Characterisation of the immune contexture of breast cancer subtypes

Lilly Anne Torland

Master Thesis

Molecular Biology

60 credits

Department of Biosciences

The Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

June 2018

Characterisation of the immune contexture of breast cancer subtypes

Lilly Anne Torland

Department of Clinical Molecular Biology
Oncogenomics
Akerhus Universitetssykehus HF

Department of Biosciences
The Faculty of Mathematics and Natural Sciences
University of Oslo

UiO : University of Oslo

AKERSHUS UNIVERSITETSSYKEHUS

2018

“Characterisation of the immune contexture of breast cancer subtypes”

Lilly Anne Torland

http://www.duo.uio.no

Trykk: Reprosentralen, Universitetet i Oslo

Abstract

Purpose: The purpose and aim of this project was to investigate immune contexture according to tumour molecular subtype, and ER status. To find the interrelationship between serum cytokine levels and specific tumour cell infiltration levels.

Material and methods: A newly released deconvolution tool called xCell, uses gene expression data to identify 64 different cell types present in the bulk tissue assessed for gene expression. We first validated the method by comparing xCell´s output with immune cell infiltration assessed by immunohistochemistry staining. The xCell profiles were then compared between the breast cancer molecular subtypes and ER status, to identify differences in the microenvironment. 26 immune cell profiles were compared to serum cytokine levels to find potential correlation between cell type levels and cytokine levels.

Results: The findings suggest clear differences of levels of immune cells in the different subtypes and ER status. ER negative samples have higher immune infiltration compared to ER positive samples. Luminal A have higher level of epithelial cells and stroma cell when compared to the other subtypes. The Luminal B has higher immune infiltration of helper T cells and higher immune infiltration when compared to Luminal A and Normal-like. The Normal-like subtype has higher levels of stem cell like cells. Her2-enriched and the Basal-like subtype have the largest amount of immune infiltration compared to the other subtypes. When correlating immune cell levels to cytokine levels, we found that levels of PDGF-bb corelate to higher levels of B cells which may reflect lymphangiogenesis.

Conclusion: These findings suggest that there are differences in the microenvironment of breast cancer subtypes. The method used to identify these findings should be better validated, as it only validates a small fraction of the cell types. In the future, a large validation using more markers, should be used. By further analysing the cell levels and infiltration profiles for individual subtypes, we could better understand the differences in the environment within a subtype. PDGF cytokine may be a good candidate for immunotherapy. These findings could potentially have prognostic and/or predictive value to deliver more targeted treatments.

Sammendrag

Formål: Formålet med denne oppgaven er å identifisere sammenhenger mellom ulike nivåer celletyper, ER status og subtyper i brystkreft. I tillegg til å identifisere en potensiell sammenheng mellom serum cytokin nivåer, og immune celle nivåer.

Materiale og metoder: For å identifisere ulike celle typer brukte vi "deconvolution" programmet xCell. Dette programmet utnytter det unike genuttrykket til en celle til å spesifikt identifisere en type celle. xCell kan ut fra genuttrykk data identifisere 64 celletyper. Dette gir unike muligheter til å se hvilke celler som er tilstede i en tumor. Ved å få tilgang til 15 datasett med genuttrykk data, brukte vi dette programmet til å lage celleprofiler. Vi brukte celleprofilene og statistiske analyser til å identifisere celletyper som er tilstede i større grad i de ulike typene bryst kreft. Vi validerte celleprofilene fra xCell ved å sammenligne "immunohistochemistry staining" med celleprofilene.

Resultater: Det var klar sammenheng mellom ulike celletyper og Brystkreft subtypene. I sammenligningen mellom ER status var det høyere infiltrasjon av immunceller i ER negative pasient prøver sammenlignet med ER positive prøver. Luminal A subtypen hadde høyere nivåer av epitelceller sammenlignet med de andre subtypene. Luminal B hadde noe immuncelle infiltrasjon spesielt av T-hjelpecellene. Normal-like hadde en større andel stamceller sammenlignet med de andre. De Basal-like og Normal-like subtypene hadde høyere immuncelle infiltrasjon sammenlignet med de andre subtypene. Her2-enriched hadde andel medfødte immunceller sammenlignet med den Basal-like subtypen. Vi fant også en sammenheng mellom cytokin PDGF nivåer og B celle nivåer. B celler og PDGF er involvert i dannelsen av nye lymfevener, og høyere nivåer PDGF kan ha en sammenheng med dette.

Konklusjon: Våre funn viser klare forskjeller mellom mikromiljøet i tumoren, og en sammenheng mellom nivåer cytokin PDGF og B celler. For å validere metoden sammenlignet vi xCell profilene med "immunohistochemistry staining", denne valideringen var noe mangelfull da den kun validerte en liten mengde av celletypene som ble benyttet. Ved å benytte en større valideringsett med flere markører ville det validert flere av celletypene som xCell benytter. Ved å sammenligne celleprofiler innad i subtypene samt sammenhengen mellom flere cytokin nivåer, kan vi potensielt finne prognostiske og prediktive markører.

Acknowledgments

The work that has been done during this master thesis and has been performed at Akershus university hospital at the department of Clinical Molecular Biology. Under the supervision of Dr. Xavier Tekpli and Professor Vessela Kristensen, who´s work and guidance have been astounding. The work was completed in the time period of 2017 to 2018.

I would like to thank my supervisors for creating a positive work environment and including me in the group. For giving me room to grow academically, answering all my questions, and for always taking me seriously. Thank you for all the time you have invested in me.
I would also like to thank you for including me in the groups retreats and team building activities, I have felt so welcome. Thank you for sending me to the NBCS contact meeting, it was highly motivating and inspiring to meet so many inspirational people and researchers.

I would especially like to thank Dr. Xavier Tekpli for selecting me as your master student and including me in your research. Thank you for all your help, for encouraging me to think critically and always lending a helping hand when I needed guidance. You are motivating and so kind and fun I felt prioritised and in good hands with you as a supervisor. So much of my academic growth is thanks to you.

I would also like to thank Vessela Kristensen for all the motivation and positive energy you spread, creating a positive and fun work environment. After all meetings with you I always felt motivated and eager to start working. I would also like to thank all the people in the group especially the team at Akershus university hospital for including me socially, making sure I never ate lunch alone and for always encouraging me.

Last, but not least, I would like to thank my family, friends and my husband for always encouraging me to do my best and for being there in all ups and downs. I would not be where I am today if not for you.

Lørenskog, April 2018
Lilly Anne Torland

Table of contents

1 List of abbreviations

aDC	Activated dendritic cell
APC	Antigen Presenting cell
BCR	B cell receptor
cDC	Conventical dendritic cells
CLP	Common lymphoid progenitor
CTL	Cytotoxic killer T cells
DC	Dendritic cell
DCIS	Ductal carcinoma in situ
DNA	Deoxyribonucleic acid
EMT	Epithelial-mesenchymal transition
ER	Oestrogen receptor
GMP	Common myeloid progenitor
HER2	Human epidermal growth factor receptor 2
HSC	Hematopoietic Stem Cell
IDC	Invasive ductal carcinoma
iDC	Immature dendritic cells
IHC	Immunohistochemistry
ILC	Invasive lobular carcinoma
LCIS	Lobular carcinoma in situ
MEP	Myeloid progenitor
MHC	Major Histocompatibility Complex
NCI	National cancer institute
NK cells	Natural killer cells
NLR	NOD- like receptors
PR	Progesterone receptor
PRR	Pattern Recognition Receptors
RNA	Ribonucleic acid
TCGA	The Cancer Genome Atlas
TCM	Central memory T cell
TCR	T cell Receptor
TD	T cell Dependent
TEM	Effector memory T cell

Tgd cell	Gamma delta T cell
Th	Helper T cells
TI	T cell Independent
TME	Tumour Microenvironment
TNBC	Triple-negative breast cancer
Treg	Regulatory T cells

2 Introduction

2.1 Cancer

Tumours arise when a cell grows abnormally and uncontrollably by bypassing the normal rules of cell division. The development of cancer is a multistep process. Accumulations of several genetic aberrations are required for the tumour to become malignant (1). The process of tumour development can be divided into stages (1).

- The first stage is the **initiation** phase, in which genetic alterations occur within the cells genetic material. These alterations can occur spontaneously or be induced by carcinogens.
- During the **promotion** stage, cells gains the ability to proliferate uncontrollably. Carcinogen promoters allow this process, during which accumulation of more mutations occur.
- In the last stage, called the **progression** phase, the tumour gains the ability to invade other tissues, it then becomes metastatic.

Normal cells are constantly subjected to signals controlling cell division. Cancer cells develop a degree of autonomy and avoid these signals resulting in uncontrolled cell division and proliferation.

2.1.1 Hallmarks of cancer

In 2000, Hanahan and Weinberg proposed a concept called the hallmarks of cancer (2). This concept summarises the biological specification and alterations in cancer. In 2011, the hallmarks were updated and extended, as illustrated below in Figure 1 (3). The concept of the hallmarks of cancer explains the processes a cell must undergo, and the challenges it overcomes to become cancerous (3).

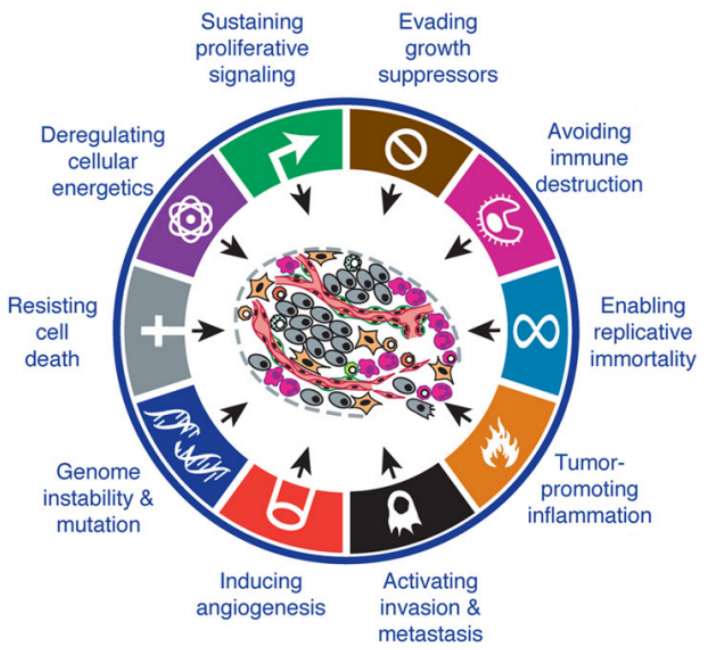

Figure 1. The 10 hallmarks of cancer proposed by Hanahan and Weinberg in 2011. Printed with permittion form "*Hallmarks of cancer: The Next generation*" *(3)*.

In order for a cancer cell to grow and divide uncontrollably the cell needs to **sustain proliferative signalling** and **evade growth suppressors** (2). While normal cells respond to growth promoting or repressing signals, cancer cells continually proliferate. Usually, if problems are detected during cell cycle or in the environment, the cell is brought into cell cycle arrest or is forced to undergo apoptosis: programmed cell death. Cancer cells adopt the ability to **avoid programmed cell death** (2). **Genome instability and mutations** are thought to be an onset in some types of cancers (3). Normal cells have restrictions on how many times they can divide, cancer cells have **enabled replicative immortality** as they may divide infinitely (2). The immune system detects pathogens and may also detect and destroy cells with altered DNA. Cancer cells express or suppress specific molecules at their surface in order to avoid immune recognition and **evade immune detection** (3)**. Tumour- promoting inflammation** is used by the cancer cells to supply growth factors, survival and promote angiogenesis (3). In tumorous cancer cells there is an increased need of nutrition and oxygen. To sustain the exponential growth and get the needed nutrients, cancer cells **induce angiogenesis**; the recruitment of blood vessels (2). Cancer cells also have altered metabolic pathways, **deregulating cellular energetics** (3). Finally, cancer cells have **invading and metastatic potential** allowing them to leave the tissue of origin and invade other tissues in the body (2).

2.1.2 Cancer genomics

Genetic alterations are a critical step in carcinogenesis. There are four main types of genetic alterations associated with cancers (4).

- **Mutations**: a permanent change to the DNA sequence. Mutations can be subcategorised into; deletions, insertions or substitutions. Mutations can be hereditary, or somatic: acquired through a person's lifetime (5).
- **Aneuploidy** is the process of loss or gain of chromosomes (6).
- **Chromosome translocation** is the fusion of two chromosomes, which may alter the function or expression of one or several genes (6).

In addition**,** amplification/deletion may play a critical role in cancers. When a gene is **amplified** the copy number increases resulting in high gene expression, and overexpression of that gene. Overexpression or amplification of certain genes that regulate the cell cycle or stimulate growth may cause cancer (6).

Genetic alterations can be passenger mutations and do not directly cause cancer. However, if mutations alter important high-risk genes, involved in crucial steps of DNA repair, apoptosis or cell proliferation, these alterations may be carcinogenic. When genetic alterations affect oncogenes or tumour suppressor genes, they may become functional and drive the carcinogenic process (7).

Oncogenes are genes with the ability to promote cancer. Before they become oncogenes, they are proto-oncogenes and are involved in proliferation and stimulating cell growth. Genetic alterations of a proto-oncogene which lead to a gain of function usually drive oncogenesis (7). Oncogenes often mutated in breast cancer are ESR1, Pik3CA, GATA3 and MAP3Ki (8).

Tumour-suppressor genes (TSG) are genes inhibiting overproliferation and uncontrolled cell division. The TSG are often involved in DNA repair, suppressing growth and promoting apoptosis. Mutations in the TSG usually lead to loss of function (7). TSG often mutated in breast cancer are TP53, CDH1, PTEN, and BRCA1/2 (9).

2.2 Breast Cancer

2.2.1 Breast anatomy

The breast is composed of fatty tissue or adipose tissue, and glandular milk-producing tissues or lobules see Figure 2 (10). The main function of the female breast is to produce milk to infants during breastfeeding. The mammary glands are exocrine glands consisting of 1-12 milk producing lobes, from the mammary glands lactiferous milk ducts branch out bringing the milk

to the nipple. In addition, to the granular tissue and the adipose tissue the breast is composed of structural tissues such as muscles and fibrous tissue, and ligaments. The breast also contains lymph nodes and blood vessels, supplying the breast with nutrients and protecting it from infections (10).

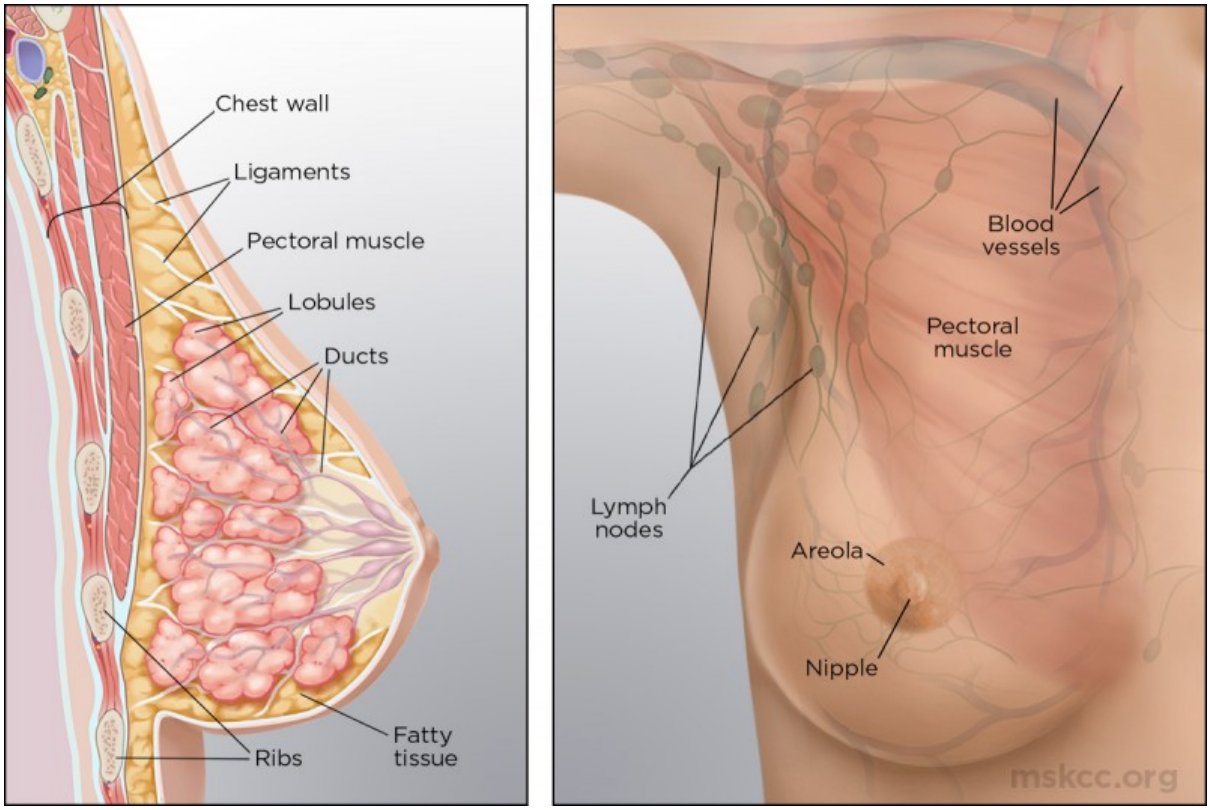

Figure 2. Anatomy of the human breast. The mammary gland consists of ducts and lobules, surrounded by adipose tissue. The breast is held in place by ligaments and muscle cells. Lymph nodes and blood vessels infiltrate the tissue. Picture modified from Memorial Solan Kettering cancer centre *(11)*.

The ratio of adipose and glandular tissue varies from person to person (10). The breast development occurs after childhood and undergoes several changes throughout a lifetime (12). Hormonal stimuli during puberty induces changes to the female breast. The adolescent breast consist of only rudimentary glandular structures which develop into primary ducts emanating from the nipple. During menopause, levels of oestrogen decreases and the breast composition changes: an increase in fatty tissue and decrease of granular tissues is observed (12).

2.2.2 Epidemiology

Breast cancer is the second most frequent type of cancer worldwide. In 2012, 1.67 million new cancer cases were reported (13). Global incidence rates vary, and higher incidence occurs in developed countries (13) (Figure 3). Studies show that migrants with an initial low incidence of developing breast cancer have an increased incidence within a few generations after moving from a country with low to higher incidence rates. This suggests that environmental factors and lifestyle are important aspects in development of the disease (14).

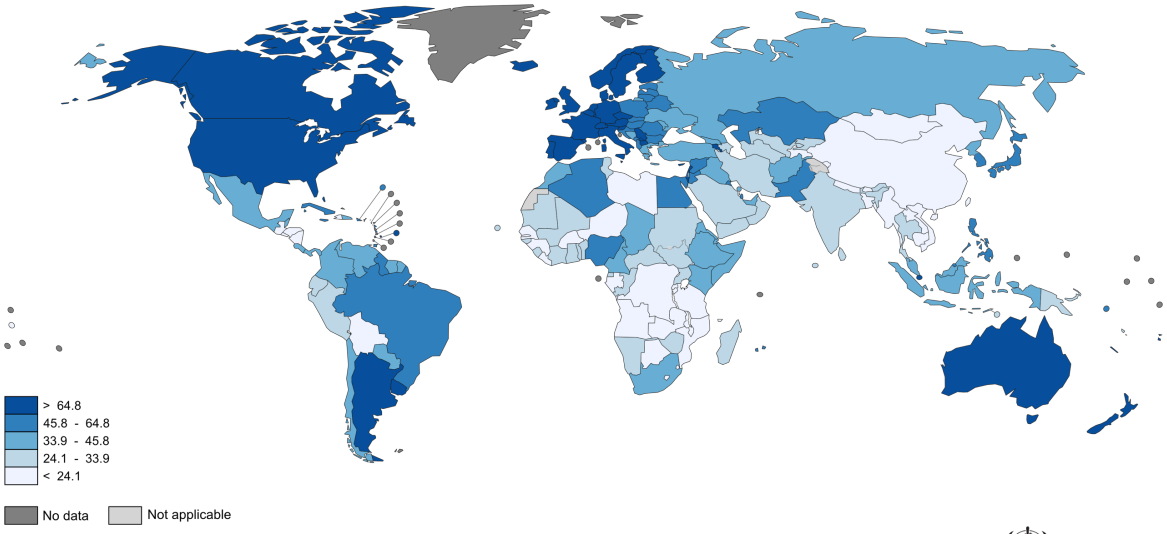

Figure 3. Estimate breast cancer incidence rates world-wide. Figure from GLOBOCAN. Printed with permission from IARC and WHO *(13)*.

In Norway breast cancer is the most common cancer among women. In 2016, 3636 women were diagnosed with breast cancer. One in every twelve women will develop breast cancer before they are 75 years old, in Norway (15). The incidence of breast cancer has been increasing rapidly the last decades. From 1957-1961, there was an average of 941 new cases per year, while between 2012-2016, 3254 new cases were reported per year. However, as the incidence rate increases, the mortality rates decrease, and about 90% of women diagnosed with breast cancer will still be alive 5 years after diagnosis which was not the case 40 years ago (15). The decrease in mortality rates is due to early detection through the mammographic screening programs, and better awareness among women. As technology and science have progressed, the hospitals and doctors are able to deliver better treatments which participate in the decrease in mortality by breast cancer (15) (16).

2.2.3 Tumour development and progression in breast cancer

Checkpoints during the cell cycle ensure that cells undergoing division are healthy and normal. However, some cells gain the ability to avoid these checkpoints, and therefore divide frequently and abnormally. The abnormal proliferation creates a mass referred to as a neoplasm (17). Breast cancers develop in the ducts; ductal carcinoma, or in the lobules; lobular carcinomas. Tumours can be *in situ* or invasive; *in situ* tumours are confined within the tissue they derived from, while invasive tumour have started to invade surrounding tissues. The conversion from normal epithelial tissue to an invasive carcinoma is thought to be through the progression of stages described in Figure 4. The progression starts as ductal or lobular hyperplasia and atypical hyperplasia where there is a small amount of abnormal growth. Further it progresses into ductal carcinoma *in situ* (DCIS) or lobular carcinoma *in situ* (LCIS), the cancer is still confined to the duct or lobule (18). The last stage of progression is invasive carcinoma in which the tumour evades the lobes or the ducts. The development and progression of breast cancer is genetically and histologically diverse and differs from person to person.

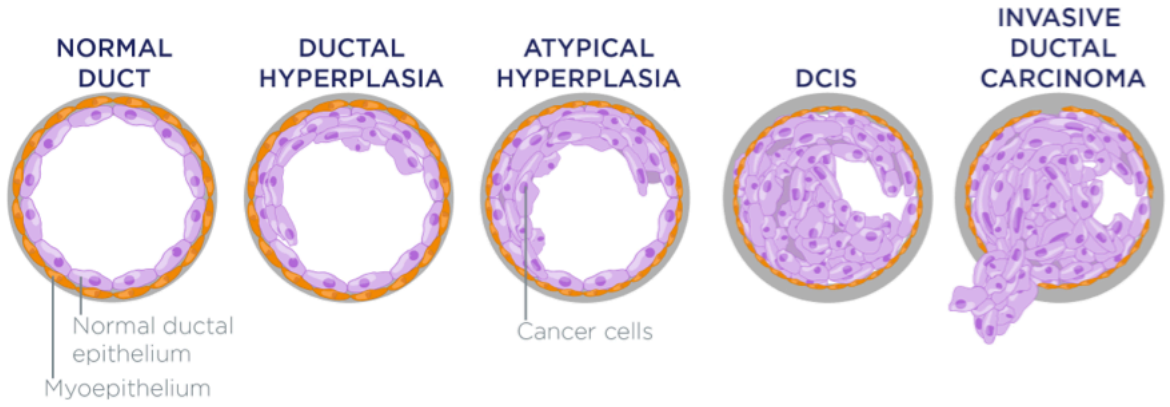

Figure 4. Developmental stages in breast cancer. Picture modified from RnCeus *(19)*.

2.2.4 Prognostic and predictive markers of breast cancer

Prognostic markers are to predict clinical outcome. A predictive marker is a factor that gives information about how a patient may responds to a specific treatment. Predictive markers help clinicians provide the right treatment to the right patient, avoiding overtreatment and sparing non-responding patients from side effects. Treatment decision is especially based on the morphology, histological grade, hormonal markers such as ER status, tumour size, and the proliferation and lymph vascular invasion status (20).

Histologic grade

Histological grade reflects how differentiated the cancer cells are compared to the normal tissue (21). The grading is based on three morphological features: (i) The degree of tubule or gland formation or the percentage that is still displaying normal structure. (ii) Nuclear pleomorphic or size and shape. (iii) Mitotic count, the number of dividing cells. The overall grade is decided from the overall score of the individual features. Histological grade is an independent prognostic factor (22).

Stage

Invasive breast cancers are classified by stage using the international TNM system. The stage is decided by three factors, the tumour size (T), the spread to the lymph nodes (N), and the spread or metastasis to other body parts (M) (23).

Molecular markers

Immunohistochemical (IHC) staining is used to account for protein expression of ER, PgR, Ki-67, and Her2. These proteins serve as prognostic and predictive markers and are important when considering treatment (24).

Hormone receptor

Oestrogen is a steroid hormone important in the reproductive cycle in women. Oestrogen binds to the oestrogen receptor on effector cells. Oestrogen receptor functions as a transcription factor and can regulate target gene expression. 70% of breast cancer are ER positive in which the oestrogen receptor is overexpressed (25). Progesterone is a steroid hormone produced by the ovaries and is important in the development of breast, and during the production of milk. Like ER the progesterone receptor (PgR) is a transcription factor. PgR is expressed by a large number of epithelial cells in the breast tumour (26).

Human epidermal growth factor Receptor (Her2)

The Her2 is a transmembrane protein that is involved in signalling proliferation, and differentiation. Her2 overexpression leads to an over activation of the cellular proliferation. About 25-30% of breast cancer cases have overexpression and amplification of the Her2 (ERBB2 gene), which is associated with poor prognosis (27).

Ki-67

The Ki-67 is a protein expressed during the stages of the cell cycle except in the G0 phase. An elevated level of this protein, is an indication of proliferation. Ki-67 levels are used to assess whether patients may receive chemotherapy (28).

