis IO. 2010. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res* 12:207.
21. Giuliano AE, Connolly JL, Edge SB, Mittendorf EA, Rugo HS, Solin LJ, Weaver DL, Winchester DJ, Hortobagyi GN. 2017. Breast Cancer-Major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin* 67:290-303.
22. Perou CM, Sørli T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO, Botstein D. 2000. Molecular portraits of human breast tumours. *Nature* 406:747-52.
23. Sørli T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Lønning PE, Børresen-Dale AL. 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 98:10869-74.
24. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, Palazzo J, Marron JS, Nobel AB, Mardis E, Nielsen TO, Ellis MJ, Perou CM, Bernard PS. 2009. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 27:1160-7.
25. Ohnstad HO, Borgen E, Falk RS, Lien TG, Aaserud M, Sveli MAT, Kyte JA, Kristensen VN, Geitvik GA, Schlichting E, Wist EA, Sørli T, Russnes HG, Naume B. 2017. Prognostic value of PAM50 and risk of recurrence score in patients with early-stage breast cancer with long-term follow-up. *Breast Cancer Res* 19:120.
26. Dai X, Xiang L, Li T, Bai Z. 2016. Cancer Hallmarks, Biomarkers and Breast Cancer Molecular Subtypes. *J Cancer* 7:1281-94.
27. Schnitt SJ. 2010. Classification and prognosis of invasive breast cancer: from morphology to molecular taxonomy. *Mod Pathol* 23 Suppl 2:S60-4.
28. Burguin A, Diorio C, Durocher F. 2021. Breast Cancer Treatments: Updates and New Challenges. *J Pers Med* 11.
29. Fisusi FA, Akala EO. 2019. Drug Combinations in Breast Cancer Therapy. *Pharm Nanotechnol* 7:3-23.
30. Pedersen RN, Esen B, Mellekjær L, Christiansen P, Ejlersen B, Lash TL, Nørgaard M, Cronin-Fenton D. 2021. The Incidence of Breast Cancer Recurrence 10-32 Years after Primary Diagnosis. *J Natl Cancer Inst* doi:10.1093/jnci/djab202.
31. Begg AC, Stewart FA, Vens C. 2011. Strategies to improve radiotherapy with targeted drugs. *Nat Rev Cancer* 11:239-53.
32. Taylor CW, Kirby AM. 2015. Cardiac Side-effects From Breast Cancer Radiotherapy. *Clin Oncol (R Coll Radiol)* 27:621-9.
33. Jarosz-Biej M, Smolarczyk R, Cichoń T, Kułach N. 2019. Tumor Microenvironment as A "Game Changer" in Cancer Radiotherapy. *Int J Mol Sci* 20.
34. Dickens E, Ahmed S. 2018. Principles of cancer treatment by chemotherapy. *Surgery (Oxford)* 36:134-138.
35. Rasmussen BB, Regan MM, Lykkesfeldt AE, Dell'Orto P, Del Curto B, Henriksen KL, Mastropasqua MG, Price KN, Méry E, Lacroix-Triki M, Braye S, Altermatt HJ, Gelber RD, Castiglione-Gertsch M, Goldhirsch A, Gusterson BA, Thürlimann B,

- Coates AS, Viale G. 2008. Adjuvant letrozole versus tamoxifen according to centrally-assessed ERBB2 status for postmenopausal women with endocrine-responsive early breast cancer: supplementary results from the BIG 1-98 randomised trial. *Lancet Oncol* 9:23-8.
36. Miller WR, Larionov AA. 2012. Understanding the mechanisms of aromatase inhibitor resistance. *Breast Cancer Res* 14:201.
 37. Schettini F, Giuliano M, Giudici F, Conte B, De Placido P, Venturini S, Rognoni C, Di Leo A, Locci M, Jerusalem G, Del Mastro L, Puglisi F, Conte P, De Laurentiis M, Puzstai L, Rimawi MF, Schiff R, Arpino G, De Placido S, Prat A, Generali D. 2021. Endocrine-Based Treatments in Clinically-Relevant Subgroups of Hormone Receptor-Positive/HER2-Negative Metastatic Breast Cancer: Systematic Review and Meta-Analysis. *Cancers (Basel)* 13.