PAM50 subtyping

During the last decades, major advances in high-throughput technologies have allowed to phenotype tumours at the molecular level. In 2000, Perou and Sørlie et al. proposed a molecular classification of breast cancers (29). This classification subdivides breast cancer in five groups: Luminal A, Luminal B, basal-like, Her2-enriched and Normal-like. The Luminal groups are usually ER positive (Table 1). When comparing patients with Luminal A and Luminal B subtypes, the Luminal B patients have a worse outcome and a higher degree of proliferation than the Luminal A patients. The Her2-enriched tumours are dominated by samples with an increased Her2 signalling pathway. Patients with Her2-enriched tumours have a worse prognosis. However, since 2000 targeted treatment for Her2-enriched breast cancer have been developed resulting in a better outcome for these patients. Basal-like tumours are usually triple negative; ER, PgR and Her2 negative. Basal-like tumours have the worst prognosis and are highly proliferative. The Normal-like group is not as well defined but consists of invasive adenocarcinomas with a gene expression profile similar to normal breast tissue (29).

By using statistical tools, 50 genes were prioritised to subtype breast cancers: the PAM50 panel. The PAM50 classification is recently used in clinical setting to identify subtype, to assess patients risk of recurrence, and to guide treatment decisions (29).

Table 1. Subtypes with ICH according to the St Gallen guidelines in 2016 *(24)*.

Molecular subtype	Surrogate subtype	ER	PgR	Her2	PI (Ki-67)
Luminal A	Luminal A- like	+	>20%	-	<20%
Luminal B	Luminal B like (HER2-negative) Luminal B like (HER2- positive)	+ +	<20% Any	- +	>20% Any
Her2 – enriched	HER2 positive	-	-	+	Any
Basal-like	Triple negative	-	-	-	Any
ER: oestrogen receptor; PgR progesterone receptor; Her2 Human epidermal growth factor receptor 2; PI proliferation index; + : positive; negative: -					

2.3 Immunology

The immune system protects the body from pathogens, through various lines of defences. Pathogens are microorganisms that can cause damages and diseases. There are four broad categories of pathogens; fungi, viruses, bacteria/archaea, and parasites. Not all microbes are pathogens, these are called commensal microorganisms and usually have a symbiotic relationship with the host. Three strategies are used to defend against microbes; avoidance, resistance and tolerance. The skin and mucosal surfaces protect the inner body from the pathogenic microbes, when these barriers are breached the pathogens can invade. When a pathogen starts to damage the body´s resistance, a response to destroy or reduce the pathogen population is initiated. The last defence is tolerance, where the cells capability to resist the pathogens is mounted. The immune system is divided into: **the innate and the adaptive immune system** (Figure 5) (30).

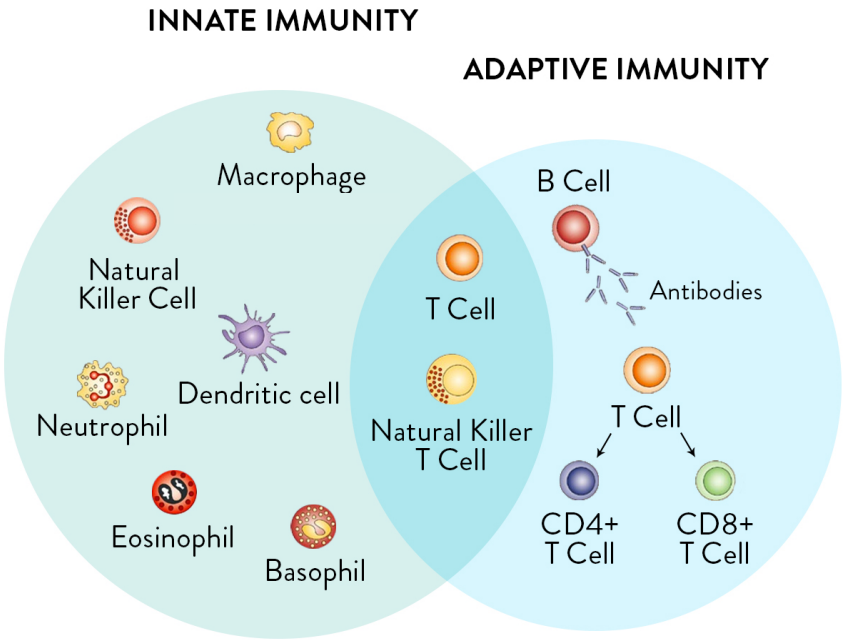

Figure 5. The different cell types found in the Innate and the Adaptive immune system. Picture modified from Nature Reviews *(31)*.

2.3.1 Innate immune system

The innate immune system is the first line of defence when a pathogenic microorganism breaches the immunogenic barriers. Sensor cells, like macrophages, dendritic cells and neutrophils detect inflammatory inducers on the pathogen surface. The inflammatory inducers are pathogen specific, for instance lipopolysaccharides are found on the surface of bacteria. Inflammatory receptors binds to unique characteristics of the pathogen, activating the innate immune system creating an inflammatory response to destroy microbes and activate other immune cells. The innate immune response is mounted quickly (32). The common myeloid progenitor (CMP) is the precursor of the macrophages, granulocytes, mast cells and dendritic cells, all of which are involved in the innate immune response (33). The cells of the innate immune system do not have specific antigen receptors, instead they express a limited number of invariant innate recognition receptors (34). The innate receptors are transcribed by specific invariable genes.

These receptors are also known as pattern recognition receptors (PRR) because they recognise pathogen specific molecules or patterns from a molecular structure (35). PRR can be divided into two groups;

- The toll like receptors (TLR), detecting structures on the extracellular surface of bacteria or the bacteria engulfed by vesicles (36).
- The NOD- like receptors (NLR) that sense intracellular invasion (37). Such cytoplasmic receptor can detect foreign RNA and DNA from viruses or other organisms.

The activation of a PRR for instance, on macrophages or neutrophils activates their phagocytic functions. PRR also amplify the innate immune system reaction through inflammatory mediators. Such inflammatory mediators, are chemokines and cytokines. Cytokines are secreted proteins relaying a signal between different immune cells therefore allowing them to "communicate". Cytokine´s signal can travel through the bloodstream or be local signals (38). Figure 6 shows a summary of the cells found in the innate immune system.

Macrophages

Macrophages monitor the environment seeking for pathogens which will activate them through the binding of an inducer to their receptors. Macrophages may reside in the tissue for years while being developed during embryonic development. Other types of macrophages develop later and are derived from circulating monocytes. Circulating monocytes migrate to the tissue where they are needed and differentiate into macrophages. Macrophages have many different functions both in the innate and the adaptive immune response. One of the main functions of

the macrophage is to engulf and destroy pathogens. The destruction of pathogens triggers the inflammatory response, which is important for both the innate and the adaptive immune system (39). There are different types of macrophages, M1 macrophages are more involved in destroying pathogens. M2 macrophages are more involved in the clean-up process, removing debree from the infected area.

Granulocytes

Granulocytes, are defined by having granules in the cytoplasm. There are three types of granulocytes: neutrophils, eosinophils and basophils. Granulocytes derive from the bone marrow and are relatively short lived. During an infection, the production of granulocytes increases, and they migrate to the site of infection. Neutrophils have phagocytotic properties, engulfing large amounts of microbes, destroying them through encapsulation in vesicles containing digesting enzymes. Eosinophils and basophils do not have phagocytic properties, however, they secrete vesicles containing toxins to destroy pathogens. Eosinophils and basophiles are efficient in destroying parasites, as these pathogens are usually too large to be engulfed (40).

Mast cells

Mast cells are derived from the bone marrow and migrate to the tissue where they further develop and mature. Mast cells contain large numbers of granules containing histamine and other inflammatory mediators. The mast cells are thought to be a line of defence directed toward parasitic invasion. Mast cells are also involved in allergic response and in cancer (41).

Dendritic cells

Dendritic cells also have phagocytic properties, but unlike the macrophages and neutrophils the dendritic cells also take up a large amount of extracellular fluid, by a process called micropinocytosis. Dendritic cells have a large membrane and after processing the matter they have engulfed they present antigens on their membranes. The main function of dendritic cells is therefore to present antigens to other immune cells in lymph nodes where they activate the adaptive immune system (42).

Natural killer Cells

Natural killer cells or NK cells, are large lymphocyte like cells. NK cells are derived from the common lymphoid progenitor (CLP). Like the other cells of the innate immune system the NK do not have antigen specific receptors but instead have receptors that are more general. (43).

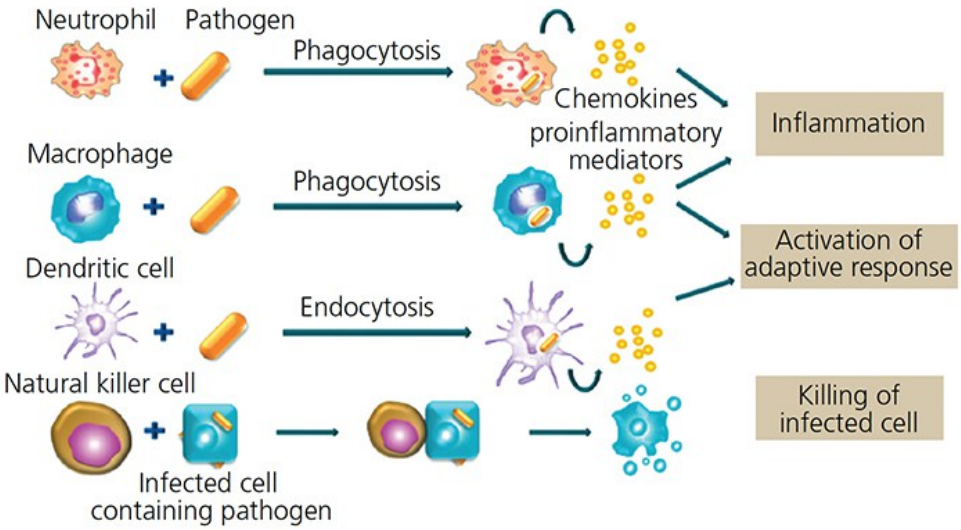

Figure 6. Cells in the Innate immune system and their functions. Picture modified from basis of disease 8 edition *(44)*.

2.3.2 Adaptive immune system

Lymphocytes are the effector cells of the adaptive immune system, they are able to produce specific and targeted antigen receptors. These receptors are produced through a highly regulated process during which different genes are recombined to produce a high diversity of receptors. Over one billion lymphocytes with unique receptors are patrolling the body seeking for a fitting antigen. When a lymphocyte with a specific receptor is activated, the information will be saved (immunological memory). Further exposition to a pathogen carrying the same antigen immunological memory will be used to mount a rapid and efficient defence. Vaccines use immunological memory to introduce the body to a modified harmless version of a pathogen. When the real pathogen invades the body, there will already be an antigen specific response. There are two major types of lymphocytes the B and the T lymphocytes (45). Figure 7 shows some of the cells found in the adaptive immune system.

T Lymphocytes

T cells develop in the thymus and migrate to lymph nodes where they further differentiate. T cells have unique cells surface receptors known as T cell receptors (TCR). The T cell receptor does not bind directly to antigens. Instead the TCR recognises antigens through major histocompatibility complex (MHC). The human MHC is called human leukocyte antigen complex (HLA). The MHC are transmembrane glycoproteins that presents antigen to the TCR. Non-activated T cells are known as **naïve T cells**, and patrol the body seeking for antigens (46). Binding to the MHC and co-stimulation activates the naïve T cells (47), they begin to proliferate and differentiate. The activated T cell gains the ability to perform specialised functions and becomes an effector T cells.

Effector T cells have the subset of various T cell types, and can differentiate to helper, killer or regulatory T cells. The Effector T cells can also be distinguished according to the type of receptor they express: the CD4 positive and the CD8 positive T cells (47).

T cells carrying the CD8 become **cytotoxic killer T cells (CTL)**. The CD8 receptor binds to antigen presented on the MHC class I molecule, expressed by all nucleated cells. Cytotoxic T cells are especially important in the defence against intracellular pathogens such as viruses. Infected cells present peptide fragments of the pathogen on the MHC class I (48).

The CD4 receptor binds to antigens found on the MHC class II molecules. The MHC class II protein is found on the antigen presenting cells (APC) including macrophages, dendritic cells and B cells (48).

- **Helper T cells (Th)** have a CD4 receptor, when activated Th cells secrete specific cytokines to modulate the immune response. There are different subtypes of helper T cells and they have different functions according to which cytokines they produce. The main subclasses are Th1, Th2, Th17 and Tfh (48).
- **Regulatory T cells (Treg)** are CD4 positive and regulate the T cells-mediated response. Treg may halt the immune reaction, by secreting immunosuppressive cytokines. Another function is to suppress autoreactive T cells that have avoided the process of negative selection in the thymus, to avoid autoimmune disease (49).

Memory T cells are long lived and have the ability to expand effector T cell populations quickly when they are reactivated by the cognate antigen and co-stimuli. The Memory T cells can be both CD4 or CD8 positive (50).

B Lymphocytes

B cells mature in the bone marrow and migrate to secondary lymphoid tissues such as the spleen and lymph nodes where they can be activated. B cells recognise antigens by using immunoglobulins or B cell receptors (BCR). The immunoglobulins are very specific to an antigen, and each B cell produces only immunoglobulin with a single specificity. When the receptor is bound to the membranes of the B cells, it is a BCR, when secreted it is an antibody (51).

B cells can be activated by T cells through thymus dependent (TD) pathway, or directly from microbial components, thymus independent (TI) pathway. Whilst in the TD pathway the antigen binds to the B cell receptor (BCR). The antigen is then endocytosed and broken down into peptides which will be presented on MHC-II molecules to activate Th cells. When the Th cell receptor bind to the B cell-MHC II-antigen complex it will initiate the Th cell to express CD40 co-stimulatory signal and cytokines, which will allow expression and differentiation of B cells. The activation of the B cell causes it to undergo proliferation, immunoglobulin class switching and somatic hyper mutation (52). When activated the B cells differentiate into different types of B cells; short lived plasmablasts, long-lived plasma cells and memory B cells. TI activation is initiated by foreign polysaccharides and unmethylated CpG DNA. Instead of being activated by T cells they are activated by cell that have toll like receptors or by crosslinking of B cell receptors (53).

B cells differentiate though highly regulated signalling cascades. **Plasmablasts** are short lived and secrete a large number of antibodies with low affinity binding. **Plasma cells** are long lived, non-proliferating antibody secreting cells which produce antibodies with high affinity binding. The antibodies bind to the antigen on the pathogen and the other end is exposed and can bind effector cells that destroy the pathogen (54).

Memory B-cells are produced during the differentiation and activation process in the germinal centres, or independently from short lived plasma cells. The Memory B cells are long lived and reside in the spleen and lymph nodes. Upon activation they can initiate a strong secondary response. During the secondary response, the memory B cells can re-enter the germinal centres and further evolve by undergoing hypermutations and affinity mutations, before differentiating into plasma cells (55).

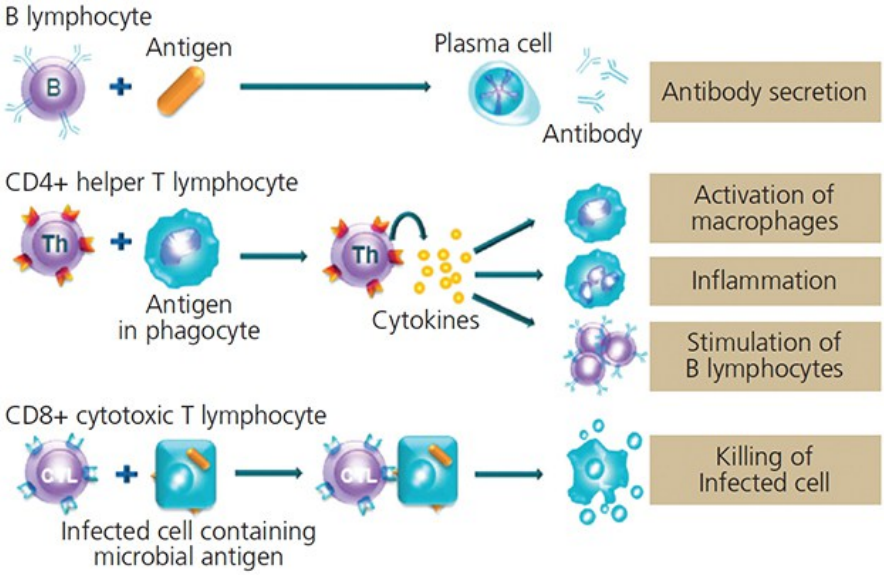

Figure 7. Cells in the Adaptive immune system and their functions. Picture modified from basis of disease 8 edition (44)**.**

2.3.3 Cytokines and Chemokines

Cytokines and chemokines are small molecules secreted by stromal, immune and tumour cells. These molecules coordinate and regulate the immune response (56). Figure 8 displays some of the cytokines released by activated Th1 and their functions. Cytokines secreted from helper T cells are also involved in regulating which type of immunoglobulin the B cell should produce (57).

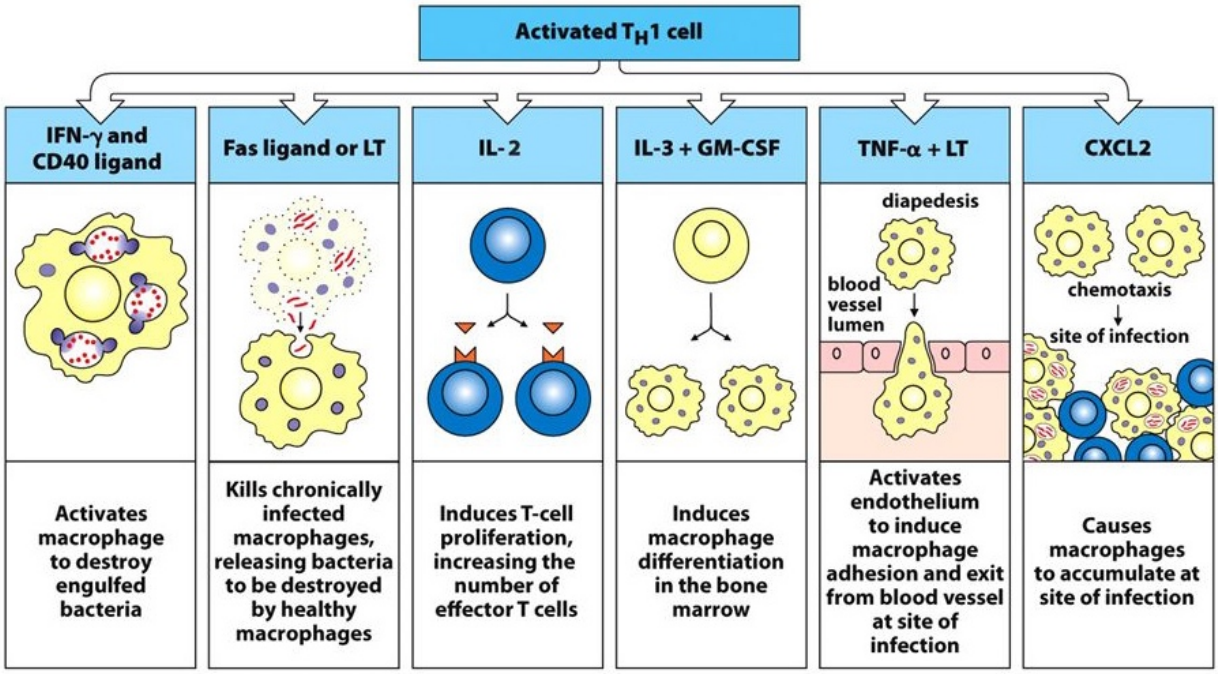

Figure 8. Functions of Th1 secreted chemokines. Picture modified from Janeway´s immunobiology textbook 6. edition.

Cytokines are also used by cells of the innate immune system to maintain and mount an inflammatory response. (58). Chemokines released by macrophages act on the surrounding tissues and instate an appropriate immune response according to pathogen. The functions of the different cytokines released by macrophages are displayed in Figure 9. The inflammatory response can act on both cells of the innate and the adaptive immune system.

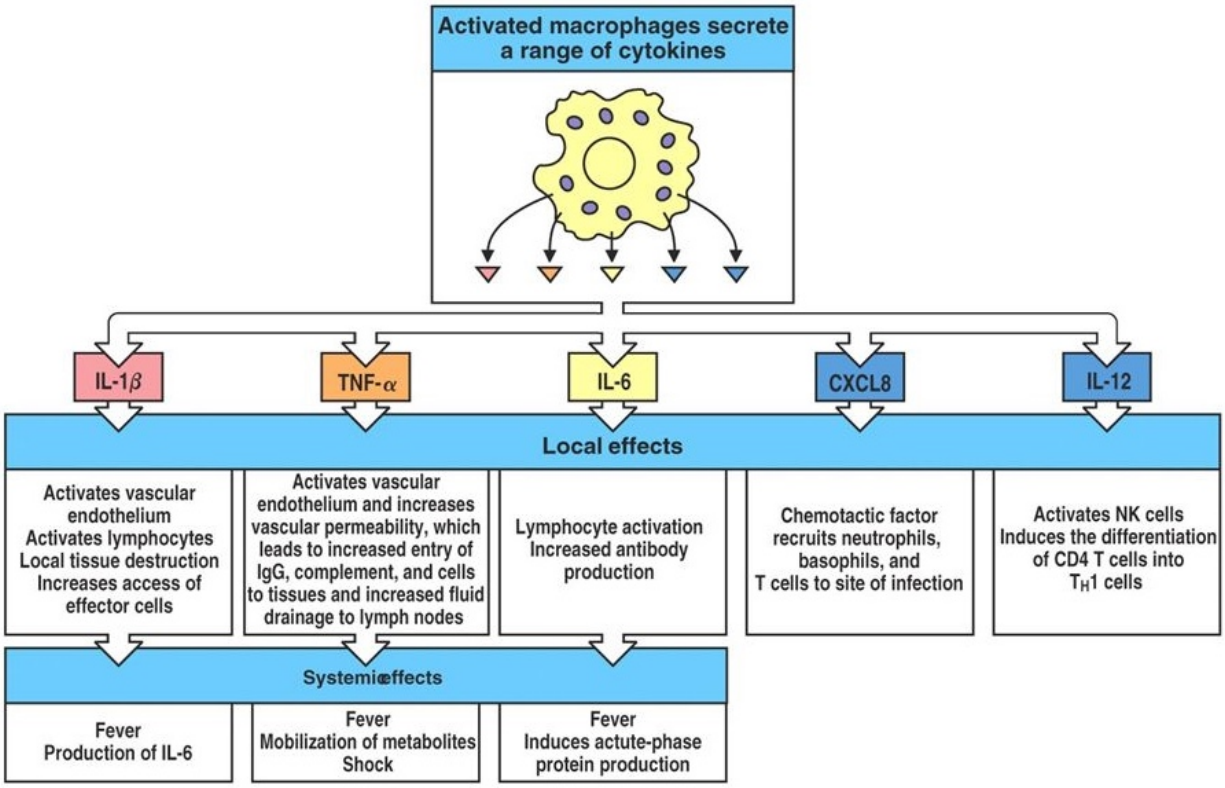

Figure 9. Functions of Macrophages secreted chemokines. Picture modified from Janeway´s immunobiology textbook 6. edition.

Table 2 summarises the cytokines, chemokines and growth factors measured in the serum of breast cancer patient during my internship, the source of cytokines and their main functions.

Table 2. Chemokines and cytokines, source of production and function *(59) (60)*.

Cytokine/ chemokine	Binds to receptor	Source of cytokine production	Function of cytokine
Eotaxin-1	CCR3	Dermal fibroblasts, airway, epithelial cells and heart cells.	Recruitment of eosinophils, basophils, Th2 subset and smooth muscle cells.
FGF-basic	FGFR-1-to FGFR-6	Adipocytes, normal and tumour cells.	Production of plasminogen activator (PA) and collagenase, induces DNA synthesis and endothelial cell proliferation, mitogenic and angiogenic activities.
G-CSF	CSF3R	Monocytes, T cells, fibroblasts and endothelial cells activated by macrophage.	Survival, proliferation, differentiation, and activates neutrophil precursors and mature neutrophils.

GM-CSF	CSF2R	Monocytes, T cells, fibroblasts, mast cells, NK cells, endothelial cells and macrophage.	Activation, proliferation, differentiation of dendritic cells, neutrophils, eosinophils and mononuclear phagocytes.
IFN-γ	IFNGR1 and IFNGR2	Activated CD4+ T lymphocytes (Th1), NKT and NK cells.	Th1 and CD8 memory cells activation and proliferation. Macrophage activation, increased expression of MHC. Suppression of Th2 and Th17.
IL-10	2α[R1] and 2β[R2] subunits= IL10R1+R2	Monocytes, macrophages, dendritic cells, Th1, Th2, Th17 and Treg.	Inhibition of Th1, stimulation of Treg, NK and B cell. Suppresses macrophage functions.
IL-12	IL-12Rβ1 + IL-12Rβ2	Monocytes/macrophages and dendritic cells	Proliferation of Th1. Activates NK cells.
IL-13	IL13R α1 and IL13R α2+IL4Rα	Th2 cells, monocytes	Downregulation of macrophages, monocyte. B lymphocyte differentiation and proliferation, increases CD23 expression, and induces IgG4 and IgE class switching.
IL-15	IL15R α + IL2R α + IL2R γ	Monocytes, epithelial cells, T lymphocytes and fibroblasts.	Recruitment and activation of T lymphocytes, expansion of B cells, proliferation of NK cells.
IL-17	IL-17RA + IL-17RC	Th17, CD4+, CD8+, gamma-delta T (γδ-T), invariant NKT and innate lymphoid cells (ILCs).	Recruitment of T cells, neutrophils, monocytes, basophils, eosinophils and inflammation induction.
IL-1β	IL-1R1 + IL1RAcP	Macrophages, Th1, and epithelial cells	Caspase1-apoptosis, Thymocytes proliferation, B-cell maturation and proliferation, and fibroblast growth factor activity, activation of macrophages. Fever inducer.
IL-1RA	(Antagonist of) IL1R	Epithelial, leukocytes, adipocytes, monocytes, macrophages, neutrophils, hepatocytes.	Inhibits the activity of interleukin-1 (alpha and beta) by binding to receptor IL1R.
IL-2	IL2Ra, IL2Rb +/- γc	Activated T cells, activated DCs.	T cell proliferation, B-cells, monocytes, lymphokine-activated killer cells, natural killer cells, and glioma cells proliferation.
IL-4	IL4R α + γc (cd132)	Macrophages, Th2 cells, basophils and mast cells, group-2 innate lymphoid cells (ILC2s).	Recruitment of mediators of cell growth, of resistance to apoptosis, and of gene activation and differentiation, stimulates B cells to produce IgE.
IL-5	IL5Rα + β (cd131)	Th2, NK, mast cells.	Eosinophil differentiation and activation and stimulation of immunoglobulin class switching to IgA, B cell proliferation.
IL-6	IL6ST + GP130 + IL6-β	Mononuclear phagocytes, T cells, and fibroblasts	Induction of mature B cells into plasma cells, Th17 activation. Fever inducer.
IL-7	IL7R α + γc	Bone marrow and thymic stromal cell, DCs, keratinocytes, hepatocytes, neurons and epithelial.	T cells homeostasis, development of pre-B and pre-T cells and early thymocytes.
IL-8	CXCR1/2	Macrophages, epithelial, airway smooth muscle cells and endothelial cells.	Tumour epithelial-mesenchymal transition (tumour-EMT), neutrophils recruitment
IL-9	IL9R α + γc	Th9, Th2, Th17, mast cells	Upregulate CD8+T, stimulate production of immunoglobulins by B cells and the proliferation of mast cells, chemokines.