 38. Hortobagyi GN, Stemmer SM, Burris HA, Yap YS, Sonke GS, Paluch-Shimon S, Campone M, Petrakova K, Blackwell KL, Winer EP, Janni W, Verma S, Conte P, Arteaga CL, Cameron DA, Mondal S, Su F, Miller M, Elmeliegy M, Germa C, O'Shaughnessy J. 2018. Updated results from MONALEESA-2, a phase III trial of first-line ribociclib plus letrozole versus placebo plus letrozole in hormone receptor-positive, HER2-negative advanced breast cancer. *Ann Oncol* 29:1541-1547.
 39. Dunbier AK, Ghazoui Z, Anderson H, Salter J, Nerurkar A, Osin P, A'Hern R, Miller WR, Smith IE, Dowsett M. 2013. Molecular profiling of aromatase inhibitor-treated postmenopausal breast tumors identifies immune-related correlates of resistance. *Clin Cancer Res* 19:2775-86.
 40. Patel HK, Bihani T. 2018. Selective estrogen receptor modulators (SERMs) and selective estrogen receptor degraders (SERDs) in cancer treatment. *Pharmacol Ther* 186:1-24.
 41. Murphy CG. 2019. The Role of CDK4/6 Inhibitors in Breast Cancer. *Curr Treat Options Oncol* 20:52.
 42. Lim JS, Turner NC, Yap TA. 2016. CDK4/6 Inhibitors: Promising Opportunities beyond Breast Cancer. *Cancer Discov* 6:697-9.
 43. Schettini F, De Santo I, Rea CG, De Placido P, Formisano L, Giuliano M, Arpino G, De Laurentiis M, Puglisi F, De Placido S, Del Mastro L. 2018. CDK 4/6 Inhibitors as Single Agent in Advanced Solid Tumors. *Frontiers in Oncology* 8.
 44. Hortobagyi GN, Stemmer SM, Burris HA, Yap YS, Sonke GS, Paluch-Shimon S, Campone M, Petrakova K, Blackwell KL, Winer EP, Janni W, Verma S, Conte P, Arteaga CL, Cameron DA, Mondal S, Su F, Miller M, Elmeliegy M, Germa C, O'Shaughnessy J. 2019. Updated results from MONALEESA-2, a phase III trial of first-line ribociclib plus letrozole versus placebo plus letrozole in hormone receptor-positive, HER2-negative advanced breast cancer. *Ann Oncol* 30:1842.
 45. Büyükkaramikli NC, de Groot S, Riemsma R, Fayter D, Armstrong N, Portegijs P, Duffy S, Kleijnen J, Al MJ. 2019. Ribociclib with an Aromatase Inhibitor for Previously Untreated, HR-Positive, HER2-Negative, Locally Advanced or Metastatic Breast Cancer: An Evidence Review Group Perspective of a NICE Single Technology Appraisal. *Pharmacoeconomics* 37:141-153.
 46. Goel S, DeCristo MJ, Watt AC, BrinJones H, Sceneay J, Li BB, Khan N, Ubellacker JM, Xie S, Metzger-Filho O, Hoog J, Ellis MJ, Ma CX, Ramm S, Krop IE, Winer EP, Roberts TM, Kim HJ, McAllister SS, Zhao JJ. 2017. CDK4/6 inhibition triggers anti-tumour immunity. *Nature* 548:471-475.
 47. Petroni G, Formenti SC, Chen-Kiang S, Galluzzi L. 2020. Immunomodulation by anticancer cell cycle inhibitors. *Nat Rev Immunol* 20:669-679.

48. Watt AC, Cejas P, DeCristo MJ, Metzger-Filho O, Lam EYN, Qiu X, BrinJones H, Kesten N, Coulson R, Font-Tello A, Lim K, Vadhi R, Daniels VW, Montero J, Taing L, Meyer CA, Gilan O, Bell CC, Korthauer KD, Giambartolomei C, Pasaniuc B, Seo JH, Freedman ML, Ma C, Ellis MJ, Krop I, Winer E, Letai A, Brown M, Dawson MA, Long HW, Zhao JJ, Goel S. 2021. CDK4/6 inhibition reprograms the breast cancer enhancer landscape by stimulating AP-1 transcriptional activity. *Nat Cancer* 2:34-48.
49. Dunn GP, Bruce AT, Ikeda H, Old LJ, Schreiber RD. 2002. Cancer immunoediting: from immunosurveillance to tumor escape. *Nat Immunol* 3:991-8.
50. Gil Del Alcazar CR, Alečković M, Polyak K. 2020. Immune Escape during Breast Tumor Progression. *Cancer Immunol Res* 8:422-427.
51. Schreiber RD, Old LJ, Smyth MJ. 2011. Cancer Immunoediting: Integrating Immunity's Roles in Cancer Suppression and Promotion. *Science* 331:1565-1570.
52. Baxevasanis CN, Fortis SP, Perez SA. 2021. The balance between breast cancer and the immune system: Challenges for prognosis and clinical benefit from immunotherapies. *Semin Cancer Biol* 72:76-89.
53. Galli F, Aguilera JV, Palermo B, Markovic SN, Nisticò P, Signore A. 2020. Relevance of immune cell and tumor microenvironment imaging in the new era of immunotherapy. *Journal of Experimental & Clinical Cancer Research* 39:89.
54. Chaudhary B, Elkord E. 2016. Regulatory T Cells in the Tumor Microenvironment and Cancer Progression: Role and Therapeutic Targeting. *Vaccines (Basel)* 4.
55. Plitas G, Konopacki C, Wu K, Bos PD, Morrow M, Putintseva EV, Chudakov DM, Rudensky AY. 2016. Regulatory T Cells Exhibit Distinct Features in Human Breast Cancer. *Immunity* 45:1122-1134.
56. Ohara M, Yamaguchi Y, Matsuura K, Murakami S, Arihiro K, Okada M. 2009. Possible involvement of regulatory T cells in tumor onset and progression in primary breast cancer. *Cancer Immunol Immunother* 58:441-7.
57. Oleinika K, Nibbs RJ, Graham GJ, Fraser AR. 2013. Suppression, subversion and escape: the role of regulatory T cells in cancer progression. *Clin Exp Immunol* 171:36-45.
58. Li JJ, Tsang JY, Tse GM. 2021. Tumor Microenvironment in Breast Cancer—Updates on Therapeutic Implications and Pathologic Assessment. *Cancers* 13:4233.
59. Gonzalez H, Hagerling C, Werb Z. 2018. Roles of the immune system in cancer: from tumor initiation to metastatic progression. *Genes Dev* 32:1267-1284.
60. Liu J, Geng X, Hou J, Wu G. 2021. New insights into M1/M2 macrophages: key modulators in cancer progression. *Cancer Cell International* 21:389.
61. Valdes-Mora F, Handler K, Law AMK, Salomon R, Oakes SR, Ormandy CJ, Gallego-Ortega D. 2018. Single-Cell Transcriptomics in Cancer Immunobiology: The Future of Precision Oncology. *Front Immunol* 9:2582.
62. Salemme V, Centonze G, Cavallo F, Defilippi P, Conti L. 2021. The Crosstalk Between Tumor Cells and the Immune Microenvironment in Breast Cancer: Implications for Immunotherapy. *Frontiers in Oncology* 11.
63. Prager BC, Xie Q, Bao S, Rich JN. 2019. Cancer Stem Cells: The Architects of the Tumor Ecosystem. *Cell Stem Cell* 24:41-53.
64. Soysal SD, Tzankov A, Muenst SE. 2015. Role of the Tumor Microenvironment in Breast Cancer. *Pathobiology* 82:142-152.
65. Gascard P, Tlsty TD. 2016. Carcinoma-associated fibroblasts: orchestrating the composition of malignancy. *Genes Dev* 30:1002-19.
66. Elwakeel E, Weigert A. 2021. Breast Cancer CAFs: Spectrum of Phenotypes and Promising Targeting Avenues. *Int J Mol Sci* 22.

67. Dudley AC. 2012. Tumor endothelial cells. *Cold Spring Harb Perspect Med* 2:a006536.
68. Ding S, Chen X, Shen K. 2020. Single-cell RNA sequencing in breast cancer: Understanding tumor heterogeneity and paving roads to individualized therapy. *Cancer Commun (Lond)* 40:329-344.
69. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K, Surani MA. 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods* 6:377-382.
70. Qian J, Olbrecht S, Boeckx B, Vos H, Laoui D, Etliloglu E, Wauters E, Pomella V, Verbandt S, Busschaert P, Bassez A, Franken A, Bempt MV, Xiong J, Weynand B, van Herck Y, Antoranz A, Bosisio FM, Thienpont B, Floris G, Vergote I, Smeets A, Tejpar S, Lambrechts D. 2020. A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. *Cell Res* 30:745-762.
71. Stubbington MJT, Lönnberg T, Proserpio V, Clare S, Speak AO, Dougan G, Teichmann SA. 2016. T cell fate and clonality inference from single-cell transcriptomes. *Nat Methods* 13:329-332.