IP-10	CXCR3	Monocytes, fibroblasts and endothelial cells (IFN-g triggered).	CXCR3 expressing leukocytes mobility and recruitment to cancer microenvironment.
MCP-1	CCR2	Monocytes, macrophages, DCs, endothelial, epithelial, fibroblasts, smooth cell.	Induce monocyte and memory T lymphocyte and NK cells migration.
MIP-1α	CCR1, CCR5	Macrophages, neutrophils (bacterial endotoxin stimulation).	Induce granulocytes, macrophages, T cells migration.
MIP-1β	CCR5	Macrophages, neutrophils.	Induce granulocytes, macrophages, monocytes, naive T cells migration.
PDGF- ββ	PDGFR-β (αβ- and/or ββ)	Activated platelets, macrophages, endothelial and muscle cells.	Promotes early endothelial cell differentiation, lymphogenesis and angiogenesis.
RANTES	CCR5	Neutrophils, monocytes, macrophages, T cells and neutrophils.	Induce eosinophils, monocytes, macrophages, and NK cells migration
TNF- α	TNFR-1 & TNFR-2	Monocytes, macrophages, lymphoid cells, mast cells, endothelial cells, fibroblasts and neurons.	Induction of fever, apoptotic cell death, cachexia, inflammation.
VEGFA	VEGFR1 & VEGFR2	Cells under hypoxia	Vasculogenesis, angiogenesis, lymphogenesis, cell migration, cell growth.

2.4 Immunity and Cancer

2.4.1 Immune surveillance and immune editing

The immune system plays a critical role in protecting the body against cancer. This became clear after observing patients with immune deficiencies. Immune-deficient patients (*i.e* after an organ transplant, or HIV infection) have a higher risk of developing cancers (61).

It is speculated that many emerging cancer cells may be eliminated before they grow into a significant detectable tumour. This concept is referred to as immune surveillance (62). Due to genetic aberrations, cancer cells may express peptides and protein which can be recognised as foreign, these are called neoantigens and allow a specific anti-cancer immune response to begin (63). The concept of immune surveillance speculates that the majority of cancer cells may be recognised and destroyed by the immune system. Only a few of the cancer cells will avoid immune surveillance though immune editing (64).

Cancer cells may avoid immune detection. One of the most common mechanism is downregulation of human leucocyte antigen complex (HLA) by the tumour. One third to half off all cancer cells lack HLA class molecules. Such cancers could however still be targeted by NK cells which can recognise and target cells lacking HLA molecules (65). A few mechanisms of immune editing, are listed in Table 3 (66) (64).

Table 3. Immune editing mechanisms to avoid immune detection and the impact these alterations have on the immune system. Table modified from; The biology of cancer, By R.A Weinberg *(67)*.

Immune check point	Impact on the immune system	Result of immune editing.
Repress tumour antigens. Not expressing HLA molecule.	Cytotoxic T lymphocytes are unable to identify the cancer cells.	Hiding Identity
Repress NKG2D ligands	By downregulating the NKG2D ligand the cancer cell hides from the NK cells. The NKG2D ligand represents a cell with dysfunctional DNA activities inducing NK cells recognition	Hide Stress
Destroying, or saturating immunocytes receptors	Destroying or saturating the immunocytes receptors, unable the NK cells or T cells to bind to the cancer cells antigens.	Inactivate immunocytes
Releasing FasL	By releasing FasL cancer cells induce apoptosis of the cytotoxic T cells	Immunocyte killing. Initiate immune cells destruction by apoptosis
Tumour inflammation	By releasing cytokines IL-10 and TGF-b, cancer cells manipulate the immune system to stop the attack and undergo apoptosis.	Immunocyte suppression ending in apoptosis
CD47 upregulation.	By up regulating CD47 expression the cancer cell can avoid phagocytose.	Avoid phagocytosis by phagocytic cells like macrophages
PD1/PDL1 upregulation.	PD1 is an immune checkpoint, When immune cells bind to PDL1 they are stimulated to undergo apoptosis. Some cancer cells express and upregulate the PDL1 to protect themselves.	Initiate apoptosis and immune cell killing.
CTLA4 expression and upregulation	CTLA4 is expressed by activated T cells and transmits an inhibitory signal to other T cells.	Downregulation of the immune response.

It is apparent that the immune system and the cancer cells interact with each other. Being able to quantify and assess the local immune response in the vicinity of a tumour will enhance our understanding of tumour host interactions.

Cancer cells develop and alter their behaviour to avoid detection and destruction. Recent therapies seek at reactivating the immune system to allow the host to eliminate cancer cells (68).

2.4.2 The Tumour Microenvironment

The microenvironment is a biological term for the structures and molecules that surround a given cell. Therefore, the tumour microenvironment (TME) describes the cells and molecules surrounding the tumour (69). TME includes blood vessels, immune cells, inflammatory cells, fibroblasts, and extracellular matrices. The TME interacts with the tumour and influences

progression and proliferation. In return tumours may co-opt their microenvironment to support growth through angiogenesis (70).

During the dynamic stages of cancer progression, the environment shifts from being dominated by self to be a cancerogenic environment driven by the cancer cells. The type, density and spatial location of the immune cells in a tumour are individual to patient and tumour. However, patterns have emerged, and can potentially be used as a prognostic marker.
The latest success of immune checkpoint inhibitors in cancer therapies have revived the interest of the cancer community in studying the tumour microenvironment. Suggesting that immune cell infiltration can potentially be used as a prognostic and/or predictive marker (70).

Recent studies have assessed the relevance of immune infiltration in regard to risk of relapse, clinical subtypes or response to therapy in breast cancers (71). By understanding the role of the tumors environment, immunotherapy may be used for specific breast cancer types.
The infiltration of T-lymphocytes in general is associated with better survival and is thought to be a positive prognostic maker in both ER positive and negative subgroups (70) (72). Higher levels of infiltrated T-lymphocytes also act as a marker for better pathological complete response rate following neoadjuvant therapy (73). Patients with an Her2-enriched tumours receiving targeted therapy against Her2 will respond better with an immunogenic microenvironment (72) .
Regulatory T cells are thought to be recruited by the tumour inflammatory response and are considered a marker for poor response and low overall survival. Similarly, the M0 and M2 macrophages are also associated with poorer outcome in both ER negative and ER positive breast cancers (70). M1 macrophages are associated with a favourable outcome in ER positive cancers (72). A higher infiltration rate of plasma cells is also associated with overall favourable outcome (70). These findings are based on two large studies using gene expression, and a gene signature-based computational method called CIBERSORT (74).

2.4.3 Methods to investigate the tumour microenvironment

The microenvironment in cancer is complex and consists of numerous types of cells. In recent years, understanding the cellular heterogeneity and the environment surrounding the tumour has become increasingly important. It is established that the clinical outcome and therapeutic response is dependent on the infiltration of immune cells (75). In that sense, new computational methods use the bulk tumour gene expression to digitally indicate which cells are present in

the microenvironment. A more conventional method used by pathologist is immunohistochemistry. Immunohistochemistry uses antibodies to label specific surface markers on immune cells to identify leukocytes.

Immunohistochemistry

Immunohistochemistry (IHC) is a microscopy-based technique used to visualise cellular structures and components. IHC uses antibodies to bind and label cellular components. This technology is used among pathologists to investigate the tumour morphology and characteristics, but also its microenvironment.

The principles of IHC is described in Figure 10. Primary antibodies are used to bind a structure or epitope with very high specificity. Secondary antibodies carry signalling molecules which can be a fluorochrome and will allow detection (76). The most common signalling molecule in pathological- clinical setting is an enzyme: peroxidase.

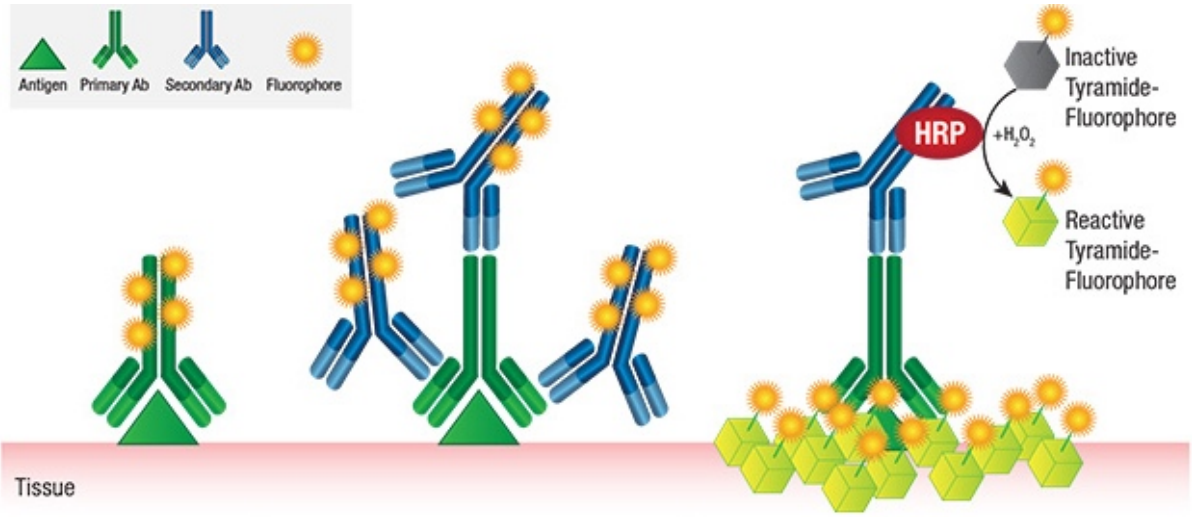

Figure 10. The principle behind immunohistochemistry. Primary antibody binds to epitope. Secondary antibody coupled with a fluorophore binds to the primary antibody. Activation of the fluorophore makes it detectable in microscope. Picture modified from Cell Signal Technology *(77)*.

Deconvolution methods

Deconvolution is the ability to decompose a complex signal into simpler signal, which compose the complex signal. In the recent years several bioinformatical deconvolution tools have become available, making it possible to digitally identify different cell types present in a tumour. Cell type identity is defined by the expression of very specific marker genes which underlie the cell function (75). In the recent years, gene signatures for specific immune cells have been reported and used to estimate specific cell type infiltration using bulk tumour expression data.

ESTIMATE and ISOLATE are two methods calculating the tumour purity and the abundancy of the main types of infiltrated immune cells. ESTIMATE has been integrated into The Cancer Genome Atlas (TCGA) and is a standard pipeline to evaluate the tumour and immune proportions (78). These two tools are useful when looking at tumour purity but gives a somewhat limited understanding of the microenvironment.

CIBERSORT characterises immune infiltration of 22 immune cells from gene expression profiles. CIBERSORT uses gene signatures, in addition to supervised machine learning frameworks through linear regression vectors to identify the proportion of each immune cell type in the microenvironment (74).
Nanodissect uses a large pre-assembled expression compendium and utilises support vector mechanisms to identify genes expressed (markers) only in very specific cell types. In that sense nanodissect may be an interesting tool when interested in a specific cell type. (79).

A recent study using the CIBERSORT reported the association between specific cell type immune infiltration in ER positive and negative breast cancer patients and overall survival (Figure 11) (72).
However, it is still unclear which immune cell type are more or less expressed according to breast cancer subtypes. In the master project we examined immune infiltration across the different subtypes within breast cancers. Using xCell the most recent deconvolution method that allows to account for 64 different cell types.

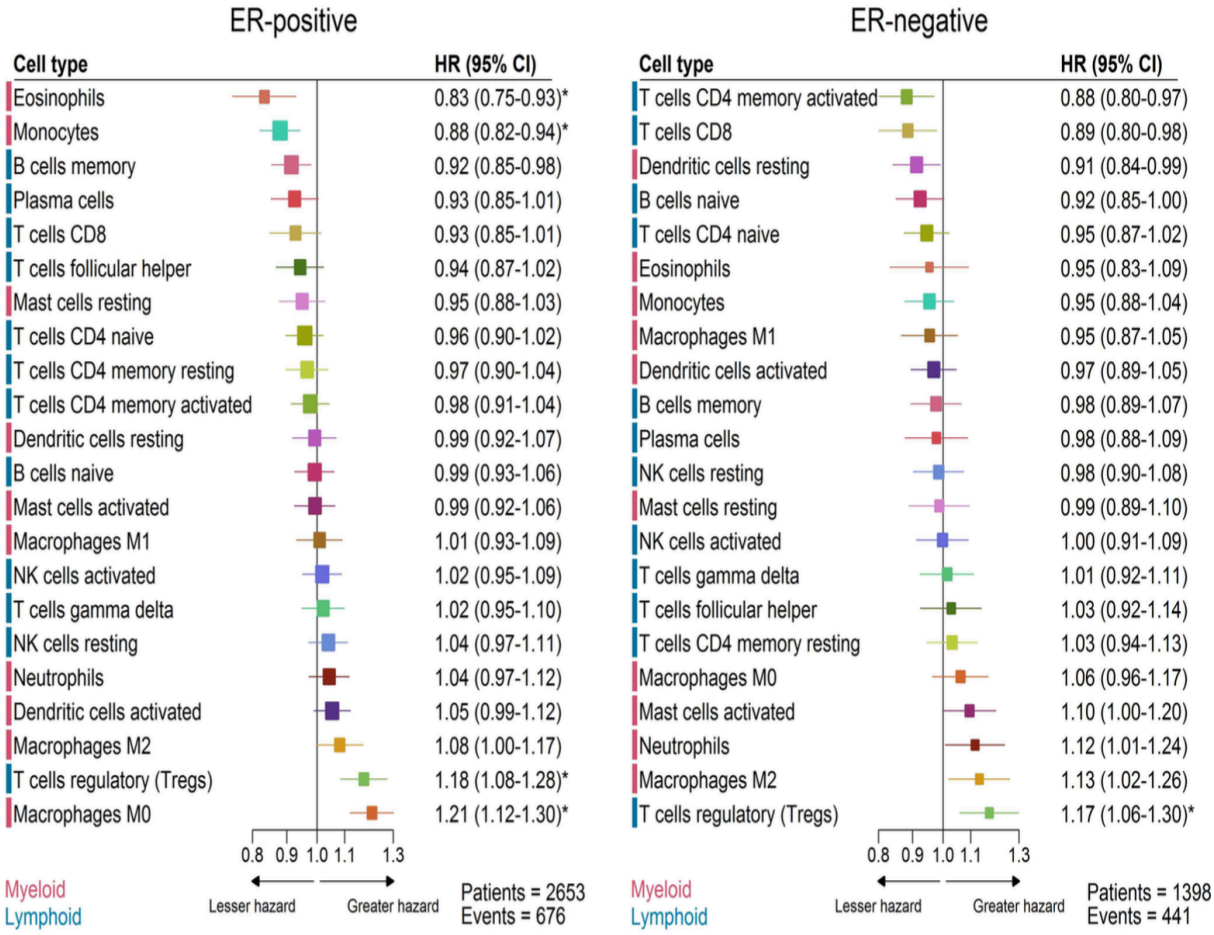

Figure 11. Immune cell infiltration and overall survival according to ER status. Figure from *"Patterns of Immune Infiltration in Breast Cancer and Their Clinical Implications: A Gene Expression Based Retrospective Study"*, By H. Raza Ali. Reprinted with permission *(72)*.

3 Aims of the study

- Investigate how the immune cell infiltration differs within ER positive and ER negative samples.
- Assess immune contexture according to breast cancer molecular subtype.
- Investigate whether cytokine serum levels are related to the levels of tumour infiltrating immune cells.

4 Material

Our study takes advantage of several breast cancer cohorts which have genome wide expression profiles. See Table 4 for a short comparison and description of all the cohorts.

Table 4. 15 cohorts were used in this study. The table shows the number of patients, a brief description and the platform used for gene expression.

Cohort	N_samples	Brief description	Platform
MAINZ	200	Study with lymph node negative patients.	Affymetrix HGU
MDACC	508	Study with samples from tumour core.	Affymetrix HGU
Metabric	1904	A large study with high representation of different types of breast cancer.	Illumina HT-12 v3, Affymetrix HGU
MicMa	104	Early stage breast cancer.	Agilent whole genome 4x44K
NeoAva	106	Clinical trial study on patients with large tumours (T2-T4).	Agilent SurePrint G3 Human GE 8x60K
OsloR	93	Early stage cancer, with operable tumours (T1-T2)	RNA- seq
Oslo2	277	Early stage cancer, with operable tumours (T1-T2)	Agilent SurePrint G3 Human GE 8x60K
STAM	856	Early stage (T1-T2) breast cancer samples.	Affymetrix HGU
STK	159	A study to look at prognostic factors defined by clinical outcome. Wide range of tumour types.	Affymetrix HGU
TAI	327	A study to look at prognostic factors defined by subtype. Wide range of tumour types.	Affymetrix HGU

TCGA	981	A large multidimensional data collection.	RNA-seq
TIF	74	Primary breast cancer patients from Denmark	Illumina HT-12 v3,
TRANSBIG	198	Patients younger than 61 and are lymph node negative. (T1-T2)	Affymetrix HGU
UPP	251	Primary breast cancer.	Affymetrix HGU
VDX	344	Study on metastasis, cohorts with differences in tumour grade and lymph node status.	Affymetrix HGU

4.1 MAINZ

Patients were treated during the time period 1988-1998 at the department of Obstetrics and Gynaecology of the Johannes Gutenberg University Mainz. At least 40% of the mass was cancer cells (80). The raw, normalised gene expression data and clinical information on the samples were collected from Gene expression omnibus (GEO), access number GEO: GSE11121.

4.2 MDACC

The MDACC study is a population-based study of 508 patients from Houston, Texas. The samples were obtained by a fine needle aspiration or core biopsy, prior to any treatment. Gene expression profiles were then performed at the department of pathology at the M.D. Anderson Cancer Centre (81). In addition to gene expression profiles, clinical and patient information were collected from GEO, access number: GEO: GSE25066.

4.3 METABRIC

The Metabric cohort is compiled by 1904 fresh frozen breast cancer specimens from tumour banks in the United Kingdom and Canada. The collection of samples is broad and in different grade, stage and molecular markers. Gene expression profiles were preformed, and clinical data collected (82). The gene expression data are deposited at EGAD with access number EGAD00010000210.

4.4 MicMa

The MicMa cohort is a sub cohort of the DNK study OSLO1 (with blood, bone marrow and tumour tissue collection) fresh-frozen tumour tissue were available from 123 patients. Patients were all diagnosed with early stage breast cancer. Genomic and transcriptome profiling has been done for 104 patient tumour samples. Cytokine serum levels were analysed. Information about tumour and patient were also collected into a clinical file and is available at GEO, with access number: GSE19536

4.5 NeoAva

The cohort is comprised of 106 patient samples with large primary breast tumours (>2,5 cm) (T2-T4). Patients were treated between November 2008 and July 2012 at Oslo University Hospital or St. Olav's Hospital, Trondheim (83). The clinical and gene expression data are available from Array Express: EMTAB-4439.

4.6 OSLOR

The OSLOR is a sub population of the Oslo2, where 93 samples were profiled with RNAseq. This data was access from Radium hospital.

4.7 OSLO2

The Oslo2 study is a consecutive study with 277 patient samples collected from hospitals in south-eastern Norway. The patients are in early stages (T1-T2) (84). Gene expression profiling was done using the Agilent SurePrint G3 Human GE 8x60K. The gene expression as well as clinical information can be collected from GEO with access number: GSE58215.

4.8 STAM

The STAM cohort consists of 255 early stage (T1-T2) breast cancer samples. The patients were diagnosed during the time period 1980-1995 from patients at the John Radcliffe hospital, Guys hospital in United Kingdom and from Uppsala university hospital in Sweden (85). The gene expression as well as clinical information can be collected from GEO with access number: GSE6532.

4.9 STK

The STK cohort contains 159 samples of population-derived breast cancer patients from Sweden. The study was set up to identify patients that will respond to treatment. The cohort is compiled of a broad range of samples (86). The gene expression and the clinical file can be collected from GEO, with access number: GSE1456.

4.10 TAI

The TAI cohort consists of 327 samples from patients. The study was performed to optimise treatment through subtyping as a prediction tool (87). The gene expression and clinical file can be accessed from GEO, with access number: GSE20685.

4.11 TCGA

The TCGA cohort is formed from a collaboration between the national cancer institute (NCI) and the national human genome research project (NHGRI) and is publicly available through the TCGA data portal. The data were collected from the TCGA data portal. Level 3 RNA-seq data were used in this study.

4.12 TIF/DCTB

The TIF cohort consists of 74 serum and tumour infiltrated fluid samples, collected from patients at the Danish centre for the Transnational breast cancer research program from 2003 and 2012. Staining and Immunohistochemistry was performed on tumour samples, and immune infiltration was measured. Cytokine profiling of 27 cytokines using the Luminex technology was performed on pre-treatment serum samples for 28 samples. The gene expression was available to us through our collaboration with the Danish group. The cytokine profiling was performed at Akershus university hospital, Oslo and these data will be published with the manuscript in preparation.

4.13 TRANSBIG

The TRANSBIG study was performed from samples from a frozen archive collected from 1980-1998. The samples were collected in 6 different hospitals in Sweden, France and England. The patients were younger than 61, they were lymph node negative and had a tumour grade between T1-T2 (88). The raw, normalised gene expression data, in addition to clinical information on the samples were collected from GEO access number GSE7390.

4.14 UPP

The UPP cohort is a population-based study compiling 215 samples collected from patients primarily from Uppsala county in Sweden. The samples were collected in the time period 1987-1889. The samples collected presented primary breast cancers **(89)**. Data file and clinical information can be accessed from GEO: with access number: GSE3494.

4.15 VDX

The VDX cohort consists of 344 samples, which were a part of a study to investigate metastasis. Gene expression profiles can be used to estimate the ability to metastasise. Three breast cancer cohorts, MSK-99, NKI-295 and EMC-344 were used in this study. The NKI-295 and EMC-344 consists of early stage breast cancer patients. The MSK-99 were comprised of more mature, larger tumours with tumour grade T2-T4, and the majority of these samples were node positive (90). The expression data and clinical files can be accessed through GEO, with access number: GSE2034/GSE5327.

5 Methods

5.1 Gene expression profiling

Two main methods can be used to profile expression: microarray and RNA-seq. This section contains a short general description on how the gene expression profiling is performed.

Tumours are usually divided into two sections on which different histological analyses are performed. RNA extraction is performed on fresh frozen tumours.

5.1.1 RNA isolation

Fresh frozen tumour samples are first lysed and homogenized to breach the plasma membrane and access the RNA from the cell and nucleus. In our lab the technique mostly used to extract RNA is the RNeasy mini spin columns method.

5.1.2 Whole genome expression microarray

Microarray gene expression analyses is based on hybridisation of RNA to specific probes on a solid surface. Before the hybridisation process the samples are labelled with a fluorophore. The samples are then amplified. The intensity of the fluorophore captured after hybridisation will reflect the amount of expression of a particular gene as illustrated in Figure 12. The gene expression from the different cohorts were done on different platforms, listed in Table 5. Most of the cohorts profiling were done by microarray. All the cohorts were normalised to be comparable regardless of the platform used.

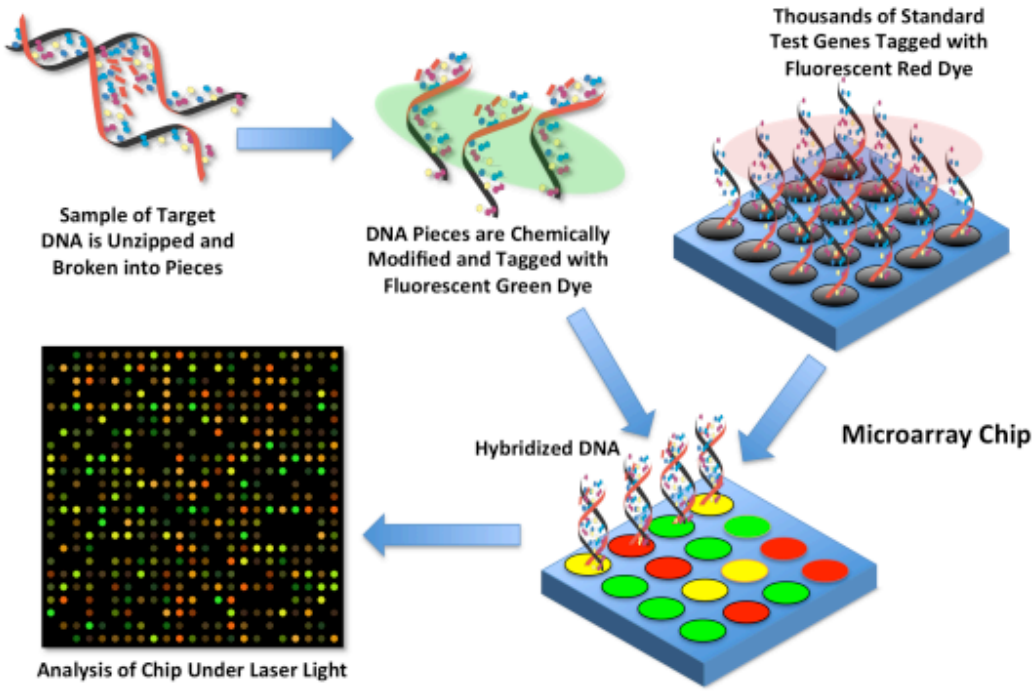

Figure 12. Microarray expression analysis. Picture modified from gene profile´s website *(91)*.

5.2 Statistical and bioinformatics analysis

5.2.1 Statistical language and computational program R

Statistical analyses were done in R, a statistical computation and graphic software. To operate the program a statistical language called R is used. The program can be downloaded for free. R provides a large assortment of statistical and graphical techniques, like classical statistical tests, classifications and clustering. In addition, several packages can be downloaded and applied to the program.

The software used during the analyses was, R version 4.3.3 © 2009-2018 RStudio, Inc.

5.2.2 Pre-processing

Expression data were normalised according to the host lab preferred method. In our lab, microarray data were log2-transformed and quantile-normalised, while RNA-seq data were analysed using cufflinks v2.2.2 with default parameters to obtain fragment per kilobase million (FPKM) value. Further, each cohort was scaled and centred which practically means that the mean and standard deviation of each gene across the sample of a cohort would equal zero. Therefore, each gene expression would be on the same scale for each cohort to allow accurate comparison of the cohorts between each other.