72. Ren L, Li J, Wang C, Lou Z, Gao S, Zhao L, Wang S, Chaulagain A, Zhang M, Li X, Tang J. 2021. Single cell RNA sequencing for breast cancer: present and future. *Cell Death Discov* 7:104.
73. Chung W, Eum HH, Lee HO, Lee KM, Lee HB, Kim KT, Ryu HS, Kim S, Lee JE, Park YH, Kan Z, Han W, Park WY. 2017. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat Commun* 8:15081.
74. Lee M-CW, Lopez-Diaz FJ, Khan SY, Tariq MA, Dayn Y, Vaske CJ, Radenbaugh AJ, Kim HJ, Emerson BM, Pourmand N. 2014. Single-cell analyses of transcriptional heterogeneity during drug tolerance transition in cancer cells by RNA sequencing. *Proceedings of the National Academy of Sciences* 111:E4726-E4735.
75. Saliba AE, Westermann AJ, Gorski SA, Vogel J. 2014. Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res* 42:8845-60.
76. Haque A, Engel J, Teichmann SA, Lönnberg T. 2017. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Medicine* 9:75.
77. Anonymous. 9 Aug 2021. Chromium Next GEM Single Cell 5' Reagents Kits v2 (Dual Index), *on* 10xGenomics. https://assets.ctfassets.net/an68im79xiti/57JaTECQNBPSpyDz8oucdi/ced6aa8eaf73d6ee18dea8fdbd945faa/CG000331_Chromium_Next_GEM_Single_Cell_5-v2_UserGuide_RevD.pdf. Accessed 12 April.
78. Genomics x. 2020. What is Cell Ranger?, *on* 10x Genomics. Accessed 03. May.
79. MarioniLab. emptyDrops: Identify empty droplets.
80. Anonymous. scDbfFinder. <https://github.com/plger/scDbfFinder>. Accessed
81. Steuart T BA, Hoffman P, Hafemeister C, Papalexi E, Mauck III WM, Hao Y, Stoeckius M, Smibert P, Satija R. 11 Jan 2022 2022. Seurat - Guided Clustering Tutorial. https://satijalab.org/seurat/articles/pbmc3k_tutorial.html. Accessed April 12.
82. Bassez A, Vos H, Van Dyck L, Floris G, Arijs I, Desmedt C, Boeckx B, Vanden Bempt M, Nevelsteen I, Lambein K, Punie K, Neven P, Garg AD, Wildiers H, Qian J, Smeets A, Lambrechts D. 2021. A single-cell map of intratumoral changes during anti-PD1 treatment of patients with breast cancer. *Nat Med* 27:820-832.
83. Wu SZ, Al-Eryani G, Roden DL, Junankar S, Harvey K, Andersson A, Thennavan A, Wang C, Torpy JR, Bartonicek N, Wang T, Larsson L, Kaczorowski D, Weisenfeld NI, Uytingco CR, Chew JG, Bent ZW, Chan C-L, Gnanasambandapillai V, Dutertre

- C-A, Gluch L, Hui MN, Beith J, Parker A, Robbins E, Segara D, Cooper C, Mak C, Chan B, Warriar S, Ginhoux F, Millar E, Powell JE, Williams SR, Liu XS, O'Toole S, Lim E, Lundeberg J, Perou CM, Swarbrick A. 2021. A single-cell and spatially resolved atlas of human breast cancers. *Nature Genetics* 53:1334-1347.
84. Fu C, Jiang A. 2018. Dendritic Cells and CD8 T Cell Immunity in Tumor Microenvironment. *Front Immunol* 9:3059.
 85. Clayton K, Vallejo AF, Davies J, Sirvent S, Polak ME. 2017. Langerhans Cells—Programmed by the Epidermis. *Frontiers in Immunology* 8.
 86. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, Chak S, Naikawadi RP, Wolters PJ, Abate AR, Butte AJ, Bhattacharya M. 2019. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nature Immunology* 20:163-172.
 87. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, Khodadoust MS, Esfahani MS, Luca BA, Steiner D, Diehn M, Alizadeh AA. 2019. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nature Biotechnology* 37:773-782.
 88. Luca BA, Steen CB, Matusiak M, Azizi A, Varma S, Zhu C, Przybyl J, Espín-Pérez A, Diehn M, Alizadeh AA, van de Rijn M, Gentles AJ, Newman AM. 2021. Atlas of clinically distinct cell states and ecosystems across human solid tumors. *Cell* 184:5482-5496.e28.