5.2.3 Molecular subclassification into PAM50 subtypes

The PAM50 molecular subclassification was performed using the Genefu R package. Genefu is a platform independent and uses gene expression data to identify the PAM50 subtypes. When ER status was not available we used a two-component Gaussian finite mixture model using maximum likelihood estimation, as previously described (92).

5.2.4 xCell

xCell is a recently published novel gene signature based on computational method, which allows to score for 64 different cell types (Figure 13) while inputting the normalised expression data. Authors harmonised 1822 pure human cell types transcriptomes from various sources and employed a curve fitting approach for linear comparison. This method relies on a group of signatures for each cell type made from multiple sources limiting contamination and creating a robust specific cell signature. This method can be used to predict the microenvironment with a high confidence level. Since the xCell method is based on gene signatures this allows us to predict the microenvironment based on the expression data from different platforms. Signature based methods have in the past had problems with distinguishing cells of similar lineage. The

xCell authors applied a novel technique to remove spillover between closely related cells (93). We used the xCell R package which can be downloaded through GitHub to score all the cohorts described above.

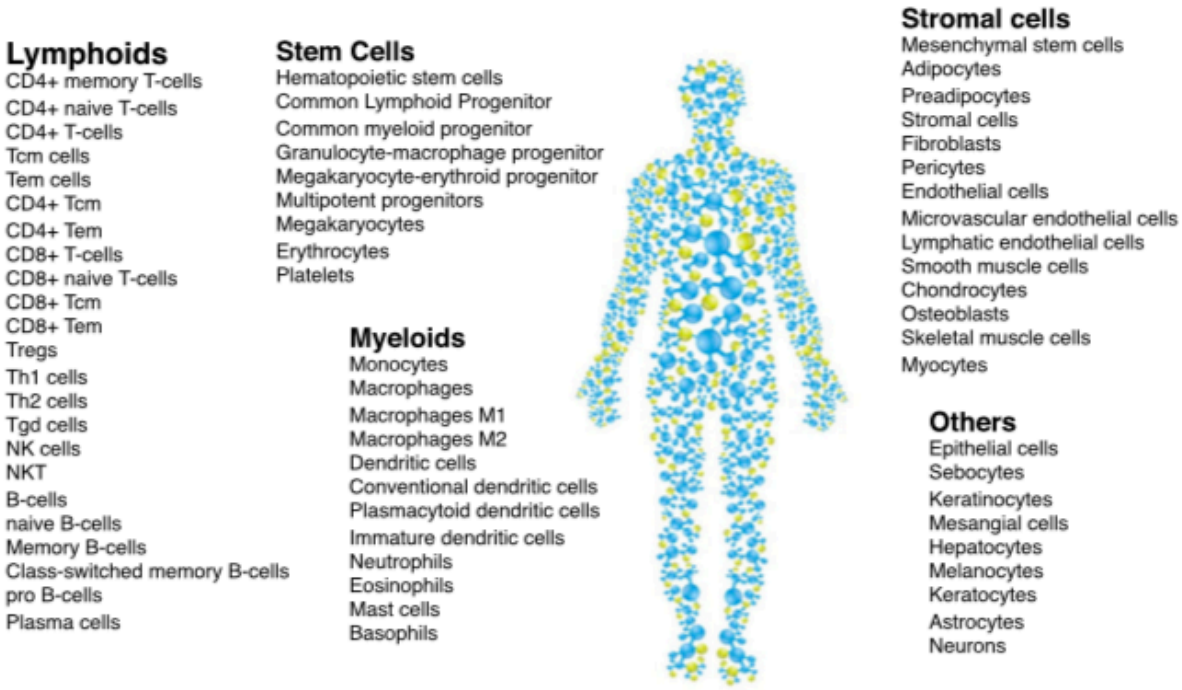

Figure 13. Cell types recognised and identified by the xCell tool. A description of all the cells profiles is described in appendix A. Picture modified from the xCell paper *(93)*.

5.2.5 Statistics

t-test

t-test is a common statistical analysis used to assess whether the means of two groups are equal to each other. The t-test assumes that both sample groups are normally distributed with equal variances. The null hypothesis is that the two means are equal. When doing statistical analysis the p value helps determine the significance of the result. Identifying if the null hypothesis in the statistical analyses is validated or not. The p value is a number between 0 and 1. A small p value typically < 0.05 indicates that the null hypothesis is rejected and that the two means are significantly different. A large p value > 0.05 indicates that the two means are likely to be equal.

Kruskal-Wallis test

The Kruskal Wallis test is a nonparametric method to test whether the median of samples are significantly different. The test does not assume normal distribution of the residuals, allowing less parametric strains on the groups to be equal. The test null hypothesis is that the median of

all groups are equal. We used the Kruskal-Wallis test to assess the significant difference between IHC and xCell scores.

Unsupervised clustering

Clustering is used to group samples, so that the samples that are most alike get grouped together. There are several different ways to cluster using different algorithms and parameters. Ward D. is a hierarchical cluster analysis using an agglomerative algorithm, assigning each sample its own cluster before combining clusters together, based on the optimal value of an objective function, ending up with one hierarchical cluster. We used this clustering method when making heatmaps, to identify clustering within cell types and subtypes across the heatmap.

Pearson correlation coefficient

Pearson correlation coefficient is a measurement of the linear correlation between two variables. The rho value of a Pearson correlation ranges from +1 and -1. +1 indicates perfect positive correlation, while -1 reflects perfect negative correlation and 0 no correlation between the two variables. We used Pearson correlation to assess the correlation between cytokine levels and xCell scores.

Correlation plots were used to visualise correlation. Corrplot is a package in R using graphical visualisation to evaluate correlation matrix. The rho value of the correlation is displayed in colours, positive correlations are displayed in red and negative correlations in blue. Colour intensity and size of the circle are proportional to the correlation.

Mann-Whitney test

Mann-Whitney test is used to assess significant differences between two independent groups. The test is nonparametric and does not assume equal distribution. The test is used when analysing dependent variables that are ordinal or continuous. We used this test in Figure 29.

Heatmap

A heatmap is a graphical representation of statistical data, where the individual values are represented by colours. Heatmaps are often used when analysing large multidimensional data sets, because it allows visualisation of many data points. We used the pheatmap package in R to visualise unsupervised clustering through heatmaps.

Boxplots

Boxplots are used to graphically illustrate groups with numerical data, displaying the spread and median of the data point within a group. We used the boxplot function in R to generate the boxplots comparing infiltration of cell types according to the ER status, subtypes, and cytokine levels. The whiskers display the maximum and minimum data points. The box itself shows the first to the third quantile. The line across the box is the median.

5.2.6 Immunohistochemistry for the TIF Cohort

Formalin fixative paraffin-embedded blocks were prepared from 2-3 various parts of the tumour tissue and was used to make Super Frost Plus slides. The tissue sections were stained CK19 (KRT19) antibody to evaluate tumour cell content and tumour stroma percentage. Antibodies recognising CD3, CD4, CD8, CD45 and CD68 were used to assess specific immune cell infiltration (94). We used the staining to validate the xCell scores.

5.3 Cytokine profiling

Cytokine profiling was performed using the multiplex Luminex xMAP platform. Luminex is a multi-analyte profiling tool used to detect and quantify multiple secreted proteins such as cytokines, chemokines and growth factors. Luminex is bead-based and provides accurate reproducible results for target analyte. It has two detecting lasers; one discriminates the microsphere type, the analyte, the other detects secondary antibody and sample quantity. In the bead-based methods, microspheres are labelled with a panel of specific antibodies. Each antibody binds to a specific target protein. The secondary antibodies bound to a biotinylated and phycoerythrin binds to the analyte. Lasers detect both the microsphere identity dye and the fluorescence, identifying the analyte and the quantity, see Figure 14. Up to 50 user defined target proteins can be profiled simultaneously, from cell culture supernatants, serum, or plasma samples. Small amount of sample volume is needed to do analyses. Bioplex is the software used to operate the Luminex machine and extract the analytical data.

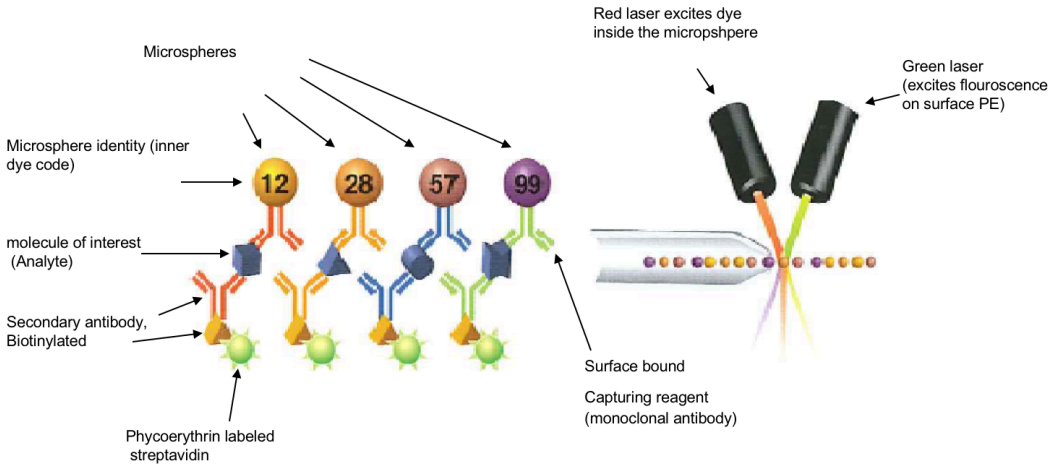

Figure 14. Principles behind Luminex platform. Microspheres labelled with a corresponding antibody bind the analyte, the molecule of interest. The secondary labelled antibody binds to the bead-protein-antibody complex. Two detection lasers excite microsphere dye and secondary antibody fluorescence. Yielding information of which analyte is present and the quantity. Reprinted with permission from Biorad.

5.3.1 Luminex calibration and validation

The machine is calibrated every month. Calibration is done to ensure that the result is exact, and the machine is operating correctly. The calibration is performed using the BioRad calibration kit. Set reagents from the kit as in Figure 15, is used according to the Bioplex calibration setup. Our calibration procedure passed. The validation is done by using the BioRad validation kit, and is done using the sample plate as in the calibration process. Reagents are applied to the appropriate wells validating; optics, carryover/fluidics, receptors and classification. During the experiment the user can also choose whether he wants to include a validation well to the setup. Validation was performed before every experiment.

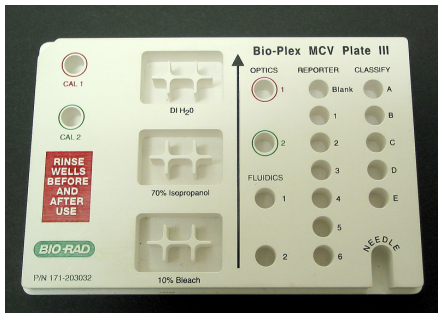

Figure 15. Picture of the validation and calibration plate. Picture taken in the lab.

5.3.2 Cytokine workflow

The protocol and the workflow (Figure 16) were followed as instructed by the supplier. Serum samples were diluted to a 1:3 ratio using a dilution solution. A series of dilutions of known concentrations of each analytes were assessed along with the unknown serum samples. These provide a standard curve to calculate the exact concentration of each analyte.

Before starting assay, protocol was made using the Bioplex assay protocol. Indicating which well the samples should be applied to. The workflow indicated the steps and in which order the steps should be completed (Figure 16).

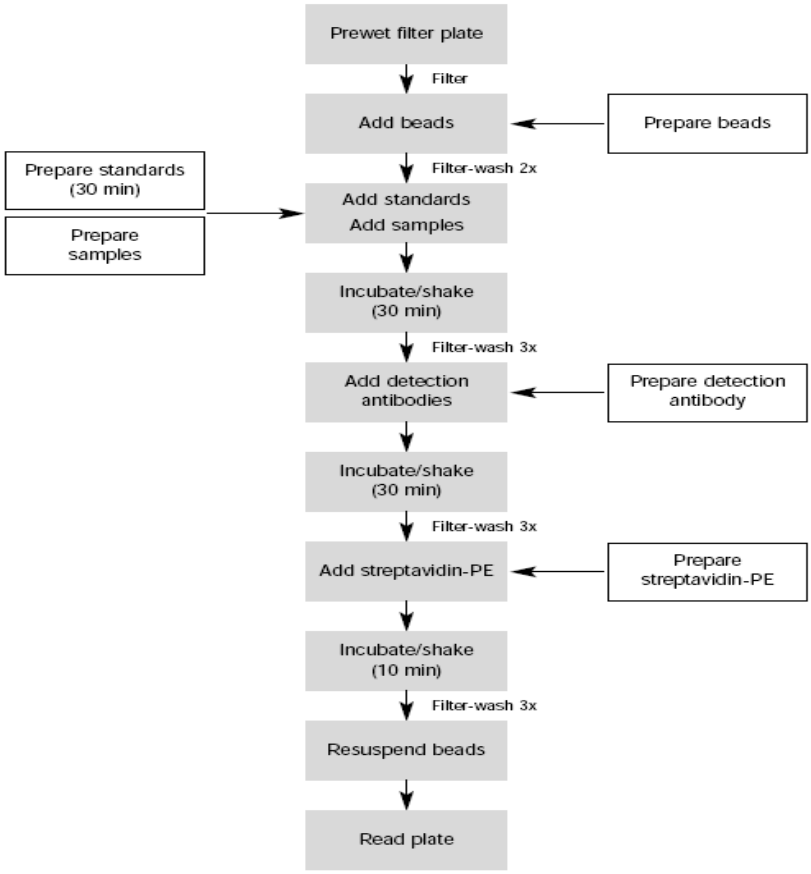

Figure 16. Workflow for cytokine profiling. The figure shows the sequence in order to do successful and accurate cytokine profiling.

5.3.3 Prewet filter plate

The magnetic washer plate bottom was prepared to hold a 96 well plate. The wells were washed by adding 100 µl wash buffer. The plate is then placed on the magnetic rack for 45 seconds to settle the beads. Paper towels were placed on the well surface before turning the plate upside down. Washing buffer was removed by plate inversion.

5.3.4 Bead preparation

Microspheres containing an identity dye code are selected to identify specific analytes; chemokines, cytokines and growth factors, Table 2. A 10x magnetic bead solution was carefully vortexed at mid speed for 30 seconds. 5175µl assay buffer pipetted out to a 15 ml tube. 574µl of the bead solution is added to the buffer tube, creating a 1x bead solution. 5µl buffer bead solution was added to each well, 45µl buffer solution was then added to make a total of 50 µl.

5.3.5 Sample and control preparation

The serum samples and the control provided by the supplier were collected filtered and centrifuged at 10 000 g, at 4° C for 10 min to remove platelets and precipitates. All dilutions were made on ice. 150 µl of sample diluting buffer was added to 50µl of samples or control, according to the Luminex setup.

5.3.6 Standard preparations

The vial containing the lyophilised standard was first inspected to ensure that the pellet was at the bottom of the vial. 500 µl of standard diluent was added to the vial and vortexed fore 1-3 seconds before incubating on ice for 30 minutes. A dilution series (standard curve) was performed as indicated in Table 5.

Table 5. Dilution series setup used under the experiments.

Standard ID	Standard Diluent (µl)	Amount to next standard tube(µl)
Reconstituted standards	500	128
S1	72	50
S2	150	50
S3	150	50
S4	150	50
S5	150	50
S6	150	50

S7	150	50
S8	150	50
S9	150	0
Blank	150	

5.3.7 Application samples and standards

The wells containing the beads were washed twice by adding 100 µl wash buffer before incubation on the magnetic plate. The plate was then turned upside-down to remove the buffer and keep the magnetic beads in the wells. Standards and samples were then applied to the wells, the plate was sealed, and placed on a filter plate on a microplate shaker and covered with aluminium foil. The speed was increased slowly to 1,100 rpm, it was maintained for 30 seconds before it was adjusted down to 700 rpm. The plate was incubated on the shaker for 60 minutes at room temperature.

5.3.8 Preparation and addition of detection antibodies

It is important that antibodies are used within 15 minutes of preparation. The required volume of detection antibody diluent was pipetted out to a 15 ml tube. 2700µl for 96 wells of detection antibodies were then vortexed for 20 seconds, at medium speed, then spun down, and the bottom phase 300µl was collected. 1,25 µl of detection antibodies diluted to a 20x with a final volume was pipetted in each well. The plate was then sealed, placed on a shaker, the speed was increased to 1100 rpm for 30 seconds and then stabilised at 700 rpm for 30 minutes.

Preparation of Streptavidin-PE

Streptavidin-PE was vortexed at 20 seconds at medium speed, then spun down and the bottom of the vial was collected. 60µl streptavidin-PE were added to 5940 µl of assay buffer and vortexed for 3 seconds.

After the incubation period with the antibodies the plate is washed three times, as previously described. 50 µl diluted streptavidin is then added by multichannel pipet. The plate was then sealed and set to incubate on a shaker for 10 minutes. After the incubation period the plate is washed three times using the same protocol as before. 125µl of assay buffer was added to each

well. The plate was incubated on a shaker at room temperature at 1100 rpm for 30 seconds. The plate was then analysed using the following protocol as seen in Table 6.

Table 6. Standard program parameters for running the Luminex platform.

Dilution factor	4
RIP: (CAL2 Low RIP target)	Standard PMT
Bead counts	50 beads per region
DD gates: low, high	5,000-25,000
Types of analysis	Quantitative, 5 PL curve fit
Unit	pg/ml

6 Results

6.1 Validation of the xCell scores using IHC

At first we compared the xCell outputs to IHC results. IHC has previously been performed on the DCTB/TIF cohort to estimate the type of leukocyte infiltration present in the tumours, notably, cytotoxic T-lymphocytes (anti-CD8+ antibodies), tumour associated macrophages (TAMs);anti-CD68 antibodies) and CD4+ antibodies were assessed. We therefore examined the relationship between intensity of CD4, CD8 and CD68 staining and xCell scores for immune cell types expressing these surface markers. We compared the pathologists scoring for each staining with the output from xCell in Figure 17. To be able to compare the IHC with xCell, all CD4, CD8, or CD68 cells, scores were pooled and compared to pathologist scoring according to staining, see Table 8. Our results indicate that xCell scoring is a good surrogate to investigate immune cell infiltration.

Table 7. Table showing the grouping of cells according to CD markers to be compared to pathologist scoring.

CD4+ staining	CD8+ staining	CD68+ staining
CD4 T cells	CD8 T cells	Macrophages
Regulatory T cells		
Th2 cells	NK cells	M1 Macrophages
Th1 cells		
Tgd cells	NKT cells	M2 Macrophages
CD4 naïve T cells	CLP	
CD4 Memory cells	CD8 naïve T cells	Monocytes
CD4 Tcm cells	CD8 Tem	
CD4 Tem cells	CD8 Tcm	

Figure 17

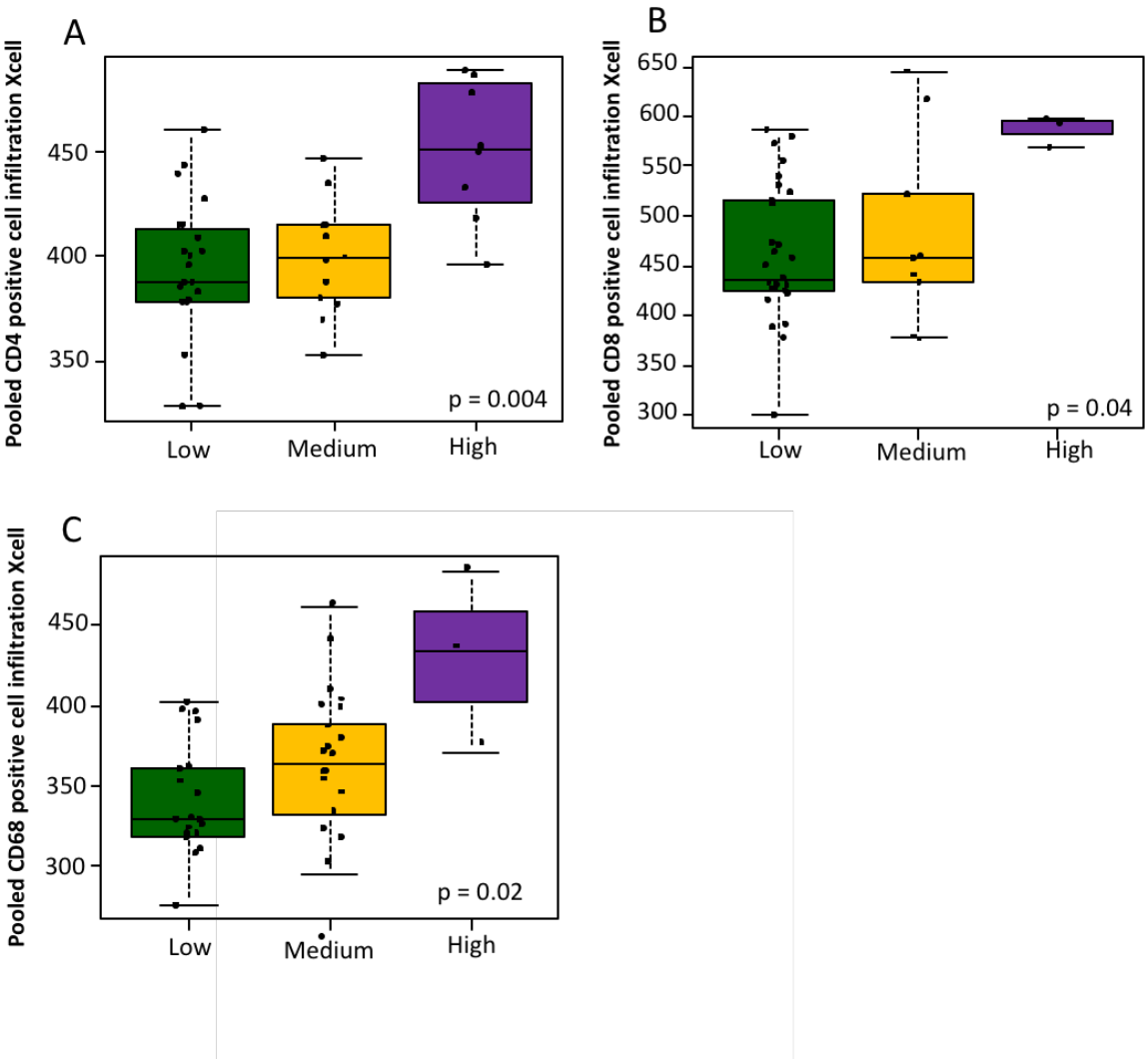

Figure 17. Comparison of xCell score to IHC staining. The x-axis indicates the amount of staining given by pathological inspection. The y-axis indicates the averaged xCell scores for the corresponding cell types as described in Table XX. **A:** CD4 positive cells, **B:** CD8 positive cells, **C:** CD68 positive infiltration.

6.2 Immune infiltration according to ER status

We compared xCell scores according to ER status. ER status is often determined by pathologist using IHC. However, for some cohorts the ER status was not available, and we also observed that the criteria to define ER positivity was not always the same (1 or 10 % of positive tumour cells may be the criteria for ER positivity). Therefore, for more homogeneity, we used a two-component Gaussian finite mixture model using maximum likelihood estimation to determine ER status.

The median of each xCell score were first computed for each cohort according to ER status, and then subjected to unsupervised clustering. The heatmap Figure 18A, shows that ER+ and ER- clustered separately, according to the median of the xCell scores. We observed higher scores in ER- samples for immune cell types (top of the heatmap), these cells are mainly CD4 and CD8 cell types, which confirm a higher overall immune infiltration in ER- samples.

While comparing the mean of each cell type according to ER status using t-test, we found cell types significantly different in all cohorts. The boxplots in Figure 18B, using the Metabric cohort shows 6 cell types which are always (all cohorts) significantly higher scores in ER negative tumours. Furthermore, in Appendix B, we summarised for each xCell type whether the comparisons of the scores across cohorts were significant or not, and if the score was higher in ER+ or ER-.

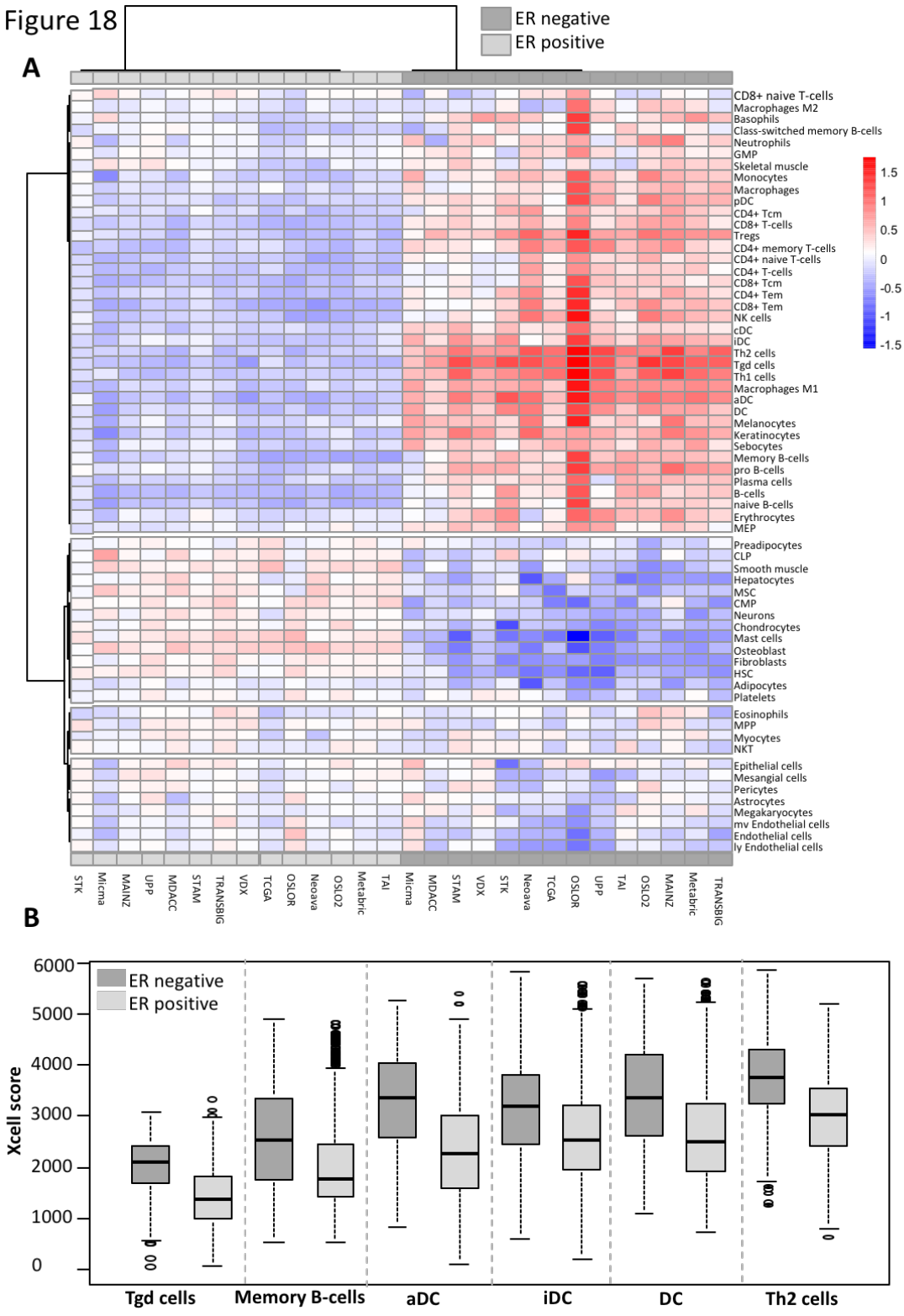

Figure 18. Comparison of xCell scores for each cell type according to ER status. **A.** Unsupervised clustering of the median of each score in each cohort according to ER status. Annotation on top and bottom of the heatmap shows ER- (dark grey) and ER + (light grey) medians. **B**. Boxplots indicating the level of immune infiltration between ER positive and negative samples in selected cell types, for the Metabric cohort. The x-axis shows the cell type. The y-axis indicates the degree of immune cell infiltration according to the xCell score.

6.3 Immune infiltration according to PAM50 subtypes

Being encouraged by such clear differences in xCell scores according to ER status, we further used the same procedure across the PAM50 subtypes. To our knowledge, it has not yet been reported which specific immune cell types may be found in a subtype compared to the other. We therefore first subtyped each cohort according to PAM50 using the Genefu package as described in material and methods. For each cohort we then determined the median of the scores for each cohort across the PAM50 subtype.

- (i) The medians are subjected to unsupervised clustering. Such analyses allowed us to visualise whether the xCell scores were sufficient to group the subtypes together, and which cell types were mainly different within subtypes. These heatmaps are shown in the main figures.
- (ii) We used t-test to seek whether each score was significantly different in each of the 15 cohorts. If the p value < 0.05 we reported in which subtype the score was higher. These results are reported in the Appendix C.
- (iii) Based on the t-test analysis, cell types which are always (all cohorts) up or down regulated in one subtype compared to the other, we reported in the form of boxplots in the main figure. In cases where not all cohorts were significant, top three results were chosen to be visualised in a boxplot (see Appendix C for a comparison of all subtypes and scores).

6.3.1 Luminal A vs Luminal B

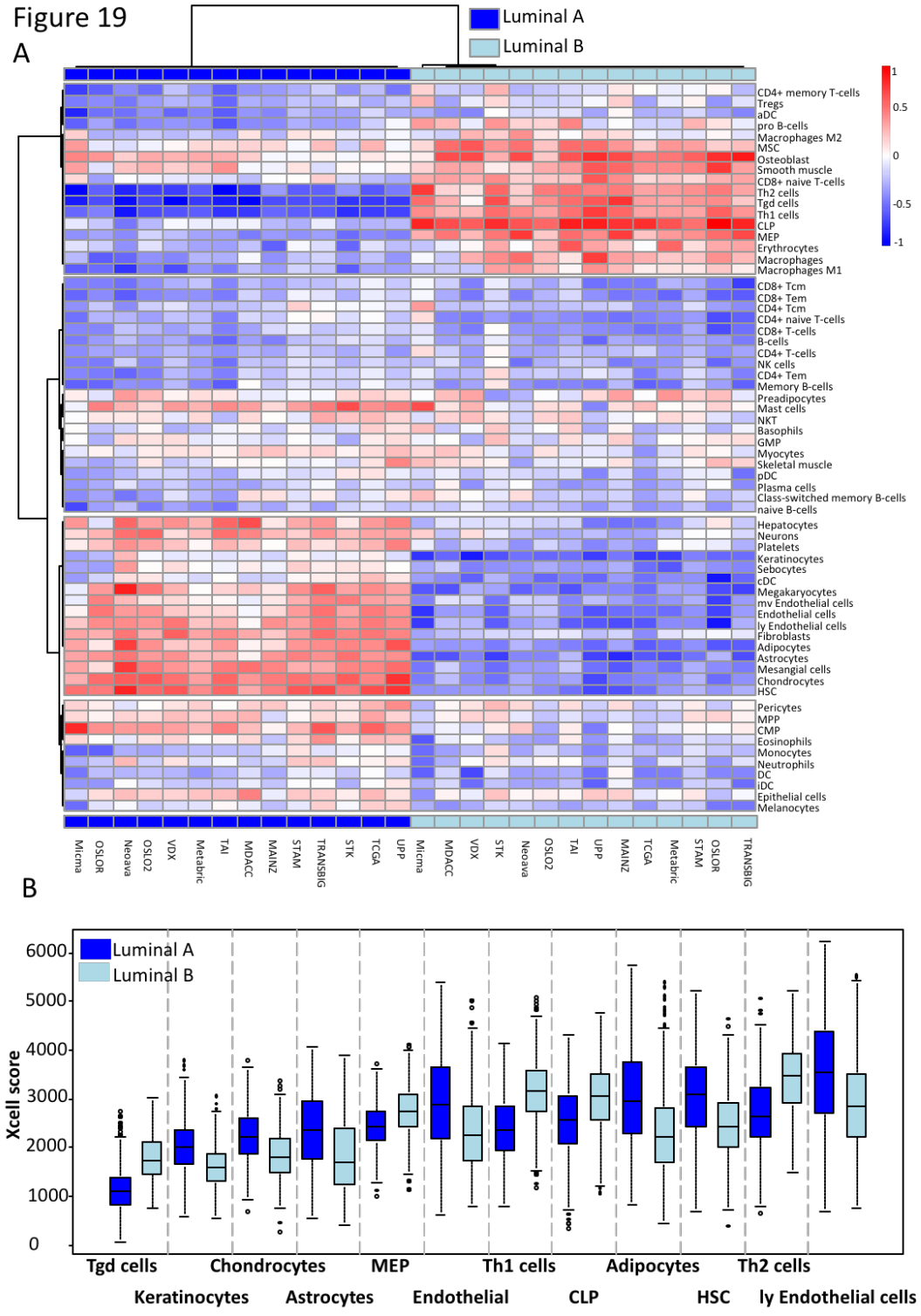

Figure 19. Comparison of xCell scores for each cell type according to Luminal A and Luminal B subtype. **A.** Unsupervised clustering of the median of each score in each cohort according to Luminal A and Luminal B subtypes. Annotation on top and bottom of heatmap shows Luminal A (blue) and Luminal B (light blue) medians. **B**. Boxplots indicating the xCell scores according to Luminal A and Luminal B in the Metabric cohort. Cell types shown in the boxplots are always up or down-regulated in a Luminal subtype compare to the other.

6.3.2 Luminal A vs Her2-enriched.

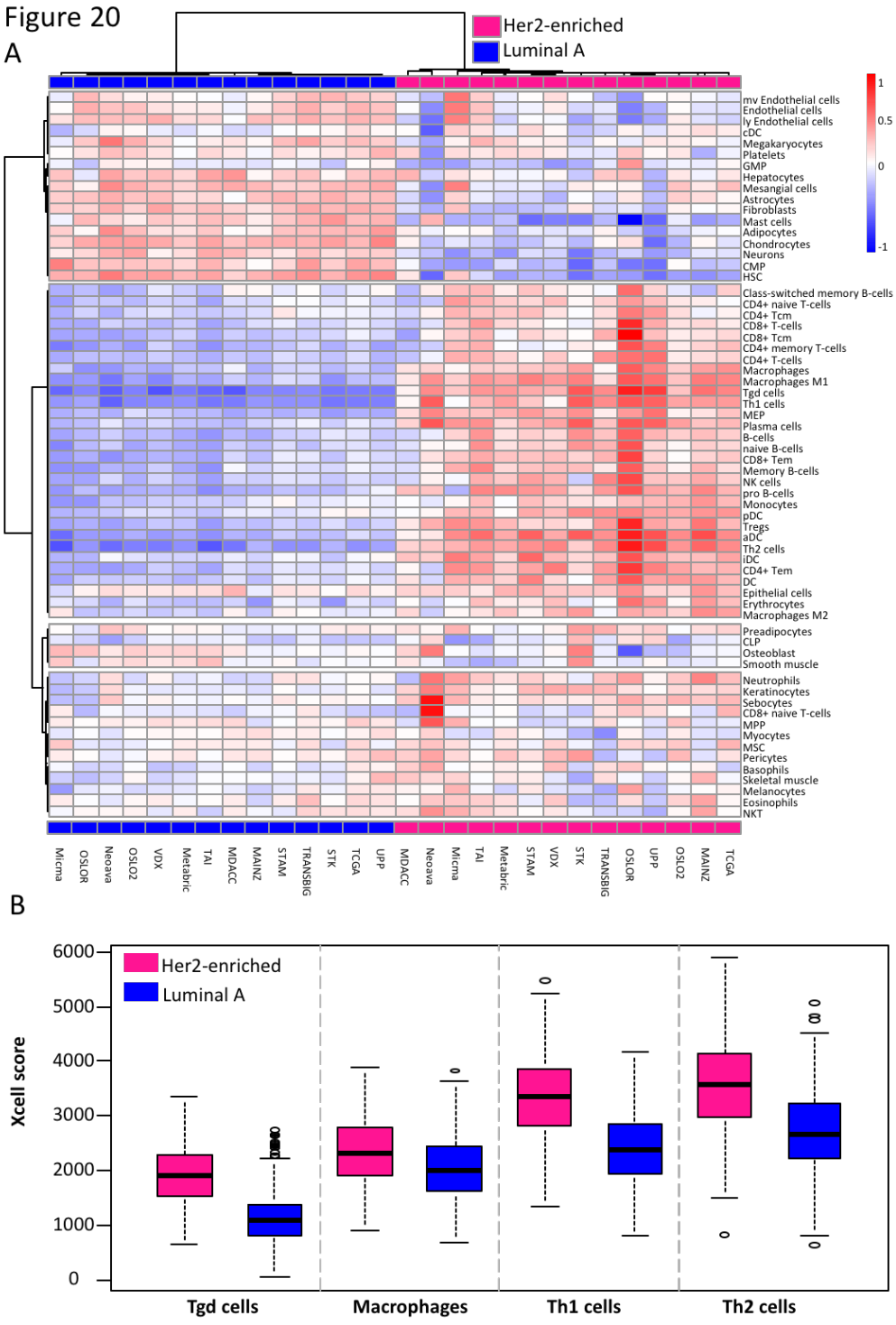

Figure 20. Comparison of xCell scores for each cell type according to Luminal A and Her2-enriched subtype. **A.** Unsupervised clustering of the median of each score in each cohort according to Luminal A and Her2-enriched subtypes. Annotation on top and bottom of the heatmap shows Luminal A (blue) and Her2-enriched (pink) medians. **B**. Boxplots indicating the xCell scores according to Luminal A and Her2-enriched in the Metabric cohort. Cell types shown in the boxplots are always up or down-regulated in Luminal or Her2-enriched subtypes.

6.3.3 Luminal A vs Basal-like

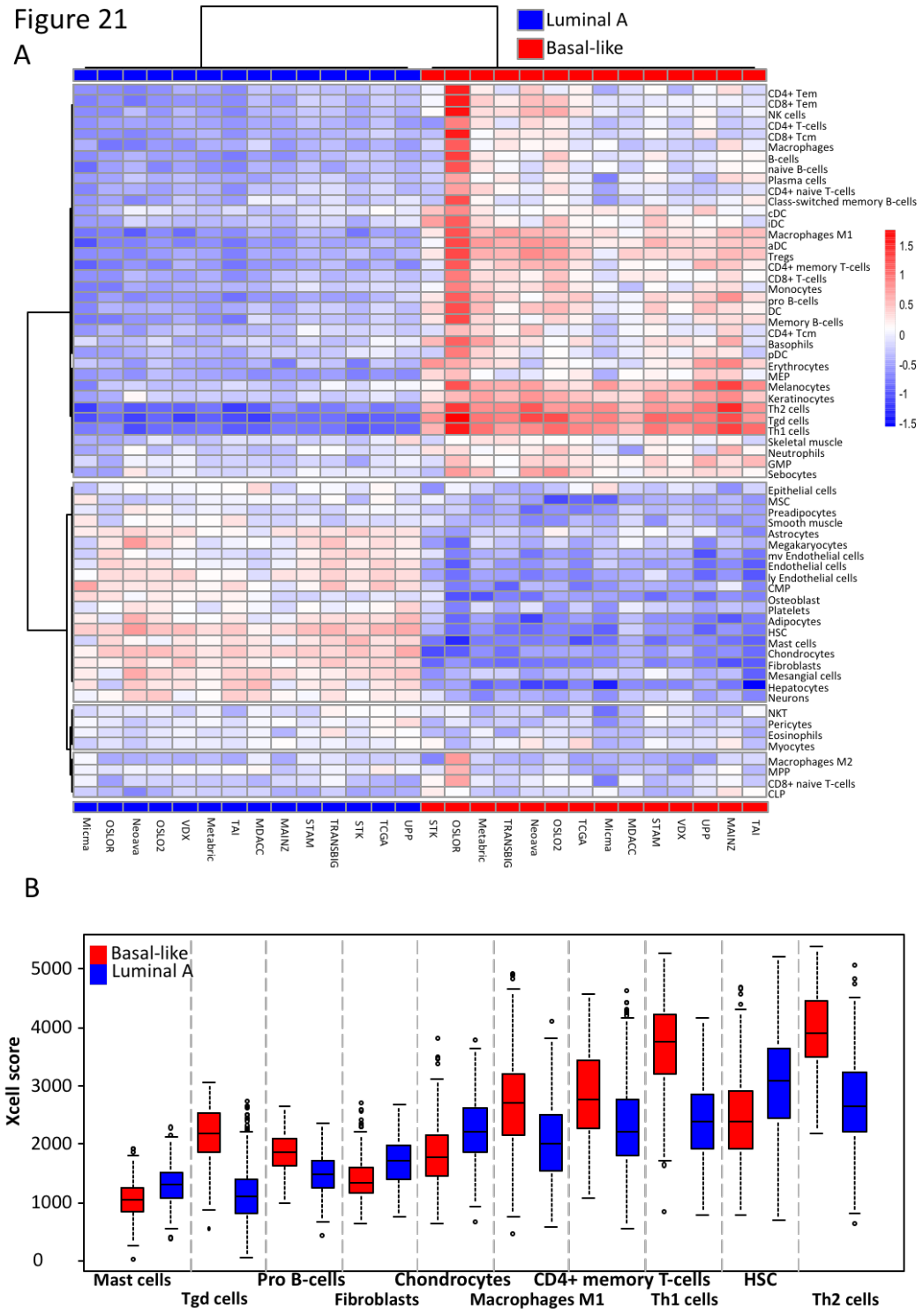

Figure 21. Comparison of xCell scores for each cell type according to Luminal A and Basal-like subtype. **A.** Unsupervised clustering of the median of each score in each cohort according to Luminal A and Basal-like subtypes. Annotation on top and bottom of the heatmap shows Luminal A (blue) and Basal-like (red) medians. **B**. Boxplots indicating the xCell scores according to Luminal A and Basal-like in the Metabric cohort. Cell types shown in the boxplots are always up or down-regulated in Luminal or Basal-like subtypes.

6.3.4 Luminal A vs Normal-like

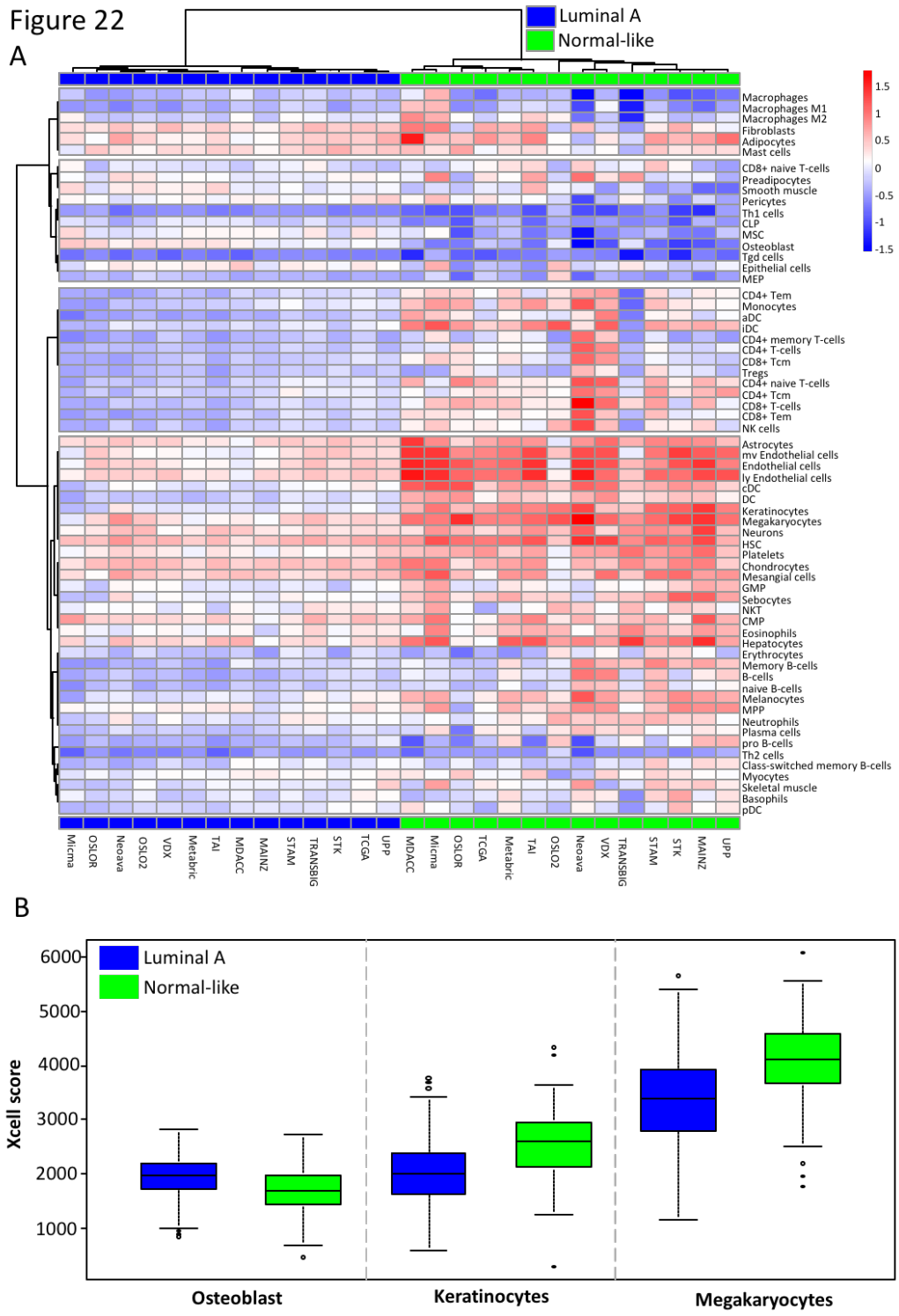

Figure 22. Comparison of xCell scores for each cell type according to Luminal A and Normal-like subtype. **A.** Unsupervised clustering of the median of each score in each cohort according to Luminal A and Normal-like subtypes. Annotation on top and bottom of the heatmap shows Luminal A (blue) and Normal-like (green) medians. **B**. Boxplots indicating the xCell scores according to Luminal A and Normal-like in the Metabric cohort. Cell types shown in the boxplots are always up or down-regulated in Luminal or Normal-like subtypes.

6.3.5 Luminal B vs Normal-like

Figure 23

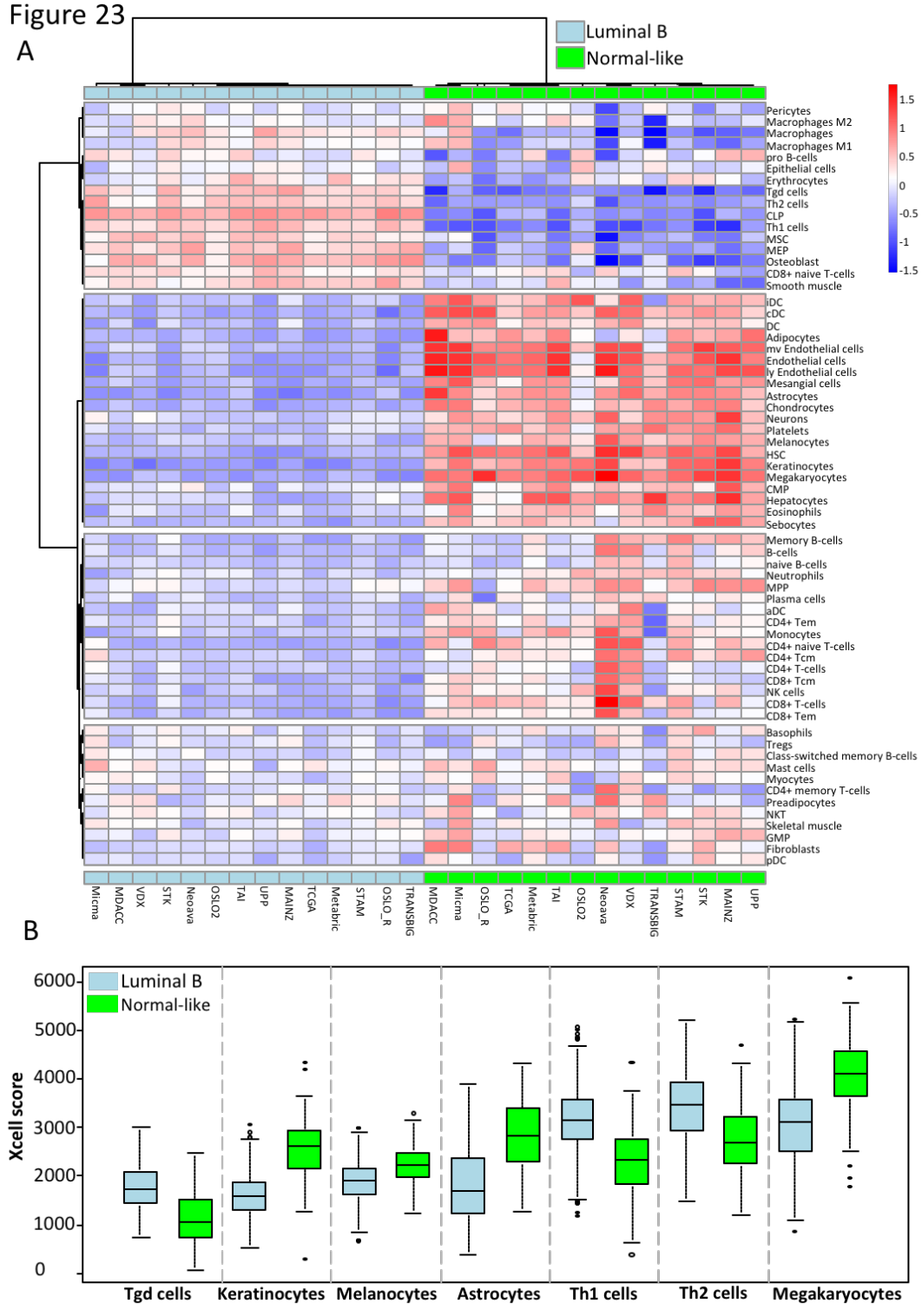

Figure 23. Comparison of xCell scores for each cell type according to Luminal B and Normal-like subtype. **A.** Unsupervised clustering of the median of each score in each cohort according to Luminal B and Normal-like subtypes. Annotation on top and bottom of heatmap shows Luminal B (light blue) and Normal-like (green) medians. **B**. Boxplots indicating the xCell scores according to Luminal B and Normal-like in the Metabric cohort. Cell types shown in the boxplots are always up or down-regulated in Luminal or Normal-like subtypes.

6.3.6 Luminal B vs Basal-like

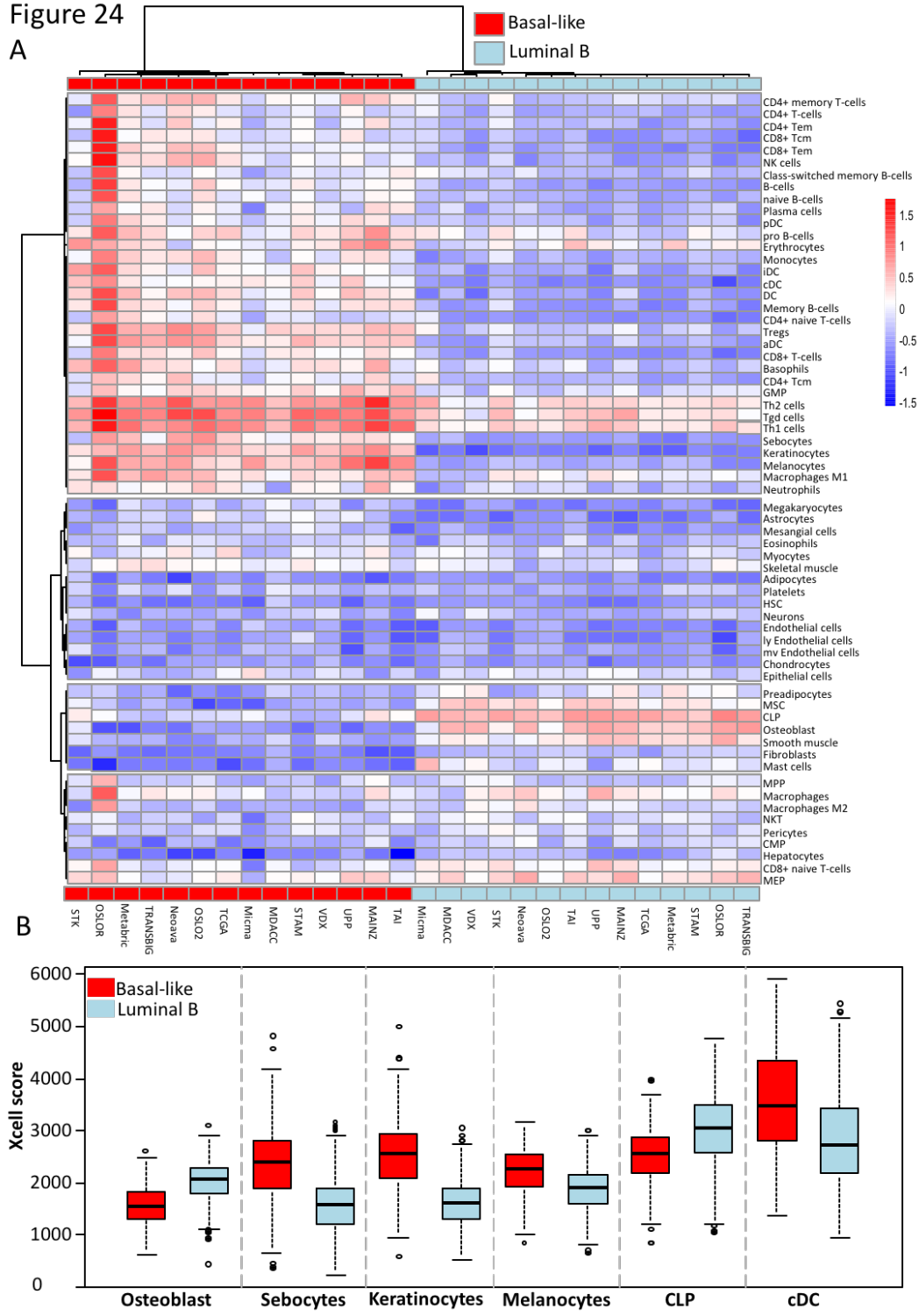

Figure 24. Comparison of xCell scores for each cell type according to Luminal B and Basal-like subtype. **A.** Unsupervised clustering of the median of each score in each cohort according to Luminal B and Basal-like subtypes. Annotation on top and bottom of heatmap shows Luminal B (light blue) and Basal-like (red) medians. **B**. Boxplots indicating the xCell scores according to Luminal B and Basal-like in the Metabric cohort. Cell types shown in the boxplots are always up or down-regulated in Luminal or Basal-like subtypes.

6.3.7 Luminal B vs Her2-enriched

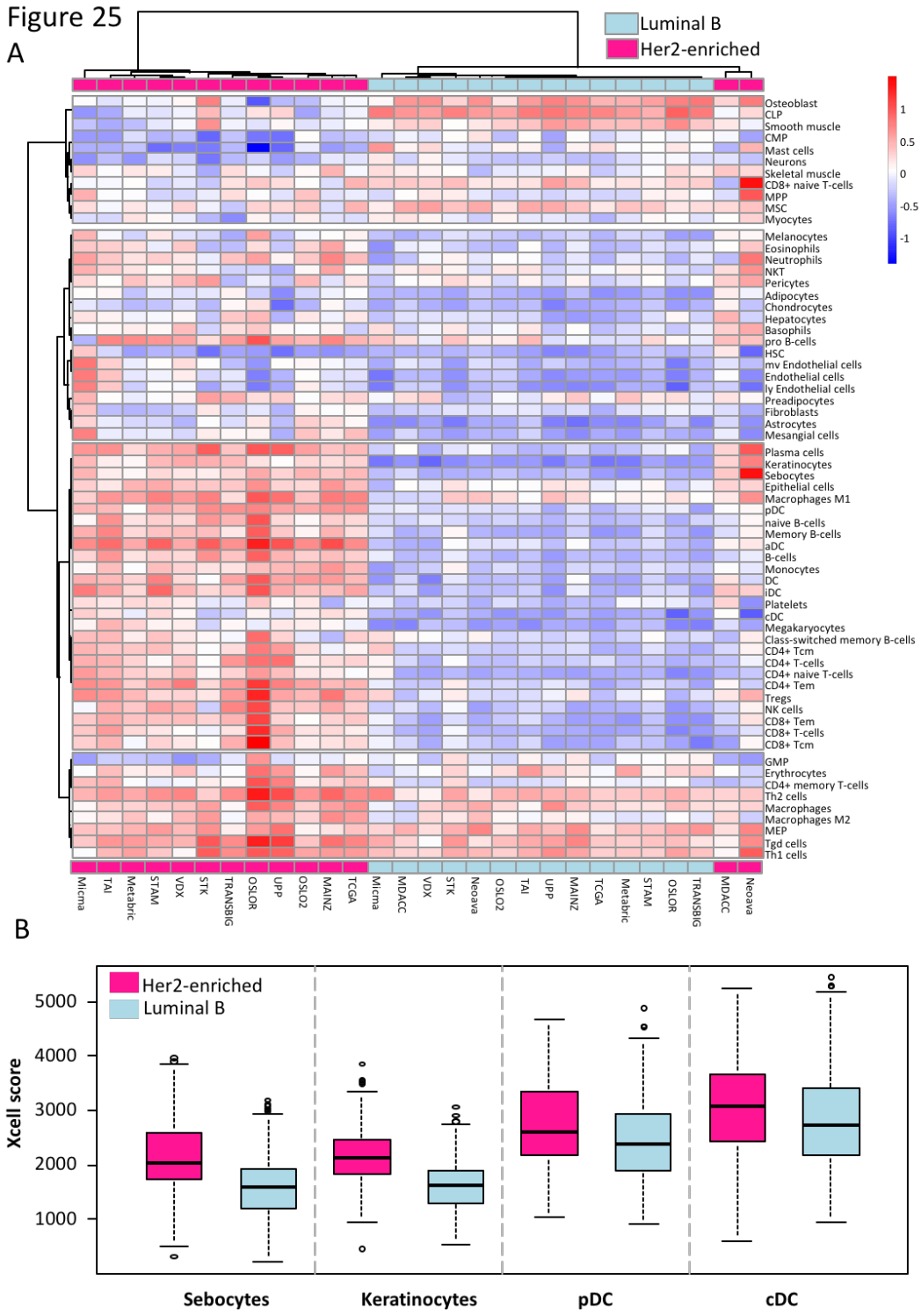

Figure 25. Comparison of xCell scores for each cell type according to Luminal B and Her2-enriched subtype. **A.** Unsupervised clustering of the median of each score in each cohort according to Luminal B and Her2-enriched subtypes. Annotation on top and bottom of the heatmap shows Luminal B (light blue) and Her2-enriched (pink) medians. **B**. Boxplots indicating the xCell scores according to Luminal B and Her2-enriched in the Metabric cohort. Cell types shown in the boxplots are always up or down-regulated in Luminal or Her2-enriched subtypes.

6.3.8 Her2-enriched vs Basal-like

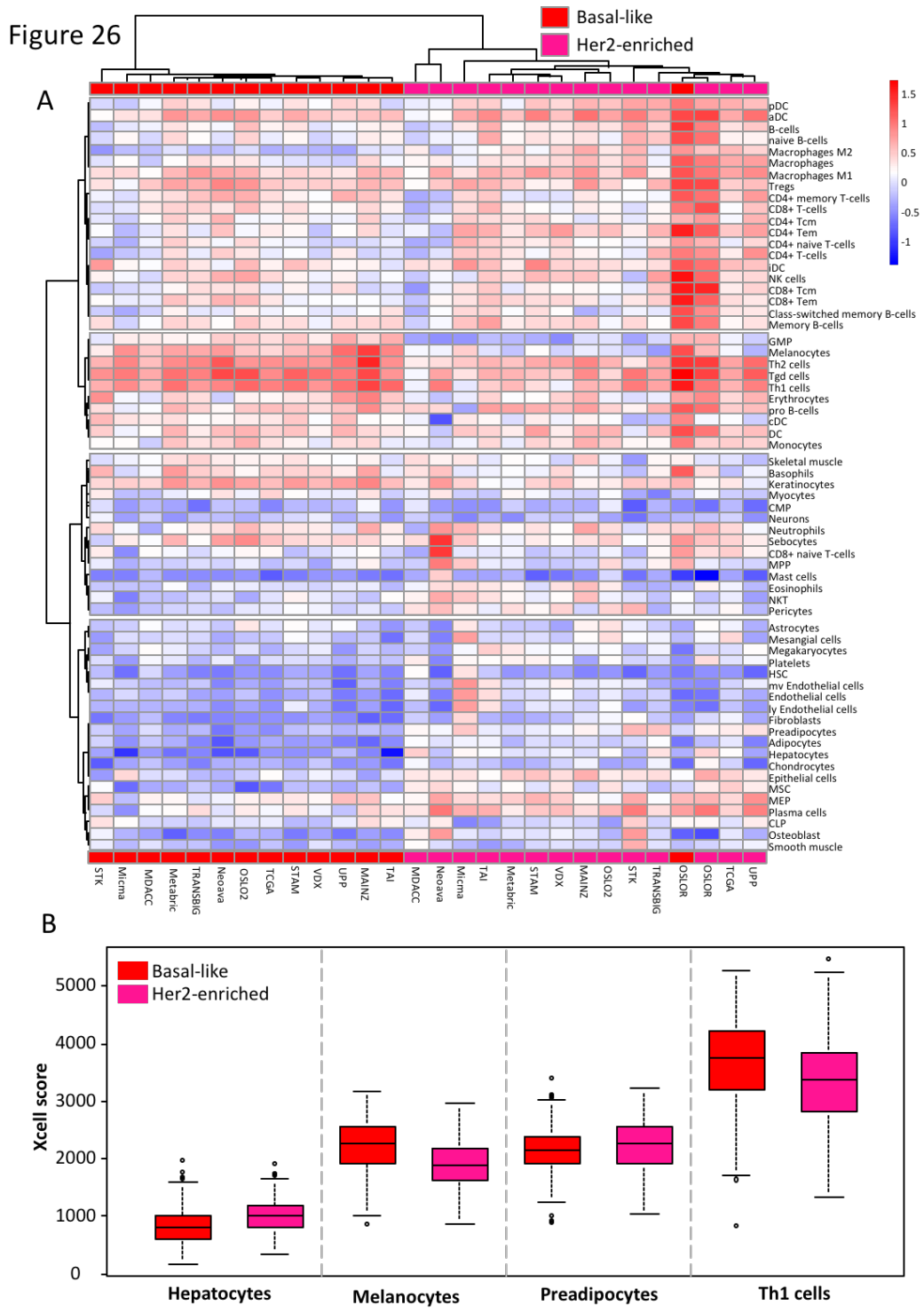

Figure 26. Comparison of xCell scores for each cell type according to Basal-like and Her2-enriched subtype. **A.** Unsupervised clustering of the median of each score in each cohort according to Basal-like and Her2-enriched subtypes. Annotation on top and bottom of the heatmap shows Basal-like (red) and Her2-enriched (pink) medians. **B**. Boxplots indicating the xCell scores according to Basal-like and Her2-enriched in the Metabric cohort. Cell types shown in the boxplots are always up or down-regulated in Basal-like or Her2-enriched subtypes.

6.3.9 Her2-enriched vs Normal-like

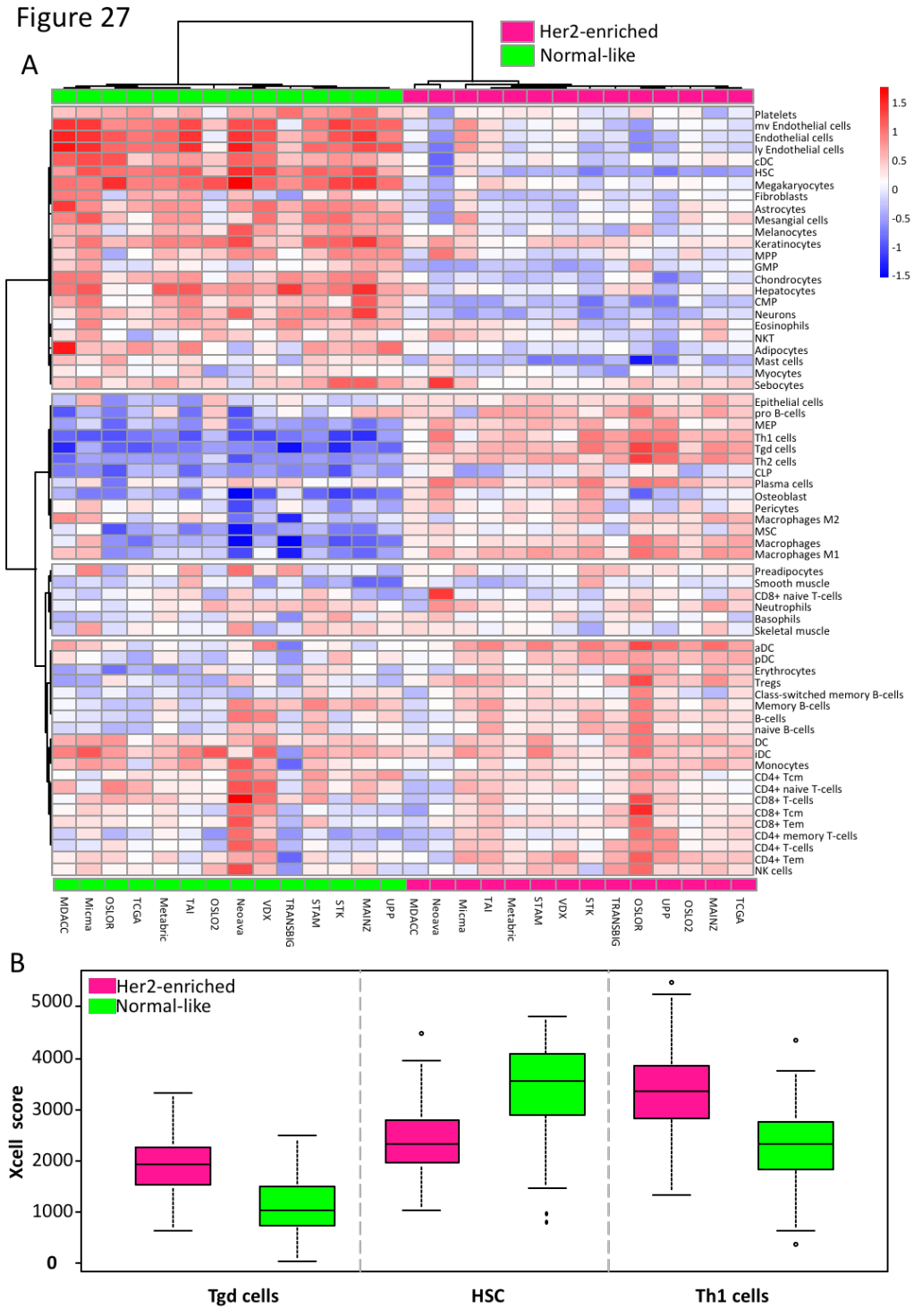

Figure 27. Comparison of xCell scores for each cell type according to Normal-like and Her2-enriched subtype. **A.** Unsupervised clustering of the median of each score in each cohort according to Normal-like and Her2-enriched subtypes. Annotation on top and bottom of the heatmap shows Normal-like (green) and Her2-enriched (pink) medians. **B**. Boxplots indicating the xCell scores according to Normal-like and Her2-enriched in the Metabric cohort. Cell types shown in the boxplots are always up or down-regulated in Normal-like or Her2-enriched subtypes.

6.3.10 Normal-like vs Basal-like

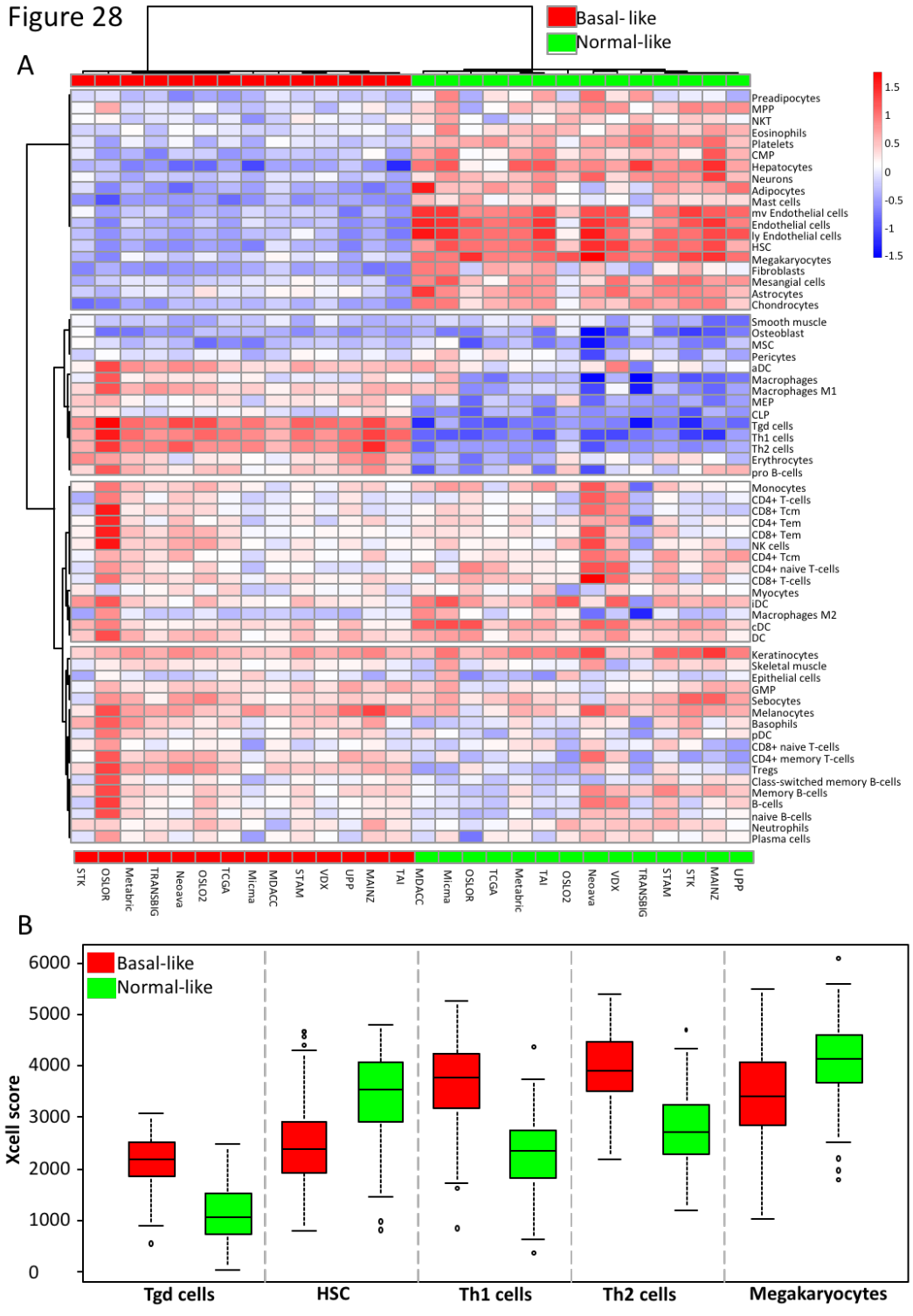

Figure 28. Comparison of xCell scores for each cell type according to Basal-like and Normal-like subtype. **A.** Unsupervised clustering of the median of each score in each cohort according to Basal-like and Normal-like subtypes. Annotation on top and bottom of heatmap the shows Basal-like (red) and Normal-like (green) medians. **B**. Boxplots indicating the xCell scores according to Basal-like and Normal-like in the Metabric cohort. Cell types shown in the boxplots are always up or down-regulated in Basal-like or Normal-like subtypes.

6.4 Cytokine serum levels in perspective of xCell scores

Finally, taking advantage of two cohorts for which we had (i) xCell scores and (ii) cytokines profile in the serum of the same patients. We investigated whether the circulating cytokines may correlate with infiltration of specific immune cells in the tumour. 98 MicMa and 25 DCTB/TIF samples were assessed for both serum cytokine levels, and xCell profiling. The 26 most common immune cell types (xCell scores) were correlated to the 27 serum cytokine levels profiled using Luminex, by using Pearson correlation heatmaps, seen in Figure 29 A and B. The heatmaps indicate that in both the MicMa and the DCTB/TIF cohorts, PDGF correlated with pro-B cells and NKT (natural killer T cells). We found that patients with the highest levels of pro-B cells (highest quartile) had significantly higher levels of PDGF in corresponding serum (Figure 29 C and D). Furthermore, samples with low NKT scores (lowest quartile) had higher levels of PDGF (Figure 29 E and F). In summary, high levels of PDGF in the serum associated with lower levels of NKT and higher levels of pro-B cells.

Figure 29

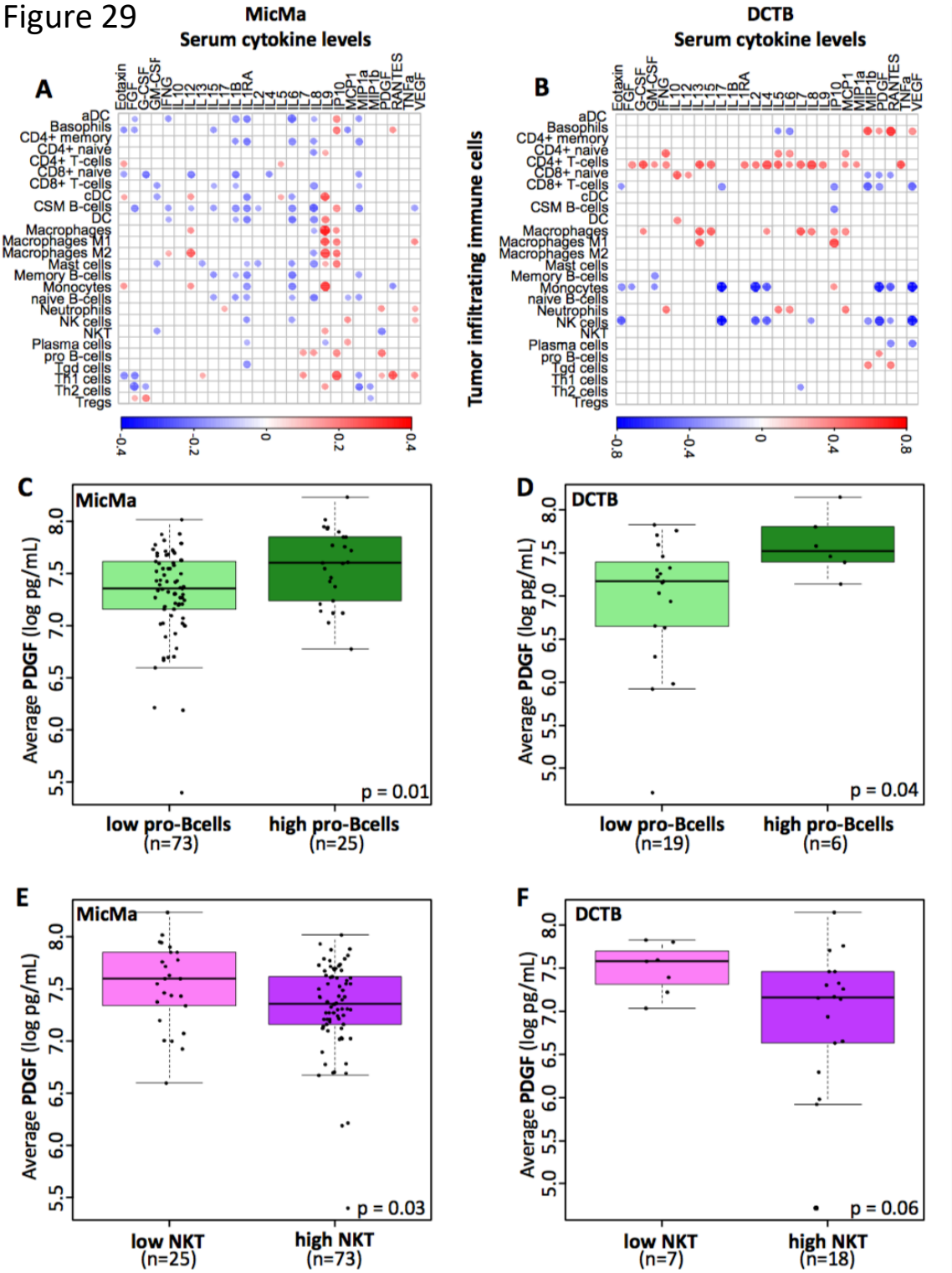

Figure 29. **Cytokine levels and infiltration immune cells.** **A** & **B** Correlation heatmaps depict only the correlations with p< 0,1 between cytokines levels and infiltration immune cells. Size of the dots reflect the strength of the Pearson correlation and colours the direction of the linear relationship. **C** & **D** Boxplots represent the average serum levels of PDGF (pg/mL) in respect to low or high (highest quantile) scores for pro-B cells infiltration at the tumour site in the MicMa (C) and DCTB (D) cohorts. **E** & **F** Boxplots of serum PDGF levels according to low (lowest quantile) and high scores for natural killer T cells (NKT) infiltration at the tumour site in the MicMa (C) and DCTB (D) cohorts. Manny-Whitney test p-values are denoted in the bottom right of each boxplot.

7 Discussion

7.1 Analysing the microenvironment

Understanding the interplay between the microenvironment and the cancer cells may give useful insight into different interactions that occur in the tumour. Higher immune infiltration is often associated with better survival rates in breast cancer. By looking at the different cell types present in different subtypes we may be able to understand more of what define each subtype.

7.1.1 xCell limitations and advantages

Immunohistochemistry and flow cytometry are traditional tools used to identify cells in the tumour microenvironment. However, these methods are somewhat limited, as they rely on a limited range of phenotypic markers and are unable to differentiate between cells with similar lineage. New deconvolution tools allowing to digitally dissect the microenvironment are not without fault and have several restrictions. xCell is a signature-based algorithm and relies on gene profiles from multiple sources, making the signature robust compared to single sample signatures. Cancer cells have highly dysregulated gene expression which can mimic that of other cells. Therefore, misinterpretation due to cancer cells abnormal gene expression could be a major issue when using deconvolution tools such as xCell.

xCell identifies 64 different cell types including immune cells from the adaptive and innate systems, hematopoietic progenitors, epithelial cells and extracellular matrix cell, all derived from thousands of expression profiles. Compared to other deconvolution algorithms, xCell signatures are superior to other methods. CIBERSORT, another deconvolution tool identifies 22 immune cell types.

Many of the cell types predicted by xCell should not be resident in the breast. For example, hepatocytes which are liver cells, and osteoblasts which are involved in bone synthesis. We therefore interpret the prediction of xCell as the fact that the cancer cells gene expression reflect the osteoblast gene signature. Therefore, as all cancer cells do not show the same degree of plasticity and expression of genes associated with non-breast resident cell types, the high degree of such cell types prediction may reflect more plastic and a stem cell like phenotype.

In addition, during the last few months, a new version of xCell that includes p values for the score has been released. This version may pinpoint the cell types not likely to be found in the environment of interest.

In our work, we validated the xCell method by immunohistochemistry staining with the CD4+, CD8+ and CD68 markers to average xCell scores of cell types carrying these markers.
We found the averaged xCell scores matching the staining´s scores given by pathological inspection. Nevertheless, this only validates a small fraction of the cell types that the xCell algorithm predicts. In addition the cohort used to validate the result is small. Alternative validation using more powerful and high throughput method such as single cell RNA-sequencing and mass spectrometer Cytof would provide better validation of the xCell algorithm. These alternative validations would be able to detect more of the components in the microenvironment. In addition a larger more detailed staining using more markers will also give a more detailed picture of the tumour environment.

Another limitation of the xCell method is that the scores we used are enrichment scores. These scores can be compared from samples (patients) to samples, but scores cannot be compared with each other. The CIBERSORT method has translated the enrichment scores into quantitative proportions, and the results can be compared between the cell types.
Additional insights could be provided in the future using CIBERSORT in the same setting as in the current study.

Despite faults xCell has addressed many problems that the deconvolution tools have faced previously. One of these faults are spillover between closely related cells, creating an algorithm to restrict spillover. xCell, unlike many other deconvolution algorithms, have gene signatures from several platforms, giving a more diverse signature. CIBERSORT for instants is limited to Affymetrix microarray based studies. Up until now, no other deconvolution tool has been able to identify as many cell types. We believe that this gives a better representation of the tumour microenvironment.

7.2 Immune infiltration according to ER status

We used xCell to compare cell scores according to ER status using unsupervised clustering. Interestingly, clustering indicates that the predicted immune infiltration could separate ER- to ER+ samples. This is mainly because ER- samples have a larger immune infiltration of CD4+ and CD8+ cell types. When comparing the mean of each cell type according to ER status using t-test, we found cell types significantly different in all cohorts. We found that gamma delta T cells (tgd cells), memory B-cells, activated dendritic cells (aDC), immature dendritic cells (iDC), dendritic cells (DC) and Th2 cells, had significantly higher means in ER- samples (Figure 18B). Helper T cells and memory B cells have previously been associated with better overall survival (72). Higher levels Tgd has been reported to be associated with longer disease free survival and overall better survival (95). Dendritic cells have little impact on survival rates (Figure 11) (72). A detailed profile of different dendritic cells and survival rates have not been assessed, and the effects of different dendritic cells in the environment had not been extensively studied. However, the high levels and interplay between dendritic cells and activated T cells would indicate an higher immune response in ER negative breast cancer, which has a worse prognosis.

7.3 Immune infiltration according to PAM50 subtypes

7.3.1 Comparison between Luminal A and Luminal B

The comparison between xCell scores and Luminal subtypes using unsupervised clustering in heatmaps (Figure 19A) showed that these two subtypes which are close in lineage could be distinguished using only the xCell scores. Clustering also showed higher levels of helper T cells and macrophages in the Luminal B subgroup. While in Luminal A, there were higher levels of endothelial cells and fibroblasts. When comparing the mean of each cell type according to Luminal subtype using t-test, we found cell types significantly different in all cohorts (Figure 19B). In the Luminal A subtype there were higher levels of Keratinocytes, Chondrocytes, Astrocytes, Endothelial cells, Adipocytes, HSC, and lymphoid endothelial cells compared to the Luminal B subtype. The presence of HSC, endothelial cells and other non-breast resident cell types, might indicate that Luminal A may have more plastic phenotype than Luminal B. Luminal B showed higher levels of Tgd cells, Th1 cells, Th2 cells, myeloid progenitor (MEP), and lymphoid progenitor (CLP); therefore, suggesting that Luminal B are

more immunogenic. Even though MEP and CLP are most commonly found in the bone marrow or the thymus, these results may indicate that a certain level of hemopoietic differentiation may be ongoing. Luminal B have a worse survival rate than Luminal A, even though Luminal B tumours are more immunogenic. This could be explained by higher levels of Tregs and M2 macrophages, which has previously been associated with worse prognosis (71) .

7.3.2 Comparison between Luminal A and Her2-enriched

The comparison between xCell scores in Luminal A and Her2-enriched subtypes using unsupervised clustering separates the two subtypes (Figure 20A). A higher level of cells associated with connective tissue was found in Luminal A, while there were higher immune cells in Her2-enriched, as previously reported (72). Therefore, our results may indicate that a main difference between Luminal A and Her2-enriched could be that Her2-enriched mainly have an immune response, while the response in Luminal A may be more stromal.

7.3.3 Comparison between Luminal A and Basal-like

The Luminal A and the Basal-like subtypes also clustered separately. Similarly to the Her2-enriched comparison with Luminal A, the separation was due to higher immune cell levels in Basal-like, while higher connective tissue was present in the Luminal A. Cells that had higher levels in the Luminal A subgroup were Mast Cells, Fibroblasts, Chondrocytes and HSC. Mast cells are tissue resident cells and are involved in inflammatory response during parasitic invasion. However, studies show that mast cells in some types of breast cancers are involved in promoting angiogenesis and tumour inflammation (96). As expected, higher levels of many immune cells, such as Tgd cells Th1 cells, Th2 cells, pro B cells, macrophages M1, and CD4+ memory T-cells were found in the Basal-like subgroup. Interestingly, in the Basal-like subgroup, we also found higher levels of Keratinocytes and melanocytes, which are basal cell types.

7.3.4 Comparison between Luminal A and Normal-like

Even though the unsupervised clustering comparing Luminal A to Normal-like separates the two subgroups, we found few differences in cell type composition. There is a slight elevation of immune cells in the Normal-like subtype. Dendritic cells present antigen to the T cells, which initiates the adaptive immune response. There are slightly higher levels of T cells in the Normal-like subtype, but there are low levels of helper T cells. These findings indicate that the

Normal-like subgroup has some sort of immune response, but very low, and maybe not strong enough to activate the adaptive immune response efficiently. Keratinocytes and Megakaryocytes are stem cells that have higher means in the Normal-like subtypes. Several studies suggests that epithelial cells in the tumour may detach and undergo epithelial-mesenchymal transition (EMT) and will then resemble stem cells (97). This may explain the stem cells in the microenvironment of the Normal-like subtype.

7.3.5 Comparison between Luminal B and Normal-like

We found more differences in cell type comparison when comparing Luminal B to Normal-like. One of the main differences seemed to be that Normal-like may be enriched in effector lymphocytes while Luminal B in Helper lymphocytes. There were also higher means for stem cells and stroma cells in the Normal-like subtype. Our result therefore, indicate an adaptive immune response in the Luminal B subtype.

7.3.6 Comparison between Luminal B and Basal-like

While comparing Luminal B to Basal-like, we observed similar behaviour as comparing the Luminal A to the Basal-like. First the higher levels of Melanocytes, Sebocytes, and Keratinocytes in the Basal-like phenotype possibly picked up the difference between the two linages the cancer originates from. In addition, a higher immune cell infiltration was found in the Basal-like, which also includes higher regulatory T cells (tregs) scores. Regulatory T cells supress the immune reaction and are correlated with worse survival (95). These results may indicate that while an immune response is well initiated in Basal-like, exhaustive amounts of immune editing associated with immune evasion as this subtype has a worse survival than Luminal B.

7.3.7 Comparison between Luminal B and Her2-eriched

The comparison between Luminal B and Her2-enriched showed surprisingly little difference between these two subgroups. The Her2-enriched subtype is thought to be more immunogenic than luminal subgroups. However, it was the first unsupervised clustering we performed which did not clearly separate the two subtypes. The presence of dendritic cells may be higher in Her2-enriched, possibly indicating higher presence of neoantigens in this subtype.

7.3.8 Comparison between Her2-enriched and Basal-like

Basal-like and Her2-enriched are both thought to be immunogenic. We definitely found little difference in the levels of immune cell types between these two subtypes. However, there were higher levels of Macrophages and M1 macrophages in the Her2-enriched subtype, while higher levels of Melanocytes, Th2 cells, Th1 cells and Tgd cells were found in the Basal-like subtype. Nevertheless, according to our analysis it was hard to find some clear cut difference between these two immunogenic subtypes. Even the cell types of the basal cell layer were not clearly enriched in the Basal-like, suggesting that Her2-enriched may also deride from basal cells.

7.3.9 Comparison between Normal-like and Her2-enriched, Basal-like

When comparing the Normal-like to Her2-enriched or Basal-like we made similar observations. There were higher levels of immune cells in Her2-enriched or Basal-like compared to Normal-like. While higher levels of non-breast residents and connective tissue cell types were found in the Normal-like which may indicate a certain degree of differentiation, plasticity and stem cell like phenotype of the Normal-like tumours compared to the two other subtypes.

7.4 Cytokine levels in correlation to Immune cell levels

Cytokines are used to transmit signals between immune cells, and have an effect on both immune and stromal cells. We investigated whether the circulating cytokines may correlate with infiltration of specific immune cells in the tumour. The 26 most common immune cell types (xCell scores) were correlated to the 27 serum cytokines (Figure 29A and B). We found positive correlation between the PDGF serum level and pro-B cell infiltration. Both B cells and PDGF are involved in lymphangiogenesis, the formation of lymph vessels (98). Therefore, our results suggesting that PDGF may act as a pro-tumorigenic cytokine. Studies indicate that inhibition of PDGF may increase chemotherapy efficiency (99). We also reported higher levels of Natural Killer T cells (NKT) that are anti-correlated with serum levels of PDGF, which may provide further clues to associate PDGF with tumour progression, as NKT cells can directly kill malignant cells (100).

8 Conclusion and future perspectives

When looking at the heterogeneity of intertumoral immune infiltrations according to breast cancer molecular subtype, we found clear differences between the subtypes. Almost all the unsupervised clustering we performed based on the median of the scores in each cohort (14 cohorts) separated the subtypes. These results showed a strong and consistent association between the breast cancer subtypes and immune/stroma cell infiltration. It is at this stage impossible to conclude whether a specific subtype "attracts" a specific microenvironment, or if an initial microenvironment shapes the subtype. It is however, not to be forgotten that both the PAM50 subtyping and xCell scores derive from gene expression data, where in the differences we picked up using xCell were intrinsically already present in the PAM50, but not clearly pinpointed.

The Luminal A subtype seems to be the subtype with the least immune infiltration. It is however the subtype with the best outline and overall survival. Thus it could also be discussed whether the presence of immune cells may shape a more heterogenous and immunodetected tumour, which may then escape or be more resistant to therapies.

In that sense, the Basal-like subtypes which shows the highest levels of overall immune infiltration also contains higher levels of Tregs, which will participate to the immunoediting process and to avoid immune destruction. Recent studies have suggested that the presence of CD8 T cells may favour the epithelial-mesenchymal transition (EMT) creating a more aggressive and differentiated tumour, which will have an increased chance to relapse (97). To be able to identify stem cell like phenotypes, associated with specific immune cells, additional experiment remain to be performed. Single sell RNA sequencing may give a better idea of the function and characteristics of the specific cell.

Using xCell to profile cell scores of 64 different cell types gave an insight to not only the immune cells, but the microenvironment complexity of the different subtypes. It would be interesting to perform the statistical analysis with the new version of xCell, as we recently checked that the xCell scores obtained were the same as the newest version. Analysis done in the new version would not only give us cell types, but also indicate whether the results were

trustworthy by using the p-value the authors implemented. This would remove potential misinterpretations of cell types.
In addition, we thought of performing a similar type of analysis using CIBERSORT. Even though the cell types obtained in CIBERSORT would be different a certain level of comparison would be possible, allowing to interpret the results with more confidence.

Finally, being able to predict or understand the tumour microenvironment, using a surrogate and readily available tissue like blood, would be extremely useful.
When investigating whether cytokine serum levels are related to the levels of tumour infiltrating immune cells, we found positive correlation between B cells scores and PDGF-bb serum cytokine levels. These results were validated in two small cohorts. Both B cells and PDGF-bb are involved in lymphangiogenesis the formation of lymph vessels. Previous studies indicate that blocking the level of PDGF increases the efficiency of chemotherapy. PDGF can prove to be a good target for immunotherapy.

Our analysis indicates clear differences in tumour microenvironment associated with breast cancer molecular subtypes. In the future, it would be interesting to assess how the microenvironment evolves during therapy and especially immunotherapy. While lots remain to be understood whether the microenvironment shapes the molecular profiles of the tumour or vice versa. It is apparent that both are interconnected in breast cancer, while immune infiltration has been traditionally associated with a better prognosis, we observe that the subtype with better outcome often associate with low immune infiltration. Such an observation suggests that high immune infiltration may also shape and drive more aggressive tumours.

Bibliography

1. Weinberg RA. The Biology of Cancer. In The nature of cancer.: Garland Science; 2014. p. 31-69.
2. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000 ;100(1):57-70.
3. Hanahan D, Weinberg R. Hallmarks of cancer: the next generation. Cell. 2011, 144(5):646-74.
4. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. Nature. 1998 ;396(6712):643-9.
5. Weinberg R. The Biology of Bancer. In Cancer- causing mutations affect the germline and stroma.: Garland Science; 2014. p. 10-13.
6. Thompson SL, Compton DA. Chromosomes and cancer cells. Chromosome Res. 2011;19(3):433-44.
7. Parham P. The Immune System. In Cancer and its interactions with the immune system.: Garland Science; 2015. p. 509- 513.
8. Weinberg R. The Biology of Cancer. In Cellular Oncogenes.: Garland Science; 2014. p. 103- 128.
9. Weinberg R. The Biology of Cancer. In Tumor Suppressor Genes.: Garland Science; 2014. p. 231-254.
10. Ramsay DT, Kent JC, Hartmann RA, Hartman PE. Anatomy of the lactating human breast redefined with ultrasound imaging. Journal of Anatomy. 2005; 206(6): 525–534.
11. Memorial , Solan , Kettering , cancer , center. [Online]. [cited 2018 03 20. Available from: https://www.mskcc.org/cancer-care/types/breast/anatomy-breast.
12. Mark D, Sternlicht , Hosein KM, Pengfei L, Werb Z. Hormonal and local control of mammary branching morphogenesis. Differentiation; research in biological diversity. 2006;74(7):365-381.
13. World Health Organisation. Globocan. [Online].; 2012 [cited 1018 03 02. Available from: http://globocan.iarc.fr/old/FactSheets/cancers/breast-new.asp.
14. Tavassoli FA, Devilee P. World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of the Breast and Female Genital Organs. Lyon: IARC Press; 2003.
15. The Norwegian Cancer Registry. Fakta om Brystkreft - Brystkreft. [Online].; 2016 [cited 2018 03 03. Available from: https://www.kreftregisteret.no/Generelt/Fakta-om-kreft/Brystkreft-Alt2/.
16. Hofvind S, Ursin G, Tretli S:SS, Møller B. Breast cancer mortality in participants of the Norwegian Breast Cancer Screening Program. Cancer. 2013;119(17):3106-12.
17. Weinberg R. The Biology of Cancer. In The Nature of cancer.: Garland Science; 2014. p. 40-53.
18. Bombonati A, Sgroi DC. The Molecular Pathology of Breast Cancer Progression. Journal of Pathology. 2011; 223(2): 307–317.
19. RnCeus. [Online]. Available from: http://www.rnceus.com/dcis/sub.html.

20. Stickeler E. Prognostic and Predictive Markers for Treatment Decisions in Early Breast Cancer. Breast Care. 2011; 6(3): 193–198.
21. Rakha EA, Ellis IO. Modern classification of breast cancer: should we stick with morphology or convert to molecular profile characteristics. Adv Anat Pathol. 2011;18(4):255-67.
22. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology. 1991;19(5):403-10.
23. Singletary SE, Greene FL, Force BT. Revision of breast cancer staging: the 6th edition of the TNM Classification. Semin Surg Oncol. 2003;21(1):53-9.
24. Harbeck N, Thomssen C, Gnant M. St. Gallen 2013: Brief Preliminary Summary of the Consensus Discussion. Breast Care. 2013;8(2):102-109.
25. Silberman A, Monica B. ER-Positive Breast Cancer: Prognosis, Life Expectancy, and More. 2016.
26. Silberstein GB, Van Horn K, Shyamala G, Daniel CW. Progesterone receptors in the mouse mammary duct: distribution and developmental regulation. Cell Growth Differ. 1996;7(7):945-52.
27. Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. J Clin Oncol. 2013;31(31):3997-4013.
28. Beresford MJ, Wilson GD, Makris A. Measuring proliferation in breast cancer: practicalities and applications. Breast Cancer Research. 2006;8(6):216..
29. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747-52.
30. Parham P. The Immune system. In Elements of the immune system and their Roles in Defense.: Garland Science ; 2015. p. 1-26.
31. Dranoff G. Cytokines in cancer pathogenesis and cancer therapy. Nat. Rev. 2004; 4, 11–22.
32. Murphy K, Weaver C. Janeway´s Immunobiology. In Principles of innate immunity.: Garland Science; 2017. p. 3-7.
33. Kenneth M, Casey W. Janeway´s Immuno Biology. In Principles of innate immunity.; 2017. p. 7-11.
34. Parham P. The immune system. In Innate immunity: The lnduces Response.: Garland Science ; 2015. p. 47- 53.
35. Murphy K, Weaver C. Janeways´s Immuno Biology. In Pattern recogniction By cells of the innate immune system.: Garland Science ; 2017. p. 77-107.
36. Parham P. The Immune System. In Toll like receptors sense the presence of the foure main groupes of pathogenic microorganisms.: Garland Science; 2015. p. 66- 71.
37. Parham P. The Immune System. In NOD-like receptors recognize bacteria degradation products in the cytoplasma.: Garland Science; 2015. p. 54-55.
38. Murphy K, Weaver C. Janeway´s Immuno Biology. In Induced innate responce to infection.: Garland Science ; 2017. p. 107-131.
39. Murphy K, Weaver C. Janeways´s Immuno Biology. In Macrophages.; 2017. p. 7.
40. Murphy K, Weaver C. Jeneway´s Immuno Biology. In Granulocytes.; 2017. p. 8.

41. Murphy K, Weaver C. Janeway´s Immuno Biology. In Mast cells.; 2017. p. 8.
42. Murphy K, Weaver C. Janeway´s immuno Biology. In Dendritic cells.; 2017. p. 8.
43. Murphy K, Weaver C. Janeway´s Immunobiology. In Innate Lymphosytes and natural killers cells are effector cells that share similarities with lymphoid lineages of the adaptive immune system.: Garland Science; 2017. p. 11.
44. Robbins , Cotran. Pathologic Basis of Disease. In.: 8th ed. Philadelphia, Penn: Saunders Elsevier; 2009.
45. Murphy K, Weaver C. Jaenway´s Immuno Biology. In Pricibples of adaptive immunity.: Garland Science; 2017. p. 11-25.
46. Murphy K, Weaver C. Janeway´s Immuno Biology. In T-cell-mediated Immunity.: Garland; 2017. p. 345-346.
47. Murphy K, Weaver C. Janeway´s Immuno Biology. In Antigen presentaion to T Lymphocytes.: Garland Science; 2017. p. 213-225.
48. Murphy K, Weaver C. Janeway´s Immuno Biology. In Priming of Naive T cells by pathogen-activated dendritic cells.: Garland Science ; 2017. p. 372-375.
49. Murphy K, Weaver C. Janeway´s Immuno biology. In Priming of naive T-cells by pathogen-activated dendritic cells..: Garland Science ; 2017. p. 379-380.
50. Murphy KWC. Janeway´s Immuno Biology. In Immunological memory.: Garland Science; 2017. p. 480-483.
51. Murphy K, Weaver C. Janeway´s Immunobiology. In Antigen recognition by Bcell and T-cells receptors.: Garland; 2017. p. 139-145.
52. Murphy K, Weaver C. Janeway´s Immuno Biology. In B-cell activation bt antigen and helper T cells..: Garland Science; 2017. p. 400-406.
53. Murphy K, Weaver C. Janeway´s Immunobiology. In B-cell activation by antigen and helper T cells.: Garland; 2017. p. 406-412.
54. Murphy K, Weaver C. Janeway´s Immunobiology. In B-cell activation by antigen and helper T cells.: Garland Science; 2017. p. 406-413.
55. Murphy K, Weaver C. Janeway´s Immunobiology. In Immunological memory.: Garland Science; 2017. p. 475- 476.
56. Parrish-Novak p, Dillon SR, Nelson A, Hammond A, Sprecher C, Gross JA, et al. Interleukin 21 and its receptor are involved in NK cell expansion and regulation of lymphocyte function. Nature. 2000 Nov 2;408(6808):57-63.
57. Murphy K, Weaver C. Janeway´s Immunobiology. In B-cell activation by antigen and helper T cells.: Garland Science ; 2017. p. 418-421.
58. Murphy K, Weaver C. Janeway´s Immunobiology. In Principles of innate immunity.: Garland ; 2017. p. 9-11.
59. Murphy K, Weaver C. Janeway´s Immunobiology. In Induced unnate responses to infection.: Garland Science; 2017. p. 112.
60. Murphy K, Weaver C. Janeway´s Immunobiology. In General Properties of effector T cells and their cytokines.: Garland Science ; 2017. p. 383-386.
61. Weinberg RA. The biology of cancer. In Cancer susceptibility and immune function.: Garland Science; 2014. p. 745-751.
62. Parham P. The immune system. In Cancer and its interactions with the immune system.: Garland Science; 2015. p. 509-510.

63. Parham P. The immune system. In Cancer and its interactions with the immune system.: Garland; 2015. p. 516-517.
64. Paraham P. The Immune System. In Cancer and its interactions with the immune system.: Garland Science ; 2015. p. 518-519.
65. Weinberg RA. The biology of cancer. In Tumor cell conunterattacks on immune cells.: Garland Science ; 2014. p. 765- 769.
66. Weinberg RA. The biology of cancer. In Immunoevasion strategies.: Garland Science; 2014. p. 761-765.
67. Weinberg RA. The biology og Cancer. In Cancer cells can evade immune detection by supperesing cell surface display of tumour antigenxs.: Garland sience p. 761-765.
68. Weinberg RA. The biology of cancer. In Mobilization of patients immune responses.: Garland; 2014. p. 786- 791.
69. Institute NC. NCI Dictionary of Cancer Terms. [Online]. [cited 2018 03 12. Available from: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/tumor-microenvironment.
70. Denkert C, Minckwitz G, Brase JC. Tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy with or without carboplatin in human epidermal growth factor receptor 2-positive and triple-negative pri- mary breast cancers. Jurnal of Clinical Oncology. 2015;33(9):983–991.
71. Loi S, Sirtaine N, Piette F, Salgado R, Viale G, VanEenoo F, et al. Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy. Journal of Clinical Oncology. 2013; (7):860-867.
72. Ali HR, Chlon L, Pharoah PDP, Markowetz F, Caldas C. Patterns of Immune Infiltration in Breast Cancer and Their Clinical Implications: A Gene-Expression-Based Retrospective Study. PLoS Med.. 2016;13(12):e1002194.
73. Denkert C, von Minckwitz G, Brase JC. Tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy with or without carboplatin in human epidermal growth factor receptor 2-positive and triple-negative pri- mary breast cancers. J Clin Onco. 2015;33(9):983–991.
74. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nature Methods. 2015;12(5):453-7.
75. Clancy T, Dannenfelser R, Troyanskaya O, Malmberg KJ, Hovig E, Kristensen V. Bioinformatics Approaches to Profile the Tumor Microenvironment for Immunotherapeutic Discovery. In.: Curr Pharm ; 2017;23(32):4716-4725.
76. Murphy K, Weaver C. Janeway´s immunobiology. In Immunohistochemistry.: Garland Science ; 2017. p. 762.
77. Cell Signaling Technology. [Online]. Available from: https://www.cellsignal.com/contents/resources-applications/fluorescent-multiplex-immunohistochemistry/fluoresence-mihc.
78. Yoshihara K, Shahmoradgoli M, Martinez E. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013; 4: 2612.
79. Ju W, Greene CS, Eichinger F. Defining cell-type specificity at the transcriptional level in human disease. Genome Res. 2013; 23: 1862-73.

80. Schmidt M, Böhm D, von Törne C, Steiner E, Puhl A, Pilch H, et al. The humoral immune system has a key prognostic impact in node-negative breast cancer. Cancer Res. 2008;68(13):5405-13.
81. Hatzis C, Pusztai L, Valero V. A Genomic Predictor of Response and Survival Following Taxane-Anthracycline Chemotherapy for Invasive Breast Cancer. JAMA. 2011;305(18):1873-1881.
82. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012;486(7403):346-52.
83. Jabeen S, Zucknick M, Nome M, Dannenfelser R, Fleischer T, Kumar S, et al. Serum cytokine levels variations in breast cancer patients during bevacizumab neoadjuvant chemotherapy. 2017.
84. Aure MR, Jernström S, Krohn M, H.K. V, E.U. D, Rødland E,KR, et al. Integrated analysis reveals microRNA networks coordinately expressed with key proteins in breast cancer. Genome Med. 2015;7(1):21.
85. Loi S, Haibe-Kains B, Desmedt C, Wirapati P, Lallemand F, Tutt AM, et al. Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen. BMC Genomics. 2008; 22;9:239.
86. Pawitan Y, Bjöhle J, Amler L, Borg AL, Egyhazi S, Hall P, et al. Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. Breast Cancer Res. 2005;7(6):R953-64.
87. Kao KJ, Chang KM, Hsu HC, Huang AT. Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization. BMC Cancer.. 2011; 11: 143.
88. Desmedt C, Piette F, Loi S, Wang Y, Lallemand F, Haibe-Kains B, et al. Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. Clin Cancer Res. 2007;13(11):3207-14.
89. Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A, et al. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. PNAS. 2005; 102 (38) 13550-13555.
90. Minn AJ, Gupta GP, Padua D, P. B, Nguyen DX, Nuyten D, et al. Lung metastasis genes couple breast tumor size and metastatic spread. Proc Natl Acad Sci U S A. 2007;17;104(16):6740-5.
91. Gene Factor. [Online]. [cited 2018 04 20. Available from: https://thegenefactor.wordpress.com/genetic-testing-2/.
92. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest. 2011;121(7):2750-67.
93. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biology. 2017;18:220.
94. Espinoza JA, Jabeen S, Batra R, Papaleo E, Haakensen V, Timmermans , et al. Cytokine profiling of tumor interstitial fluid of the breast and its relationship with lymphocyte infiltration and clinicopathological characteristics. 2016 Oct 24;5(12):e1248015.

95. Bense RD, Sotiriou C, Piccart-Gebhart MJ, Haanen JB, van Vugt MA, de Vries EG, et al. Relevance of tumour-infiltrating immune cell composition and functionality for disease outcome in breast Cancer. J Natl Cancer Inst. 2016; 13;109(1).
96. Cimpean AM, Tamma R, Ruggieri S, Nico B, Toma A, Ribatti D. Mast cells in breast cancer angiogenesis. Crit Rev Oncol Hematol. 2017 Jul;115:23-26.
97. Cimpean AM, TR, Ruggieri S, Nico B, Toma A, Ribatti D. Immune promotion of epithelial-mesenchymal transition and generation of breast cancer stem cells. Cancer Res.. 2010 April 15; 70(8): 3005–3008.
98. L.K. D, Karempudi P, Luther SA, Ludewig B, Harris NL, I. Interactions between fibroblastic reticular cells and B cells promote mesenteric lymph node lymphangiogenesis. Nat Commun. 2017;8:367.
99. Pietras K ea. Inhibition of PDGF receptor signaling in tumor stroma enhances antitumor effect of chemotherapy. Cancer Res. 2002;62:5476-5484.
100. Terabe MBJA. NKT cells can directly kill malignant. Cancer Res. 2008 ; 101: 277–348.
101. Alberts B, Johnson AD, Lewis J, Morgan D, M. R, Roberts K, et al. Molecular biology of the cell. In.; 6.ed.

Appendix A

Table 8. The 64 cell types xCell calculates profiles for. The cell type names and abbreviations xCell uses, Cell type and a short description of cell function is described *(101)*.

Cell	Abbreviation	Cell type	Short description of function:
CD4 positive memory T-cell	CD4+ memory T-cell	Lymphoid	After an primary response a T cell is stored and used to mount a secondary response at a later time.
CD4 positive Naïve T-cells	CD4+ Naïve T-cells	Lymphoid	T helper cell that has not been activated by antigen.
CD4 positive T-cells	CD4+ T-cells	Lymphoid	T helper cell, used to activate B cells and mount the adaptive immune response.
Central memory T cell	Tcm cells	Lymphoid	Memory lymphocytes that circulate between on blood and tissue. Ned to be reactivated in the secondary lymphoid tissues in order to become effector T cells.
Effector memory T cell	Tem cells	Lymphoid	Memory lymphocytes that recirculate between blood and peripheral tissues. Maturate rapidly into effector T cells upon secondary stimuli.
CD4 positive Centeral memory T cell	CD4+ Tcm	Lymphoid	Central memory T cells that when stimulated differentiate into CD4+ T cells.
CD4 positive Effector memory T cell	CD4+Tem	Lymphoid	Effector memory T cells, when stimulated differentiate into CD4+ T cells.
CD8 positive T-cells	CD8+ T-cells	Lymphoid	Cytotoxic T cell, kill virus infected cells.
CD8 positive naïve T-cells	CD8+ naïve T-cells	Lymphoid	Cytotoxic T cell that has not been activated.
CD8 positive Centeral memory T cell	CD8+ Tcm	Lymphoid	Central memory T cells that when stimulated differentiate into CD8+ T cells
CD8 positive Effector memory T cell	CD8+ Tem	Lymphoid	Effector memory T cells, when stimulated differentiate into CD4+ T cells.
Regulatory T cells	Tregs	Lymphoid	Regulates the immune response, by supressing the functions of other immune cells
Th1 cells	Th1 cells	Lymphoid	A subset of helper T cells. Involved in macrophage activation and stimulating B cells to produce antibodies.

Th2 cells	Th2 cells	Lymphoid	A subset of helper T cell. Involved in stimulating B cells to produce antibodies.
Tgd cells	Tgd cells	Lymphoid	T cells that have a TCR made up of gamma delta chains, rather than the more common alfa beta chains.
Natural killer cells	Nk cells	Lymphoid	Natural killer cells, kills virus infected cells, dose not have antigen specific receptors like T and B cells.
Natural killer T cells	NKT	Lymphoid	Kills infected cells, using specific antigen receptors.
B-cells	B-cells	Lymphoid	Antigen specific lymphocyte.
Naïve B-cells	Naïve B-cells	Lymphoid	Inactivated B cell.
Memory B-cells	Memory B-cells	Lymphoid	After a primary response a pro B cell is kept in the bone marrow, until reactivation by secondary response.
Class-switched memory B-cells	Class-switched memory B-cells	Lymphoid	B cell with a specific immunoglobulin, reactivated upon secondary response.
Pro B-cells	Pro B-cells	Lymphoid	During a stage in B cell development in then B-cell have surface markers proteins but have not yet completed heavy chain gene rearmament.
Plasma cells	Plasma cells	Lymphoid	Antigen secreting B cell.
Monocytes	Monocytes	Myeloid	Precursor of tissue macrophages.
Macrophages	Macrophages	Myeloid	Phagocytotic cells present in most tissues, that phagocytise pathogens.
Macrophages M1	Macrophages M1	Myeloid	Macrophages activated by the phagocytosis of pathogens, inflammatory response.
Macrophages M2	Macrophages M2	Myeloid	Macrophages that are activated by a parasite invasion, stimulate tissue repair.
Dendritic cells	DC	Myeloid	A cell that takes up pathogens and presents antigen to T cells.
Conventional dendritic cells	cDC	Myeloid	Dendritic cell that takes up antigen in the peripheral tissue, and are activated by antigens.
Plasmacytoid dendritic cells	pDC	Myeloid	Dendritic cells that's main function is to produce antiviral interferons.

Immature dendritic cells	iDC	Myeloid	Immature Dendritic cells are not fully developed.
Neutrophils	Neutrophils	Myeloid	Phagocytises and kills pathogens.
Eosinophils	Eosinophils	Myeloid	A large granule riche cell, protects agents parasitic invasion.
Mast cells	Mast cells	Myeloid	A large granule riche cell, protects agents parasitic invasion.
Basophils	Basophils	Myeloid	A large granule riche cell, protects agents parasitic invasion.
Hematopoietic stem cells	HSC	Stem cells	Stem cells that gives rise to blood cells, found in the bone marrow and in the peripheral blood.
Common Lymphoid progenitor	CLP	Stem cells	Stem cells that give rise to different lymphoid cells.
Common myeloid progenitor	MSC	Stem cells	Stem cells that give rise to different myeloid cells.
Granulocyte-Macrophage progenitor	GMP	Stem cells	Stem cells that give rise to macrophages.
Megakaryocyte-erythroid progenitors	MEP	Stem cells	Stem cell that gives rise to megakaryocytes and erythrocytes.
Multipotent progenitors	MPP	Stem cells	Stem cells that have the ability to differentiate into all cell types found in the blood but can no longer self-renew themselves.
Megakaryocytes	Megakaryocytes	Stem cells	A large bone marrow cell that is responsible for the production of platelets.
Erythrocytes	Erythrocytes	Stem cells	Red blood cells.
Platelets	Platelets	Stem cells	Involved in clotting.
Mesenchymal stem cells	Mesenchymal stem cells	Stem cells	Stem cell that can differentiate into a variety of stroma cells.
Adipocytes	Adipocytes	Stroma cells	Fat cell, energy storage.
Preadipocytes	Preadipocytes	Stroma cells	A collection of adipocytes.
Stroma cells	Stroma cells	Stroma cells	A collective term of connective tissue.
Fibroblasts	Fibroblasts	Stroma cells	Synthesises collagen and structural components in extracellular matrix.
Pericytes	Pericytes	Stroma cells	Contractile cells that line the capillaries an venules, and wrap around endothelial cells.

Endothelial	Endothelial	Stroma cells	Endothelial cells line the interior surface of vessels.
Lymphatic endothelial cells	ly Endothelial cells	Stroma cells	Endothelial cells line the interior surface of lymph vessels.
Microvascular endothelial cells	mv Endothelial cells	Stroma cells	Endothelial cells line the interior surface of capillaries.
Smooth muscle cells	Smooth muscle cells	Stroma cells	Muscle cells that are not satiated and smooth.
Chondrocytes	Chondrocytes	Stroma cells	Cells found in cartilage.
Osteoblasts	Osteoblasts	Stroma cells	Cells found in bone.
Skeletal muscle cells	Skeletal muscle cells	Stroma cells	Skeletal muscle cells, involved in movement.
Myocytes	Myocytes	Stroma Cells	Muscle cell.
Epithelial cells	Epithelial cells	Epithelial cells	Epithelial cells line the outer surface of organs and blood vessels.
Sebocytes	Sebocytes	Epithelial cells	Sebum producing epithelial cells
Keratinocytes	Keratinocytes	Skin cell	Keratinocytes are found in the outer most layer of the skin
Mesangial cells	Mesangial cells	Kidney cell	Mesangial cells are found in the kidney.
Hepatocytes	Hepatocytes	Liver cell	Hepatocytes cells are found in the liver.
Melanocytes	Melanocytes	Nerve cell	Nerve cell found in the skin.
Keratocytes	keratocytes	Stroma cell	Keratocytes are a type of fibroblast and are involved in mainlining the extracellular matrix.
Astrocytes	Astrocytes	Nerve cell	A type of nerve cell commonly found in the brain and in the spinal cord.
Neurones	Neurones	Nerve cell	A type of nerve cell, used to rally signals.

Appendix B

Table 9. ER status and cell type infiltration profiles. Table showing for which cohort the score for a specific cell type is significantly different: dark grey: significant t-test and higher score in ER negative samples, light grey: t-test significant and score higher in ER positive samples, Insignificant t-test are left blank.

Type	MAINZ	MDACC	Metabric	Micma	Neoava	OSLOR	OSLO2	STAM	STK	TAI	TCGA	TRANSBIG	UPP	VDX
aDC														
DC														
iDC														
Memory B-cells														
Tgd cells														
Th2 cells														
Keratinocytes														
Macrophages M1														
Th1 cells														
Mast cells														
Osteoblast														
B-cells														
CD4+ memory T-cells														
CD8+ T-cells														
Melanocytes														
Monocytes														
pDC														
pro B-cells														
Tregs														
Chondrocytes														
Fibroblasts														
CD4+ Tcm														
CD4+ Tem														
CD8+ Tem														
cDC														
Erythrocytes														
naive B-cells														
NK cells														
Plasma cells														
HSC														
Sebocytes														
CMP														
Hepatocytes														
MSC														
CD4+ naive T-cells														
Macrophages														
Neutrophils														
Neurons														
Smooth muscle														
Adipocytes														
Basophils														
CD8+ Tcm														
MEP														
CD4+ T-cells														
Class-switched memory B-cells														
GMP														
Megakaryocytes														
CLP														
Skeletal muscle														
Myocytes														
Endothelial cells														
ly Endothelial cells														
Eosinophils														
CD8+ naive T-cells														
Mesangial cells														
Pericytes														
Preadipocytes														
Epithelial cells														
MPP														
mv Endothelial cells														
Macrophages M2														
Astrocytes														
NKT														
Platelets														

Appendix C

Table 10. Luminal B vs Luminal A. Table showing for which cohort the score for a specific cell type is significantly different: dark blue: significant t-test and higher score in Luminal A samples, light blue: t-test significant and score higher in Luminal B samples, Insignificant t-test are left blank.

Type	MAINZ	MDACC	Metabric	Micma	Neoava	OSLOR	OSLO2	STAM	STK	TAI	TCGA	TRANSBIG	UPP	VDX
CLP														
MEP														
Tgd cells														
Th1 cells														
Th2 cells														
Adipocytes														
Astrocytes														
Chondrocytes														
Endothelial cells														
HSC														
Keratinocytes														
ly Endothelial cells														
Hepatocytes														
Megakaryocytes														
Mesangial cells														
Macrophages														
Macrophages M1														
pro B-cells														
Fibroblasts														
CD4+ memory T-cells														
mv Endothelial cells														
Erythrocytes														
Osteoblast														
cDC														
CMP														
Sebocytes														
Neurons														
CD8+ naive T-cells														
MSC														
Platelets														
aDC														
Smooth muscle														
Eosinophils														
Tregs														
Macrophages M2														
Preadipocytes														
CD4+ naive T-cells														
Epithelial cells														
Mast cells														
Class-switched memory B-cells														
naive B-cells														
Melanocytes														
Pericytes														
NKT														
CD4+ Tcm														
pDC														
CD8+ T-cells														
iDC														
Memory B-cells														
MPP														
B-cells														
CD4+ Tem														
DC														
Neutrophils														
Plasma cells														
Myocytes														
CD4+ T-cells														
GMP														
Basophils														
CD8+ Tcm														
CD8+ Tem														
Monocytes														
NK cells														
Skeletal muscle														

Table 11. Luminal A vs Her2-enriched. Table showing for which cohort the score for a specific cell type is significantly different: dark blue: significant t-test and higher score in Luminal A samples, pink: t-test significant and score higher in Her2-enriched samples, Insignificant t-test are left blank.

Type	MAINZ	MDACC	Metabric	Micma	Neoava	OSLOR	OSLO2	STAM	STK	TAI	TCGA	TRANSBG	UPP	VDX
Macrophages M1														
Tgd cells														
Th1 cells														
Th2 cells														
aDC														
Plasma cells														
Mast cells														
CD4+ memory T-cells														
CD4+ Tem														
Macrophages														
naive B-cells														
pro B-cells														
Tregs														
CMP														
HSC														
B-cells														
DC														
iDC														
pDC														
Neurons														
CD4+ T-cells														
Memory B-cells														
MEP														
NK cells														
Adipocytes														
Chondrocytes														
Fibroblasts														
ly Endothelial cells														
CD8+ Tem														
Keratinocytes														
Macrophages M2														
Monocytes														
Endothelial cells														
CD4+ naive T-cells														
CD8+ T-cells														
CD8+ Tcm														
Neutrophils														
Class-switched memory B-cells														
Megakaryocytes														
CD4+ Tcm														
Erythrocytes														
Sebocytes														
mv Endothelial cells														
Astrocytes														
Mesangial cells														
CLP														
NKT														
Basophils														
Hepatocytes														
MPP														
Osteoblast														
Smooth muscle														
CD8+ naive T-cells														
Eosinophils														
Pericytes														
Platelets														
Melanocytes														
Preadipocytes														
cDC														
Epithelial cells														
Skeletal muscle														
GMP														
MSC														
Myocytes														

Table 12. Luminal A vs Basal-like. Table showing for which cohort the score for a specific cell type is significantly different: dark blue: significant t-test and higher score in Luminal A samples, red: t-test significant and score higher in Basal-like samples, Insignificant t-test are left blank.

Type	MAINZ	MDACC	Metabric	Micma	Neoava	OSLOR	OSLO2	STAM	STK	TAI	TCGA	TRANSBIG	UPP	VDX
CD4+ memory T-cells														
Macrophages M1														
pro B-cells														
Tgd cells														
Th1 cells														
Th2 cells														
Chondrocytes														
Fibroblasts														
HSC														
Mast cells														
aDC														
DC														
Melanocytes														
Tregs														
Osteoblast														
B-cells														
Erythrocytes														
Keratinocytes														
Memory B-cells														
naive B-cells														
Adipocytes														
Hepatocytes														
ly Endothelial cells														
CD8+ T-cells														
NK cells														
CMP														
Endothelial cells														
Mesangial cells														
mv Endothelial cells														
MEP														
Neurons														
CD4+ Tcm														
GMP														
Plasma cells														
CD8+ Tem														
Monocytes														
Sebocytes														
MSC														
Basophils														
CD4+ Tem														
Macrophages														
Neutrophils														
Megakaryocytes														
Smooth muscle														
CD4+ naive T-cells														
cDC														
Class-switched memory B-cells														
iDC														
Astrocytes														
Platelets														
Preadipocytes														
pDC														
Skeletal muscle														
Eosinophils														
MPP														
CD4+ T-cells														
CD8+ naive T-cells														
CD8+ Tcm														
Macrophages M2														
Epithelial cells														
NKT														
Pericytes														
CLP														
Myocytes														

Table 13. Luminal A vs Normal-like. Table showing for which cohort the score for a specific cell type is significantly different: dark blue: significant t-test and higher score in Luminal A samples, green: t-test significant and score higher in Normal-like samples, Insignificant t-test are left blank.

Type	MAINZ	MDACC	Metabric	Micma	Neoava	OSLOR	OSLO2	STAM	STK	TAI	TCGA	TRANSBIG	UPP	VDX
Keratinocytes														
Megakaryocytes														
Osteoblast														
iDC														
CD4+ naive T-cells														
ly Endothelial cells														
Melanocytes														
CD8+ T-cells														
cDC														
DC														
mv Endothelial cells														
Astrocytes														
CD4+ Tcm														
Endothelial cells														
HSC														
Sebocytes														
aDC														
Platelets														
CD4+ Tem														
CD8+ Tem														
Mesangial cells														
Monocytes														
CLP														
MSC														
Smooth muscle														
Hepatocytes														
Memory B-cells														
MPP														
NK cells														
pro B-cells														
Epithelial cells														
B-cells														
Eosinophils														
Chondrocytes														
Adipocytes														
Basophils														
CD4+ T-cells														
CMP														
GMP														
Macrophages M1														
naive B-cells														
Neurons														
Neutrophils														
Plasma cells														
Preadipocytes														
Erythrocytes														
Macrophages M2														
Macrophages														
Th1 cells														
CD4+ memory T-cells														
CD8+ Tcm														
Skeletal muscle														
pDC														
Pericytes														
Tgd cells														
CD8+ naive T-cells														
Class-switched memory B-cells														
Fibroblasts														
Tregs														
Mast cells														
MEP														
NKT														
Myocytes														
Th2 cells														

Table 14. Luminal B vs Normal-like. Table showing for which cohort the score for a specific cell type is significantly different: light blue: significant t-test and higher score in Luminal B samples, green: t-test significant and score higher in Normal-like samples, Insignificant t-test are left blank.

Type	MAINZ	MDACC	Metabric	Micma	Neoava	OSLOR	OSLO2	STAM	STK	TAI	TCGA	TRANSBIG	UPP	VDX
Astrocytes														
Keratinocytes														
Megakaryocytes														
Melanocytes														
Tgd cells														
Th1 cells														
Th2 cells														
Hepatocytes														
HSC														
Sebocytes														
cDC														
Chondrocytes														
Endothelial cells														
ly Endothelial cells														
mv Endothelial cells														
Platelets														
CLP														
Adipocytes														
CD4+ naive T-cells														
CD8+ T-cells														
iDC														
Mesangial cells														
MEP														
Osteoblast														
CMP														
DC														
Eosinophils														
Smooth muscle														
CD4+ Tcm														
MPP														
Neurons														
Macrophages M1														
Memory B-cells														
CD8+ Tem														
Fibroblasts														
Monocytes														
aDC														
B-cells														
Neutrophils														
Plasma cells														
Preadipocytes														
pro B-cells														
Erythrocytes														
Basophils														
CD4+ Tem														
CD8+ Tcm														
NK cells														
Macrophages M2														
CD4+ T-cells														
GMP														
Mast cells														
naive B-cells														
Skeletal muscle														
Pericytes														
Class-switched memory B-cells														
Myocytes														
NKT														
pDC														
CD4+ memory T-cells														
CD8+ naive T-cells														
Epithelial cells														
Tregs														

Table 15. Luminal B vs Basal-like. Table showing for which cohort the score for a specific cell type is significantly different: light blue: significant t-test and higher score in Luminal B samples, red: t-test significant and score higher in Basal-like samples, Insignificant t-test are left blank.

Type	MAINZ	MDACC	Metabric	Micma	Neoava	OSLOR	OSLO2	STAM	STK	TAI	TCGA	TRANSBIG	UPP	VDX
Keratinocytes														
cDC														
Melanocytes														
Sebocytes														
CLP														
Osteoblast														
aDC														
CD8+ T-cells														
Tgd cells														
Th2 cells														
MSC														
Smooth muscle														
DC														
Memory B-cells														
Th1 cells														
Tregs														
Mast cells														
B-cells														
CD4+ naive T-cells														
CD4+ Tcm														
Astrocytes														
CD8+ Tem														
GMP														
iDC														
pro B-cells														
Hepatocytes														
Basophils														
CD4+ memory T-cells														
Megakaryocytes														
Monocytes														
naive B-cells														
NK cells														
Preadipocytes														
Macrophages M1														
Plasma cells														
CMP														
Fibroblasts														
CD8+ Tcm														
Neutrophils														
pDC														
Macrophages M2														
CD4+ Tem														
CD4+ T-cells														
Erythrocytes														
Mesangial cells														
Myocytes														
Adipocytes														
Eosinophils														
CD8+ naive T-cells														
Class-switched memory B-cells														
Skeletal muscle														
Pericytes														
MEP														
NKT														
Epithelial cells														
ly Endothelial cells														
Platelets														
MPP														
Endothelial cells														
mv Endothelial cells														
Chondrocytes														
HSC														
Macrophages														
Neurons														

Table 16. Luminal B vs Her2-enriched. Table showing for which cohort the score for a specific cell type is significantly different: light blue: significant t-test and higher score in Luminal B samples, pink: t-test significant and score higher in Her2-enriched samples, Insignificant t-test are left blank.

Type	MAINZ	MDACC	Metabric	Micma	Neoava	OSLOR	OSLO2	STAM	STK	TAI	TCGA	TRANSBIG	UPP	VDX
iDC														
Keratinocytes														
pDC														
Sebocytes														
aDC														
cDC														
DC														
Megakaryocytes														
Plasma cells														
CLP														
B-cells														
Mast cells														
Osteoblast														
CD4+ naive T-cells														
CD4+ Tem														
Memory B-cells														
Astrocytes														
Endothelial cells														
Mesangial cells														
naive B-cells														
NK cells														
pro B-cells														
Adipocytes														
CD4+ T-cells														
CD8+ T-cells														
CD8+ Tcm														
CD8+ Tem														
Monocytes														
Neutrophils														
Tregs														
Smooth muscle														
CD4+ Tcm														
Eosinophils														
Epithelial cells														
ly Endothelial cells														
Macrophages M1														
mv Endothelial cells														
Tgd cells														
Class-switched memory B-cells														
Th2 cells														
MSC														
CD4+ memory T-cells														
Hepatocytes														
Macrophages M2														
Platelets														
Pericytes														
Chondrocytes														
Macrophages														
Melanocytes														
Th1 cells														
Fibroblasts														
HSC														
CMP														
GMP														
Myocytes														
NKT														
Skeletal muscle														
Erythrocytes														
CD8+ naive T-cells														
MEP														
Neurons														
Basophils														
Preadipocytes														
MPP														

Table 17. Her2-enriched vs Basal-like. Table showing for which cohort the score for a specific cell type is significantly different: red: significant t-test and higher score in Basal-link samples, pink: t-test significant and score higher in Her2-enriched samples, Insignificant t-test are left blank.

Type	MAINZ	MDACC	Metabric	Micma	Neoava	OSLOR	OSLO2	STAM	STK	TAI	TCGA	TRANSBIG	UPP	VDX
Melanocytes														
Th1 cells														
Hepatocytes														
Preadipocytes														
GMP														
Macrophages M2														
Osteoblast														
Keratinocytes														
NKT														
MSC														
Th2 cells														
Adipocytes														
Epithelial cells														
Fibroblasts														
Macrophages														
mv Endothelial cells														
Tgd cells														
Endothelial cells														
Basophils														
Chondrocytes														
Eosinophils														
Pericytes														
Plasma cells														
CLP														
Erythrocytes														
Tregs														
iDC														
CD4+ Tem														
CMP														
ly Endothelial cells														
Mast cells														
Megakaryocytes														
Platelets														
Astrocytes														
CD4+ memory T-cells														
CD8+ T-cells														
cDC														
Neurons														
aDC														
Sebocytes														
Macrophages M1														
Mesangial cells														
pDC														
CD4+ naive T-cells														
CD8+ naive T-cells														
DC														
Monocytes														
Myocytes														
Skeletal muscle														
CD4+ T-cells														
CD4+ Tcm														
HSC														
MEP														
Smooth muscle														
B-cells														
CD8+ Tcm														
CD8+ Tem														
Class-switched memory B-cells														
Memory B-cells														
MPP														
naive B-cells														
Neutrophils														
NK cells														
pro B-cells														

Table 18. Her2-enriched vs Normal-like. Table showing for which cohort the score for a specific cell type is significantly different: green: significant t-test and higher score in Normal-like samples, pink: t-test significant and score higher in Her2-enriched samples, Insignificant t-test are left blank.

Type	MAINZ	MDACC	Metabric	Micma	Neoava	OSLO_R	OSLO2	STAM	STK	TAI	TCGA	TRANSBIG	UPP	VDX
HSC														
Tgd cells														
Th1 cells														
Megakaryocytes														
Th2 cells														
Chondrocytes														
Endothelial cells														
ly Endothelial cells														
mv Endothelial cells														
Neurons														
Astrocytes														
CMP														
Hepatocytes														
Macrophages														
Mast cells														
Melanocytes														
Adipocytes														
Mesangial cells														
Platelets														
CLP														
Macrophages M1														
MEP														
Keratinocytes														
MSC														
Osteoblast														
Tregs														
cDC														
Fibroblasts														
GMP														
aDC														
pro B-cells														
MPP														
Smooth muscle														
Epithelial cells														
Plasma cells														
CD4+ naive T-cells														
CD8+ T-cells														
CD8+ naive T-cells														
CD4+ memory T-cells														
Erythrocytes														
Macrophages M2														
Sebocytes														
Basophils														
DC														
Eosinophils														
Monocytes														
Preadipocytes														
B-cells														
NK cells														
pDC														
iDC														
Myocytes														
NKT														
CD4+ T-cells														
Class-switched memory B-cells														
Memory B-cells														
naive B-cells														
Pericytes														
CD4+ Tcm														
CD4+ Tem														
CD8+ Tcm														
CD8+ Tem														
Neutrophils														
Skeletal muscle														

Table 19. Normal-like vs Basal-like. Table showing for which cohort the score for a specific cell type is significantly different: green: significant t-test and higher score in Normal-like samples, red: t-test significant and score higher in Basal-like samples, Insignificant t-test are left blank.

Type	MAINZ	MDACC	Metabric	Micma	Neoava	OSLOR	OSLO2	STAM	STK	TAI	TCGA	TRANSBIG	UPP	VDX
HSC														
Megakaryocytes														
Tgd cells														
Th1 cells														
Th2 cells														
Chondrocytes														
Hepatocytes														
Endothelial cells														
ly Endothelial cells														
mv Endothelial cells														
Astrocytes														
CMP														
Platelets														
Fibroblasts														
Adipocytes														
Mast cells														
Mesangial cells														
Neurons														
Macrophages M1														
MEP														
CLP														
Macrophages														
Tregs														
MPP														
Preadipocytes														
Erythrocytes														
Eosinophils														
CD4+ memory T-cells														
pro B-cells														
aDC														
NKT														
CD4+ naive T-cells														
iDC														
Sebocytes														
MSC														
Osteoblast														
CD8+ T-cells														
cDC														
CD4+ T-cells														
Epithelial cells														
Macrophages M2														
Pericytes														
Basophils														
CD4+ Tcm														
CD4+ Tem														
CD8+ Tcm														
Monocytes														
Myocytes														
B-cells														
CD8+ naive T-cells														
CD8+ Tem														
Class-switched memory B-cells														
GMP														
Memory B-cells														
naive B-cells														
NK cells														
pDC														
Skeletal muscle														
Smooth muscle														
DC														
Keratinocytes														
Melanocytes														
Neutrophils														
Plasma cells														