 89. Doebel T, Voisin B, Nagao K. 2017. Langerhans Cells - The Macrophage in Dendritic Cell Clothing. *Trends Immunol* 38:817-828.
 90. Peuker CA, Yaghobramzi S, Grunert C, Keilholz L, Gjerga E, Hennig S, Schaper S, Na IK, Keller U, Brucker S, Decker T, Fasching P, Fehm T, Janni W, Kümmel S, Schneeweiss A, Schuler M, Lüftner D, Busse A. 2022. Treatment with ribociclib shows favourable immunomodulatory effects in patients with hormone receptor-positive breast cancer-findings from the RIBECCA trial. *Eur J Cancer* 162:45-55.
 91. Haq F, Mahoney M, Koropatnick J. 2003. Signaling events for metallothionein induction. *Mutat Res* 533:211-26.
 92. Gomulkiewicz A, Podhorska-Okolow M, Szulc R, Smorag Z, Wojnar A, Zabel M, Dziegiel P. 2010. Correlation between metallothionein (MT) expression and selected prognostic factors in ductal breast cancers. *Folia Histochem Cytobiol* 48:242-8.
 93. Fu J, Lv H, Guan H, Ma X, Ji M, He N, Shi B, Hou P. 2013. Metallothionein 1G functions as a tumor suppressor in thyroid cancer through modulating the PI3K/Akt signaling pathway. *BMC Cancer* 13:462.
 94. Nagel WW, Vallee BL. 1995. Cell cycle regulation of metallothionein in human colonic cancer cells. *Proc Natl Acad Sci U S A* 92:579-83.
 95. Hoffman P. SL. 01.11.2022 2022. Cell-Cycle Scoring and Regression. https://satijalab.org/seurat/articles/cell_cycle_vignette.html. Accessed
 96. She L, Qin Y, Wang J, Liu C, Zhu G, Li G, Wei M, Chen C, Liu G, Zhang D, Chen X, Wang Y, Qiu Y, Tian Y, Zhang X, Liu Y, Huang D. 2018. Tumor-associated macrophages derived CCL18 promotes metastasis in squamous cell carcinoma of the head and neck. *Cancer Cell International* 18:120.
 97. Chen J, Yao Y, Gong C, Yu F, Su S, Chen J, Liu B, Deng H, Wang F, Lin L, Yao H, Su F, Anderson Karen S, Liu Q, Ewen Mark E, Yao X, Song E. 2011. CCL18 from Tumor-Associated Macrophages Promotes Breast Cancer Metastasis via PITPNM3. *Cancer Cell* 19:541-555.
 98. Su S, Liao J, Liu J, Huang D, He C, Chen F, Yang L, Wu W, Chen J, Lin L, Zeng Y, Ouyang N, Cui X, Yao H, Su F, Huang J-d, Lieberman J, Liu Q, Song E. 2017.

- Blocking the recruitment of naive CD4+ T cells reverses immunosuppression in breast cancer. *Cell Research* 27:461-482.
99. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, Zhang F, Mundlos S, Christiansen L, Steemers FJ, Trapnell C, Shendure J. 2019. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 566:496-502.
 100. Qian B-Z, Li J, Zhang H, Kitamura T, Zhang J, Campion LR, Kaiser EA, Snyder LA, Pollard JW. 2011. CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. *Nature* 475:222-225.
 101. Oh S, Kim H, Nam K, Shin I. 2017. Glut1 promotes cell proliferation, migration and invasion by regulating epidermal growth factor receptor and integrin signaling in triple-negative breast cancer cells. *BMB Rep* 50:132-137.
 102. Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, Giacomello S, Asp M, Westholm JO, Huss M, Mollbrink A, Linnarsson S, Codeluppi S, Borg Å, Pontén F, Costea PI, Sahlén P, Mulder J, Bergmann O, Lundeberg J, Frisén J. 2016. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353:78-82.
 103. Merritt CR, Ong GT, Church SE, Barker K, Danaher P, Geiss G, Hoang M, Jung J, Liang Y, McKay-Fleisch J, Nguyen K, Norgaard Z, Sorg K, Sprague I, Warren C, Warren S, Webster PJ, Zhou Z, Zollinger DR, Dunaway DL, Mills GB, Beechem JM. 2020. Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nature Biotechnology* 38:586-599.



Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitenskapelige universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway