# Tumor cell heterogeneity profiling using single-cell analysis

Emma Jonasson

Department of Laboratory Medicine

Institute of Biomedicine

Sahlgrenska Academy, University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2020

Cover illustration by Emma Jonasson

To my family

# ABSTRACT

Cancer is a diverse disease with large variations between tumor types and patients regarding tumor progression and prognosis. Additionally, most individual tumors are heterogeneous, containing subpopulations of cells with various characteristics. Numerous factors affect the differences observed between tumor cells, such as variations in genetics, epigenetics, cellular states and the microenvironment surrounding the individual cells. One clinically relevant subpopulation, commonly referred to as cancer stem cells, consists of cells with stem cell characteristics. These are present in many tumor types and are known to be important for tumor development and treatment resistance. The tumor microenvironment is a key factor affecting the cellular phenotype, including the cancer stem cell subpopulation. Analysis at cell population level will not capture the true variations between individual cells. Instead, single-cell analysis offers new means to study and understand cellular and molecular differences between tumor subpopulations. The main objective of this thesis was to study tumor cell heterogeneity in myxoid liposarcoma and breast cancer with the help of single-cell gene expression analysis methods. We could generate a flexible workflow to measure gene expression, including the assessment of total mRNA amounts in each cell, using several diverse approaches developed from already existing protocols. Subsequently, we combined a number of functional cell culture methods to enrich for tumor cells with characteristic cellular properties together with single-cell gene expression profiling methods, to match phenotype with the corresponding transcription pattern. Single-cell analysis of myxoid liposarcoma cells, sorted based on the cell-cycle, identified a number of genes previously not reported as cell-cycle regulated and defined two subgroups of cells within the G1 phase. In the same tumor type, we identified a subpopulation of cells with cancer stem cell- and chemotherapy resistance properties associated with an active JAK-STAT signaling pathway. Here, a combination treatment of chemotherapy and JAK-STAT inhibition was *in vitro* shown to be more effective against tumor cells than chemotherapy alone. In breast cancer cells, we identified a number of potential biomarkers overexpressed in a subpopulation of cells with cancer stem cell characteristics. We also developed a new *in vivo*-like culture system based on decellularized human tumors to study the effect of the microenvironment on breast cancer cells. We demonstrated that the gene expression profiles of cells cultured in these patient-derived scaffolds closely mimic the profiles of *in vivo* cells. Furthermore, gene expression patterns changed differently depending on the patient-derived scaffold, which could be linked to patient recurrence. In conclusion, we developed single-cell analysis methods as well as a new *in vivo*-like model system. Furthermore, we identified genes and pathways connected to different subpopulations of myxoid liposarcoma or breast cancer cells that potentially can be used as biomarkers and future drug targets.

**Keywords**: tumor heterogeneity, single-cell analysis, cancer stem cells

# POPULÄRVETENSKAPLIG SAMMANFATTNING

Cancer är en av de huvudsakliga dödsorsakerna i världen. Prognos och behandlingsrespons beror på cancertyp men kan även variera mellan patienter med samma cancersjukdom. Det har också visat sig att en och samma tumör kan variera i sammansättning och bestå av flera grupper av cancerceller med specifika egenskaper. En hypotes är att det finns en mindre population av cancerceller med stamcellsegenskaper, de så kallade cancerstamcellerna, som är viktiga för tumörtillväxt och behandlingsresistens. Flera faktorer påverkar skillnaden mellan de enskilda cellerna i tumören, såsom genetiska förändringar och den omgivande miljön. Genom att analysera enskilda celler, bland annat genom att mäta genuttryck, kan man undersöka hur olika grupper av cancerceller skiljer sig åt. Vi har vidareutvecklat existerande metoder för att mäta genuttryck i enskilda celler. Våra studier har fokuserat på två olika cancertyper, bröstcancer och myxoid liposarkom, där den senare är en ovanlig cancertyp som vanligen uppkommer i lårets muskelvävnad. För att studera dessa har vi använt oss av cellinjer, modellsystem baserade på celler från tumörer som i ett laboratorium kan fås att dela sig oändligt många gånger. Genom att odla cellinjer på varierande sätt och använda oss av infärgningsmetoder har vi kunnat anrika celler med specifika egenskaper. Vi har sedan analyserat genuttrycket hos dessa celler. Genom att använda oss av dessa och andra metoder har vi kunnat upptäcka en tidigare okänd grupp i en myxoid liposarkom-cellinje med cancerstamcellsegenskaper vilka var förknippade med uttryck av gener som tillhör JAK-STAT, en specifik signalväg som är inblandad i tumörutveckling. En kombination av ett läkemedel riktat mot denna signalväg och en vanlig typ av cellgift som används för att behandla patienter drabbade av myxoid liposarkom visades vara mer effektivt i myxoid liposarkom-cellinjer än bara cellgiftbehandlingen. I bröstcancer har existensen av cancerstamceller varit känd i flera år men de biomarkörer som idag används för att identifiera dessa varierar och är inte uttryckta i alla cancerstamceller. Genom att kombinera specifika odlings- och infärgningsmetoder av en bröstcancercellinje med encellsanalys kunde vi identifiera flera gener som var högre uttryckta i celler med mer cancerstamcellsegenskaper. Vi undersökte också effekten av olika tumöromgivningar på bröstcancerceller. Vi utvecklade ett nytt odlingssystem där vi tvättade bort celler från tumörbitar för att sedan låta cancercellinjer växa i kvarvarande material. Vi kunde se att celler odlade i dessa strukturer blev mer tumörlika i jämförelse med celler odlade i konventionella odlingsflaskor. Vi kunde också se att sättet vissa genuttryck förändrades hos de odlade cellerna varierade och kunde kopplas till återfall hos patienterna som tumörbitarna tagits från. Sammanfattningsvis så har vi tagit fram metoder för att studera genuttryck på encellsnivå och även tagit fram ett nytt cellodlingssystem. Vi har identifierat gener och signalvägar specifika för olika cellpopulationer som, efter mer studier, potentiellt skulle kunna användas vid utveckling av nya metoder för diagnostik och behandling.

# LIST OF PAPERS

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I.    Kroneis T, **Jonasson E**, Andersson D, Dolatabadi S, Ståhlberg A. Global preamplification simplifies targeted mRNA quantification. *Scientific reports, 2017. 7:45219.*

II.   **Jonasson E**, Andersson L, Dolatabadi S, Ghannoum S, Ståhlberg A. Total mRNA quantification in single cells. *Manuscript*

III.  Karlsson J, Kroneis T, **Jonasson E**, Larsson E, Ståhlberg A. Transcriptomic Characterization of the Human Cell Cycle in Individual Unsynchronized Cells. *Journal of Molecular Biology, 2017. 429(24):3909-24.*

IV.   Dolatabadi S, **Jonasson E**, Lindén M, Fereydouni B, Bäcksten K, Nilsson M, Martner A, Forootan A, Fagman H, Landberg G, Åman P, Ståhlberg A. JAK–STAT signalling controls cancer stem cell properties including chemotherapy resistance in myxoid liposarcoma. *International Journal of Cancer, 2019. 145(2):435-49.*

V.    **Jonasson E**, Ghannoum S, Persson E, Karlsson J, Kroneis T, Larsson E, Landberg G, Ståhlberg A. Identification of Breast Cancer Stem Cell Related Genes Using Functional Cellular Assays Combined With Single-Cell RNA Sequencing in MDA-MB-231 Cells. *Frontiers in genetics, 2019. 10:500.*

VI.   Landberg G, Fitzpatrick P, Isakson P, **Jonasson E**, Karlsson J, Larsson E, Svanström A, Rafnsdottir S, Persson E, Gustafsson A, Andersson D, Rosendahl J, Petronis S, Ranji P, Gregersson P, Magnusson Y, Håkansson J, Ståhlberg A. Patient-derived scaffolds uncover breast cancer promoting properties of the microenvironment. *Biomaterials, 2020. 235:119705.*

Additional publications not part of this thesis:

i. Ståhlberg A, Gustafsson CK, Engtröm K, Thomsen C, Dolatabadi S, **Jonasson E**, Li CY, Ruff D, Chen SM, Åman P. Normal and functional TP53 in genetically stable myxoid/round cell liposarcoma. *PLoS ONE, 2014. 9(11):e113110.*

ii. Safavi S, Jarnum S, Vannas C, Udhane S, **Jonasson E**, Tomic TT, Grundevik P, Fagman H, Hansson M, Kalender Z, Jauhiainen A, Dolatabadi S, Stratford EW, Myklebost O, Eriksson M, Stenman G, Schneider-Stock R, Ståhlberg A, Åman P. HSP90 inhibition blocks ERBB3 and RET phosphorylation in myxoid/round cell liposarcoma and causes massive cell death in vitro and in vivo. *Oncotarget, 2016. 7(1):433-45.*

iii. Åman P, Dolatabadi S, Svec D, **Jonasson E**, Safavi S, Andersson D, Grundevik P, Thomsen C, Ståhlberg A. Regulatory mechanisms, expression levels and proliferation effects of the FUS-DDIT3 fusion oncogene in liposarcoma. *The Journal of pathology, 2016. 238(5):689-99.*

iv. Lindén M, Thomsen C, Grundevik P, **Jonasson E**, Andersson D, Runnberg R, Dolatabadi S, Vannas C, Luna Santamaría M, Fagman H, Ståhlberg A, Åman P. FET family fusion oncoproteins target the SWI/SNF chromatin remodeling complex. *EMBO Rep, 2019. 20(5):e45766.*

# CONTENT

x

# ABBREVIATIONS

| | |
|---|---|
| CSC | Cancer stem cell |
| ECM | Extracellular matrix |
| EMT | Epithelial-mesenchymal transition |
| ER | Estrogen receptor |
| MLS | Myxoid liposarcoma |
| PCA | Principal component analysis |
| PDS | Patient-derived scaffold |
| qPCR | Quantitative polymerase chain reaction |
| RPM | Reads per million |
| SOM | Self-organizing map |
| t-SNE | t-distributed stochastic neighbor embedding |

# INTRODUCTION

Cancer is one of the leading causes of death worldwide and both incidence and mortality is increasing [1]. During the last decades, the understanding of the biology of cancer have markedly increased, but little of this has been translated into new prevention and treatment strategies [2]. Cancer is a group of diseases where cells, through mainly genetic changes, gain properties that allow them to proliferate more than normal cells. The amount and type of genetic changes varies between tumor types, patients and during the course of tumor progression [3]. The main treatments of cancer include surgery to remove the tumor, radiotherapy directed at the tumor site and chemotherapy; cytotoxic drugs that aims to kill highly proliferative cancer cells while causing little harm to normal cells. As our knowledge about altered signaling pathways has increased, new possibilities for development of directed treatments have increased accordingly, though with varying results [2, 4]. Both prognosis and treatment response commonly varies between patients with the same cancer type and there are several issues associated with today's treatments. Apart from side effects and lack of treatment effect, there is also a problem with treatment resistance, when relapses are occurring despite initial response to treatment [5]. To complicate things further, cancer cells within one tumor differ from each other where some cells seem to be more important for treatment resistance [6]. Increased knowledge about tumor cell heterogeneity could be the key to understanding the difference in treatment response between patients and to design new and more efficient treatments.

## CANCER SUBTYPES

Cancer is divided into subtypes, normally depending on their tissue of origin [2]. The most common subtype is carcinoma, which comprises tumors arising from epithelial tissues in different parts of the body. Further subtypes include; sarcomas from different types of connective tissue, hematopoietic cancers such as leukemias and lymphomas from blood-forming tissues, neuroectodermal tumors from cells of the nervous system, and melanomas from the pigmented cells of the skin and retina [2]. Both the amount and type of genetic alterations varies between different cancer types [3]. Apart from point mutations, other chromosomal changes include changes in chromosome number as well as deletions, inversions and translocations, the latter often resulting in fusion of two genes to create a fusion oncogene. It is thought that just a few of the genetic changes are important drivers of tumor progression whereas many are considered passengers [3].

In this thesis, we have focused on two different cancer types, breast cancer and myxoid liposarcoma (MLS). While breast cancer on average have around 30 mutations [3] and consists of several different subtypes [7, 8], MLS is driven by specific fusion oncogenes

and harbors few other genetic changes [9, 10] leading to less genetic variations between patients.

## Breast cancer

Breast cancer is the most common cancer type affecting women worldwide and it is one of the main causes of cancer-related deaths in women. The risk of developing breast cancer before the age of 75 is 5%. The incidence is higher in more developed countries, mostly due to a higher prevalence of known risk factors, for example related to reproduction, hormone intake and nutrition [1]. In Sweden, the corresponding value is 10%, according to the Swedish Cancer Society [11]. Breast cancer is a heterogeneous disease with several subtypes connected with different prognosis and treatment. Even though the overall survival of breast cancer has increased due to early detection and improved treatment, with a current 5-year survival rate of 89% in the US [12] and 90% in Sweden [11], there are subtypes with worse prognosis [7].

Currently, breast cancer can be divided into four intrinsic subtypes; Luminal A, Luminal B, HER2-enriched and basal-like. Studies have also shown heterogeneity within these subtypes [13-15]. The subtypes are further defined based on the expression pattern of proliferation marker Ki67, HER2, and the hormone receptors estrogen receptor (ER) and progesterone receptor (PgR) [8]. The basal-like subtype usually lacks expression of HER2 and the hormone receptors and is often called triple-negative breast cancer, although not all triple-negative tumors are basal-like. These expression patterns, together with other factors such as the size and location of the tumor as well as age and health status of the patient are guiding treatment decisions. Treatment of breast cancer includes surgery and radiation therapy as well as different types of systemic treatment including chemotherapy, endocrine therapy (for treatment of ER positive tumors) and HER2-directed treatment (for treatment of tumors with HER2 amplification) [8]. Triple-negative tumors, lacking specific treatment, are often aggressive with a poor outcome [16].

## Myxoid liposarcoma

Myxoid liposarcoma is a rare type of tumor belonging to the group of soft tissue sarcomas, which are malignant tumors arising in soft tissues such as fat or muscle in different parts of the body. Most MLS tumors are situated in the muscle of the thigh. The incidence of soft tissue sarcomas is less than 1% of all malignant tumors, although more common in children. MLS affects mainly young adults and accounts for approximately 5% of soft tissue sarcomas in adults. [17].

Myxoid liposarcoma is characterized by specific chromosomal translocations resulting in fusion oncogenes, most commonly t(12;16)(q13;p11) resulting in a fusion between the gene *FUS* and the transcription factor gene *DDIT3*. Some cases instead has the translocation t(12;22)(q13;q12) resulting in the fusion oncogene *EWSR1-DDIT3* [17-19].

Myxoid liposarcoma is one of a group of genetically related tumors that are caused by fusion oncogenes formed between one of the FET genes *FUS*, *EWSR1* or *TAF15* combined with different transcription factor-coding genes (Figure 1). The three FET genes can in many cases be replaced with each other whereas the transcription factor partners are specific for each tumor type [20, 21].
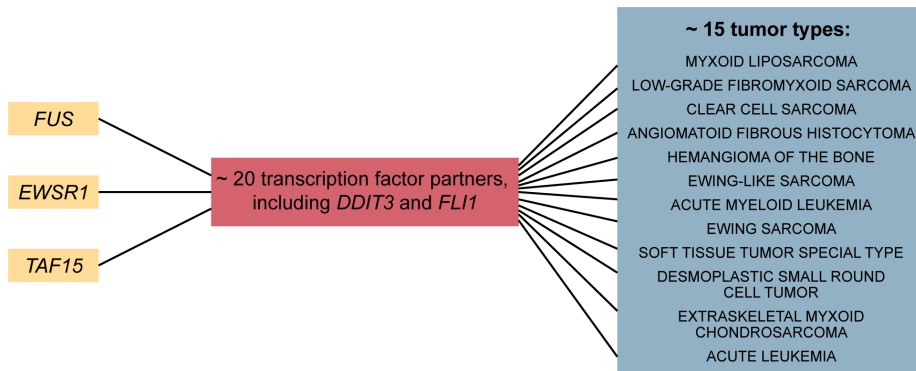


Figure 1. Tumors defined by FET fusion oncogenes. Any of the three FET genes, *FUS*, *EWSR1* or *TAF15*, is fused to one of more than 20 transcription factors partners, resulting in a specific tumor type defined by the transcription factor. Adapted from [21].

Apart from the fusion oncogene, MLS tumors have few other genetic changes [9, 22]. Histologically, MLS is composed of round or oval-shaped cells together with lipoblasts in a myxoid surrounding containing a branching vascular pattern [17]. Some tumors display a round-cell morphology which is connected to poorer prognosis and tumors with more than 5% round cells significantly correlates with increased metastasis and death. A retrospective study of 418 cases showed a 5-year disease-specific survival rate of 91% in pure myxoid tumors and 79% in round-cell tumors [23]. Myxoid liposarcoma is mainly treated with surgery, often combined with radiotherapy. Some patients with high-risk disease, based on size of the tumor and round-cell morphology, are also treated with chemotherapy [24].

## TUMOR CELL HETEROGENEITY

The human body is composed of trillions of cells that are of different types and with various functions. Also within each cell type, there are variations between individual cells [25]. The same is true for tumors, which can contain several subpopulations of cells with different characteristics. The cell composition of a tumor can vary in different parts of the tumor tissue and can alter over the course of time, which can affect the therapy response. For instance, some cell types can be treatment resistant or can acquire resistance over time [26].

## Intrinsic and extrinsic heterogeneity

The intra-tumor heterogeneity originates from different sources as illustrated in Figure 2. [25, 27, 28]. Several of these sources are intrinsic, such as the genomic aberrations within each cell. Many cancer cells have much higher mutation rates than normal cells that may result in considerable variation in mutations between cells during tumor progression [27]. Apart from the genetic variability, epigenetic programming also has a major impact. The epigenome involve modifications to chromatin which in turn leads to regulation of gene expression, and epigenetic changes are somatically heritable. Epigenetic profiles are influenced by mutations, and mutations affecting chromatin-regulating enzymes are present in many cancer types. For instance, 20% of all human tumors have mutations in genes encoding components of the SWI/SNF chromatin remodeling complex [29]. Among other things, the epigenome defines the differentiation state of a cell [30]. Hierarchies of stem-like cells as well as more differentiated cells, similarly to normal tissues, have been identified in several tumor types with great impact for tumor progression [28]. This difference in cellular states will influence the cellular phenotype, including the transcriptional profile. The same is true for more dynamic cellular states, for example different cell cycle phases, which can lead to considerable variation between cells when observing gene expression patterns [31]. Furthermore, stochastic gene expression, originating from the fact that mRNA is synthesized in random bursts, can result in more transient changes in a cell that give rise to variability between individual cells [32]. Apart from the intrinsic sources, there are extrinsic factors leading to cell heterogeneity coming from interactions with the microenvironment. The microenvironment consists of stromal cells, such as fibroblasts and immune cells, surrounding vasculature, extracellular matrix (ECM) and secreted molecules. The microenvironment of a tumor is an important driver of cancer progression and induced changes in the surrounding are necessary for the cancer cells to be able to proliferate, for example to receive enough nutrients [33-35]. Adding to this, there are several connections between intrinsic and extrinsic factors, for example that cells with certain mutations survive more easily in certain environments. This leads to the fact that some mutations are common in many cancer types and occur early during tumor development whereas some more tissue-specific mutations shows up later in cancer progression [27].
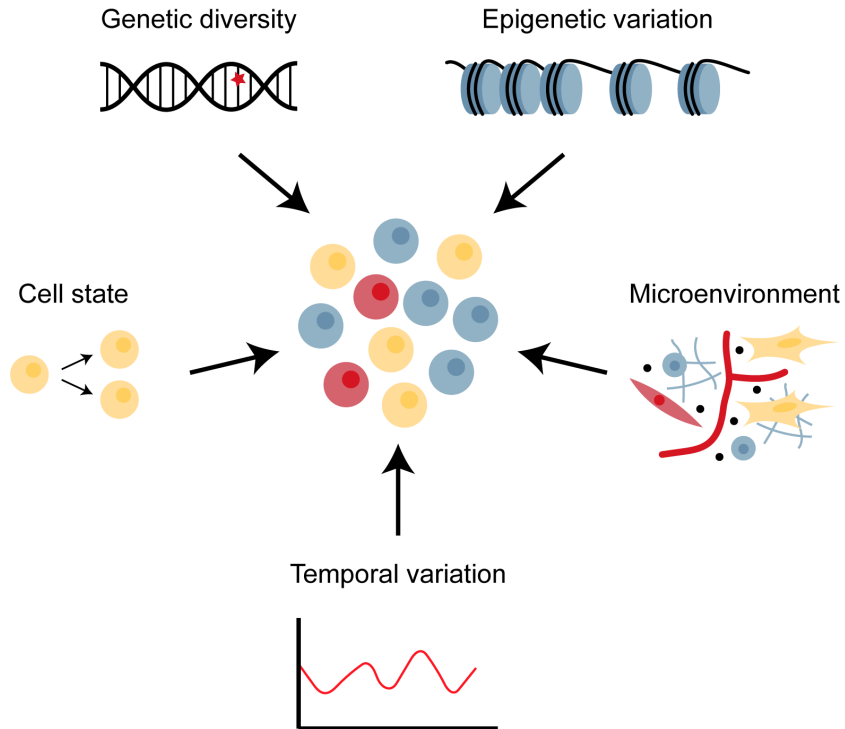
Figure 2. Tumor cell heterogeneity. Different factors affect the variations observed between tumor cells.

## Tumor heterogeneity models

Two models are commonly used to explain the formation of the main subpopulations important for tumor progression. The clonal evolution theory describes a model where a tumor progresses from one cell that have acquired advantageous genetic changes [36]. Due to genetic instability, more mutations will occur when the cells proliferate, which in turn can result in the formation of additional subgroups, or clones (Figure 3A). Due to the non-genetic features that affect the cell variation, other factors will influence the selection of cellular subpopulations and this model cannot completely explain tumor heterogeneity [28]. The other model suggests the presence of a small proportion of cells called cancer stem cells (CSCs) believed to drive tumor progression and therapy resistance [37-39]. Similarly to normal stem cells, the CSCs have self-renewing capacity and can give rise to progenitor cells and more differentiated cells (Figure 3B). These two models are not mutually exclusive but could together explain heterogeneity within tumors [38, 39].
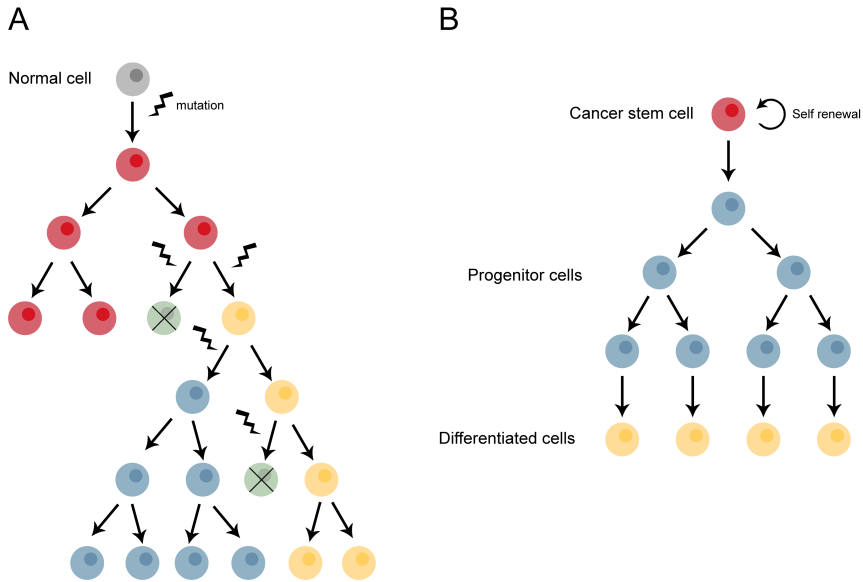
Figure 3. Tumor heterogeneity models. (A) The clonal evolution theory explains different subpopulations of tumor cells. Lightning bolts indicate mutational events and crossed-over green cells are non-surviving clones. Adapted from [36]. (B) The cancer stem cell model explains a hierarchical organization of different cell types. Adapted from [39].

## Cancer stem cell properties and functional assays

Cancer stem cells (CSCs) are believed to be resistant to many current treatments and despite that therapies target the tumor bulk, the small population of CSCs can survive, proliferate and eventually cause a relapse. It is therefore important to further characterize this subpopulation to be able to develop new treatment strategies that target these cells [6]. It has been shown that CSCs share properties with normal stem cells, and one suggestion is that they originate from normal stem cells that, through acquiring mutations, develop malignant phenotypes [6, 40]. However, studies have also demonstrated that there exist plasticity in the CSC model, indicating that more differentiated cancer cells can dedifferentiate into a CSC state [41, 42]. Dedifferentiation has for example been connected to epithelial-mesenchymal transition (EMT), a trait known to be associated with metastases [42, 43].

In order to study CSCs, several of their properties are used in functional assays to enrich for this subpopulation or to assess the proportion of CSCs in a larger cell population [39]. Cancer stem cells are also referred to as tumor-initiating cells based on their unique ability to cause tumor formation. Therefore, the xenograft assay, grafting tumor cells into immunodeficient mice, is often used to detect CSCs [44, 45]. Another widely used assay is the sphere formation assay, where single cells are cultured in non-adherent conditions

leading to sphere formation of the cells that survive anoikis, death induced by loss of contact with the ECM or neighboring cells. This type of assays include the mammosphere formation assay used in breast cancer research which has been shown to enrich for cells with self-renewing properties and the ability to differentiate [46]. Normal stem cells commonly show a slow-cycling, or even quiescent, phenotype, a trait which has also been shown to be present in the CSC population [40, 47, 48]. This property can be assessed with label-retention assays, where the ability of a cell to maintain a high intensity of an incorporated dye indicates few cell divisions [40]. Such label-retention assays can be used in cancer research to select for CSCs [47, 49, 50]. Further, normal stem cells also express ABC transporters, responsible for transporting compounds across membranes, which is speculated to be a mechanism for drug resistance in CSCs [6]. The ABC transporters also efflux dyes, such as Hoechst dye, and cells expressing these transporters are present in the so called side population (SP) during flow cytometry analysis. The SP cells have been shown to have increased CSC properties and this technique is therefore used to select for CSCs [51, 52].

## Tumor microenvironment

Not only cancer cells determine the course of tumor development. Interactions of cells with the microenvironment is crucial and deregulation of this interaction is needed for cancer cells to be able to proliferate [33]. The microenvironment change during tumor progression and influences all stages including invasion and metastasis [53]. This influence is mediated through several components of the microenvironment, including stromal cells and ECM [35]. Important stromal cells are immune cells, vascular cells and cancer-associated fibroblasts. Studies have shown that different types of immune cells can have either tumor-promoting or tumor-suppressing effects in various tumor types depending on cancer type and external stimuli [53]. It has also been shown that chronic inflammation is correlated to increased cancer risk [54]. Another important cancer-related process is angiogenesis, blood vessel formation, induced to supply the proliferating tumor cells with oxygen and nutrients. Angiogenesis is mediated by several stromal compartments, apart from the vascular cells. Furthermore, specific fibroblasts, termed cancer-associated fibroblasts, provide the tumor with signaling molecules that influence a number of tumor processes, including angiogenesis and immune responses [53]. Apart from the stromal cells, the ECM has many important functions, including creating a physical scaffold for the tumor and to allow signaling between cells. The ECM also promotes many cancer-related properties, such as invasion and migration, as well as creating a CSC niche. There are continuous interactions between the ECM and the different cell types present in the tumor and coordination between all these components is needed for cancer growth and metastasis [34].

# TUMOR MODEL SYSTEMS

When studying cancer biology, cell lines derived from specific tumor types are commonly used. Cell lines have many advantages as they can proliferate indefinitely, are easy to handle and can be expanded to high amounts, allowing many types of analyses to be performed. However, culturing cell lines in monolayer cultures also has limitations. For example, classical cell culture techniques select for certain cellular properties leading to a more homogenous population and cell lines have in some cases been demonstrated to change genotype and phenotype after long time in culture. Additionally, normal cell line cultures do not take into account the contribution of the tumor microenvironment [55]. *In vivo* mouse models better recapitulate human tumors. These models include cell line-derived xenografts where cancer cells are injected into mice, and patient-derived xenografts (PDXs), which are created from human tumor pieces being engrafted into mice. Finally, mice can also be genetically engineered to generate specific genotypes or phenotypes. However, all these mouse models have different advantages and disadvantages regarding costs and difficulty to establish, time consumption, and lack of resemblance to the human microenvironment [56]. Further, the ethical aspect of animal experiments must be considered. To better resemble the situation *in vivo* and to generate more tissue-like architectures *in vitro*, various types of three-dimensional (3D) model systems have emerged [57, 58]. These are normally based on either inducing aggregation of cells into 3D growth patterns or generating a 3D structure in which cells can be cultured. Culturing cells on non-adherent plates can cause spheroid formation which can lead to increased *in vivo*-like behavior in terms of cell-cell contact and formation of nutrient and oxygen gradients [58]. More recently, organoid cultures have gained interest. Organoids are formed as a result of spontaneous reassembly of cultured stem cells or primary cells, in combinations with added factors. This generates cellular structures that include various tissue-specific cell types and to some extent resemble the real organ which the cells were derived from [59, 60]. However, organoid cultures lack vital information of the tumor microenvironment and are normally dependent on matrigel to provide a supporting surrounding matrix [61]. Matrigel is a tumor-derived basement membrane extract from mice, which contains ECM proteins. Although matrigel has been shown to induce a more *in vivo*-like morphology in cells cultured therein, as well as promoting differentiation and stem cell survival [62], it suffers from batch variability and its composition is not fully known [57]. Three-dimensional cultures can also be carried out by adding cells on scaffolds created from different types of hydrogels or synthetic materials. The structures range from a thick gel layer to more advanced scaffold structures to provide physical support for the cells [58]. Recently, scaffolds derived from human tissues have been used to better recapitulate the *in vivo* microenvironments, as they provide unique architectures and cell-ECM connections [63-65].

# SINGLE-CELL GENE EXPRESSION ANALYSIS

It is important to consider the heterogeneity within a tumor when studying tumor cells. As illustrated in Figure 4A-C, different gene expression patterns among individual cells can result in similar results when measured as an average of a population. When performing bulk analysis, subpopulations of cells can therefore not be detected. Additionally, a change in gene expression within a small subpopulation, such as the CSCs, may be hidden by the larger population, since the molecules of the smaller population will be in minority. Single-cell analysis can be applied to more accurately define and profile each subpopulation.
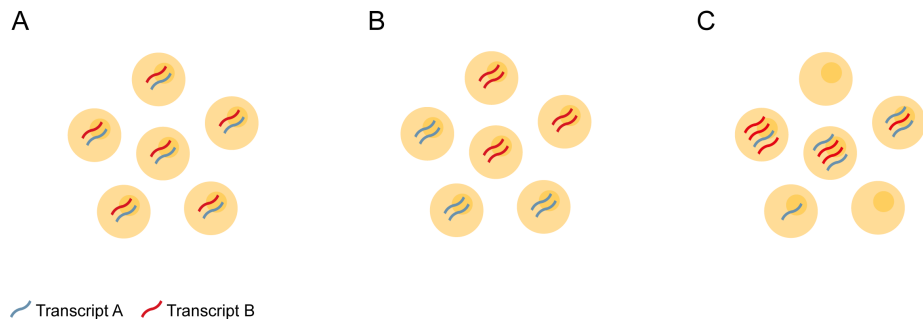


Figure 4. Gene expression heterogeneity. Even though the gene expression average in a cell population is the same, it can vary between the individual cells. (A) Expression of both gene *A* and gene *B* are the same in all cells. (B) Some cells only express gene *A*, while others only gene *B*. (C) Some cells express higher levels of both genes compared to other cells.

Single-cell gene expression analysis has been used in many studies of different topics, including cell type identification [66, 67], lineage patterns [68, 69] tumor microenvironment [70, 71], and therapy resistance in cancer [72, 73].

## Experimental methods

One of the first examples of single-cell gene expression analysis was performed in 1992, a gene expression profiling of rat neurons [74]. Many technological advances followed and in 2009, single-cell RNA sequencing could be performed [75]. At present, many single-cell RNA sequencing protocols are available with different advantages and for different applications [76]. For analysis of thousands of cells, droplet-based systems can be used [77, 78]. Here, individual cells are encapsulated in droplets together with beads containing primers with unique barcode sequences. One barcode sequence is specific for each individual cell, thereby allowing all cells to be processed together in a single reaction. Another barcode sequence, referred to as a unique molecular identifier (UMI), is specific for each individual transcript. The use of UMIs, which are common also in other type of RNA sequencing protocols, results in improved quantification of the number of

molecules as well as reduced noise [79]. A drawback of these methods is that sequencing is only performed at one end of the transcripts, causing reduced sensitivity [76]. Full-length protocols such as Smart-seq2 [80], on the other hand, been shown to have superior sensitivity compared to many other single-cell RNA sequencing protocols [79] and can also be used for detection of isoforms or splicing events [76]. However, the library preparation needs to be performed on each cell separately as they are not individually labelled and UMIs are not included. Apart from RNA sequencing, a widely used tool for single-cell gene expression analysis is quantitative PCR (qPCR), which often is more sensitive, more accessible and with a simpler data analysis workflow compared to RNA sequencing [81, 82]. However, qPCR is limited to the number of genes that can easily be analyzed and the targets needs to be decided beforehand.

Both RNA sequencing and qPCR normally relies on similar initial protocol steps, including single-cell isolation, cell lysis, reverse transcription and preamplification (Figure 5), but the different steps can be performed in various ways [76, 82]. Analysis using qPCR can be performed without a preamplification step if few genes, intermediately or highly expressed, are to be quantified [83].
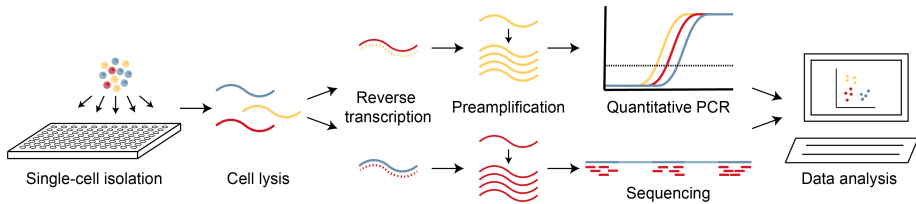


Figure 5. Single-cell gene expression analysis workflow. An overview exemplifying the different experimental steps needed when performing qPCR or RNA sequencing on single cells.

## Data analysis

Analysis of single-cell qPCR data can be performed with qPCR analysis tools established for bulk analysis, with some small but important modifications. The two main differences are data normalization and handling of missing data. Traditional qPCR generally relies on reference genes for normalization but for single cells, due to the stochastic gene expression, no transcripts can be considered stably expressed across samples. Instead, data is usually analyzed as expression per cell [82]. Missing data in bulk analysis normally depends on technical errors whereas for single-cell analysis, missing data is often due to expression values below the limit of detection. For downstream analysis, missing values can be handled either  by setting a Cq (cycle of quantification) value higher than the highest value among the other cells, or by impute a value based on expression of the other analyzed genes [84]. For single-cell RNA sequencing, the amount of data received is much larger leading to challenges in data analysis and new analysis tools are continuously emerging [85]. There are several ways to normalize single-cell RNA sequencing data that

will take different sources of bias into account, and what method to use depends on biological and technical aspects [86]. Methods developed for bulk RNA sequencing data that normalizes the samples according to variations in sequencing depth is commonly used. One example is to normalize the raw count for each transcript to the total number of transcripts in that sample (reads per million, RPM). Still, these methods might not be optimal for single-cell analyses, due to the variation between single cells [85, 86]. One aspect is that this type of normalization will not assess gene expression per cell but will rather report gene expression of each gene in relation to the total amount of RNA in each sample. In cases where the total amount of RNA varies between samples, normalization per cell can be possible using spike-in RNA sequences that are added to each cell lysate before downstream processing. However, challenges with using spike-ins remain, for example due to technical bias and difficulties in protocol optimization [86]. Compared to targeted approaches, global single-cell RNA sequencing data will contain expression values for tens of thousands of genes, including a lot of missing data and low counts that will not be informative and may add noise to downstream analysis. In these cases, data is normally filtered to reduce the gene number and complexity. One way is to select genes that have high variation relative to the mean between the different samples, in order to only include genes in the analysis that are important for explaining the differences between cells [87]. For all kinds of single-cell gene expression data as well as for bulk data, downstream analysis normally includes a dimensionality reduction technique to allow visualization of the multi-dimensional data, including principal component analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE). For single-cell analysis, there is also commonly an interest to group cells based on similarities in gene expression profiles, performed with various kinds of clustering techniques [85]. These identified groups can be further defined by identifying genes that are regulated between them, using differential expression analysis [88]. Additionally, pseudo-time analysis, which orders cells along a trajectory based on their gene expression profiles, is often used to study differentiation paths or cell cycle progression, among other things [89].

# AIM

The overall aim of this thesis was to study tumor cell heterogeneity using single-cell gene expression analysis. By combining functional cellular assays with single-cell gene expression profiling we are able to match cellular phenotype with specific molecular profiles. Individual tumor cells differ from each other due to many different factors, such as the genetic and epigenetic profiles, the cellular state, stochastic variability and interactions with the microenvironment. Some subpopulations of tumor cells, such as cancer stem cells (CSCs), have important roles in tumor progression, metastasis and therapy resistance. Studying tumor cell heterogeneity and characterizing different subpopulations will therefore be important in order to find new biomarkers and targets for future therapies.

The specific aims were:

**Paper I:** To develop a global single-cell preamplification protocol that can be applied to both targeted qPCR analysis and global sequencing approaches.

**Paper II:** To develop a simple and versatile method that can be used to quantify the total amount of mRNA in a single cell.

**Paper III:** To determine gene expression changes during cell cycle progression using single-cell analysis.

**Paper IV:** To identify and characterize myxoid liposarcoma cells with CSC properties using several functional assays and gene expression analysis.

**Paper V:** To identify and characterize ER-negative breast cancer cells with CSC properties by combining functional cellular enrichment assays with single-cell RNA sequencing.

**Paper VI:** To determine the microenvironmental influence on breast cancer cells using an *in vivo*-like culture system based on cell-free patient-derived scaffolds.

# RESULTS AND DISCUSSION

This thesis is based on data from six papers where two covers development and refinement of methods used to profile tumor cell heterogeneity at single-cell gene expression level. In the remaining four papers, combinations of functional assays and single-cell gene expression analyses were used to study different aspects of tumor cell heterogeneity, including differences due to cell cycle phase, characterization of CSCs and effect of the microenvironment, in myxoid liposarcoma and breast cancer.

## GENE EXPRESSION WORKFLOW DEVELOPMENT

To gain more information about the cellular profile of individual cells, we examined the possibility of combining several types of gene expression analyses in a flexible manner. Single-cell gene expression analysis protocols normally include a preamplification step in order to receive enough transcripts to be detected with downstream analysis. Quantitative PCR analysis normally use target-specific preamplification protocols [90-92] that amplifies chosen targets of interest, generally using pools of specific primers. In contrast, RNA sequencing relies on global protocols [80, 93, 94] that theoretically amplifies all transcripts in your sample. In paper I, the aim was to develop a global preamplification strategy that can be used to profile single cells using both a global sequencing technique as well as a targeted qPCR approach. This would allow for a flexible workflow where samples can initially be screened by analysis of limited number of genes using qPCR after which cells of interest can be further analyzed with either additional qPCR or RNA sequencing. To examine this, we compared different reverse transcription and preamplification protocols for gene expression profiling using qPCR. Here, a global approach that generates, amplifies and analyzes full-length cDNA established from the Smart-seq2 protocol [80] was compared to a target-specific approach based on the targeted analysis of cDNA preamplified using multiplex PCR [91] (Figure 6). The two strategies were evaluated using high-throughput qPCR. The reverse transcription of the Smart-seq2 protocol is initiated by an oligodT primer that hybridizes to polyadenylated RNA at the 3'-end. At the 5'-end the reverse transcriptase switch template molecule using a template-switching oligo (TSO). With both primers containing an identical adapter sequence this generates full-length cDNA molecules that can be amplified using a single primer (Figure 6). The targeted approach instead uses a combination of oligodT primers and random hexamers during reverse transcription, followed by preamplification using multiplex PCR with a pool of primers targeting a predefined panel of genes (Figure 6).
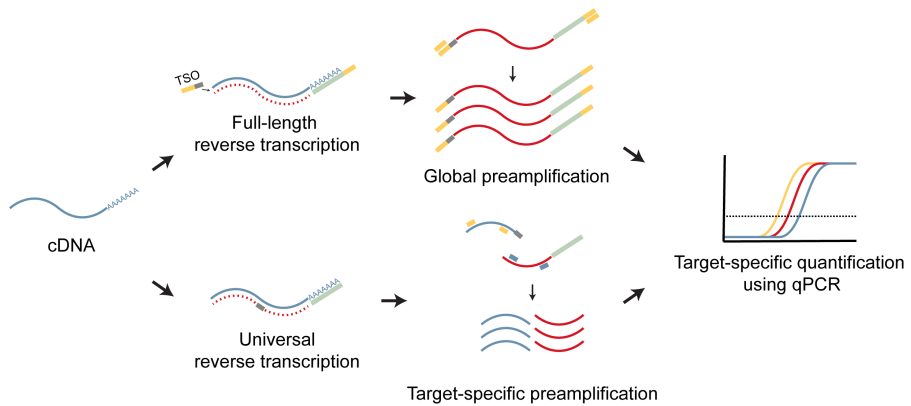
Figure 6. Preamplification protocols. Protocols for the global and target-specific preamplification strategies used in paper I. The global protocol attaches an adapter-sequence on both ends of the cDNA molecule during reverse transcription and amplification is performed using a primer directed against this sequence. The target-specific protocol uses a mix of oligodT and random hexamers in reverse transcription and amplification is performed with a pool of primers directed against specific genes. Adapted from paper I.

We compared yield and reproducibility between the two methods based on expression analysis of 96 genes using high throughput qPCR. Target-specific preamplification generated both higher yield and reproducibility than global preamplification. However, yield and reproducibility are related as the variation will be increased for low expression values [84] which is also shown in our data. This indicates that the decreased reproducibility for global preamplification was, at least to some extent, due to the lower yield. The lower yield seen in the global protocol is largely explained by only using oligodT primers, while the target-specific protocol uses a mix of oligodTs and random hexamers, which generates comparably larger yield [95]. Furthermore, global preamplification relies on incorporation of the adapter sequence on both ends of the cDNA transcript, which is not needed for target-specific preamplification. This incorporation may be incomplete due to inefficiencies in the template-switching mechanism. Finally, preamplification efficiency is also likely decreased due to the longer full-length transcripts. Another drawback of the full-length protocol, not evaluated in this paper, is that intact polyadenylated RNA is required, whereas the target-specific protocol is compatible with samples of poorer integrity as well as non-polyadenylated RNA. A number of these limitations may potentially be overcome by using a global preamplification protocol that relies on an alternative priming strategy [96-99]. Some of the drawbacks with the target-specific preamplification protocol is that it requires time to develop, optimize and validate predefined gene panels. In contrast, the global preamplification protocol is the same every time regardless of the genes of interest. Hence, the global strategy is more versatile, since any polyadenylated transcript can be quantified after the initial experiment, either by qPCR

or RNA sequencing. Our conclusions regarding the advantages of the two methods are summarized in Figure 7.



Figure 7. Advantageous properties of global and target-specific preamplification, respectively, in combination with targeted qPCR analysis.

Finally, to assess if global preamplification may be used for gene expression profiling with qPCR on small sample sizes, single-cell analysis was performed. This experiment demonstrated that the number of cells expressing a certain gene correlated with the average expression value of that gene. Furthermore, the biological variability observed between single cells was considerably larger than explained by technical noise, indicating that this method is able to distinguish the heterogeneity between single cells. Additionally, when simulating the reduced yield resulting from global preamplification on a previously published single-cell dataset generated using the target-specific strategy, the same subpopulations could still be detected when performing principal component analysis. In summary, we concluded that our global preamplification protocol based on the Smart-seq2 protocol offers a simple and flexible workflow that can successfully be used for targeted mRNA quantification of even limited sample sizes.

It has been shown that the total amount of RNA varies depending on several biological factors such as cell type [94], cell size [100], cell cycle state [101, 102] and aging [103]. Studies have also shown that specific proteins can affect the transcriptome globally, for example MYC (also known as c-myc) in both tumor cells and other cell types [104, 105], and MECP2 in neurons [106]. The aim of paper II was to develop a simple method to quantify the total amount of mRNA in single cells enabling studies of variations in RNA amount that may be related to tumor cell properties. Apart from the potential biological implications, the heterogeneity in RNA amount is also important to consider when analyzing gene expression data as many normalization methods assume equal amounts of RNA in each sample. One indirect method to handle this is to add known concentrations of spiked-in RNA standards to each cell, such as the ones developed by the External RNA Controls Consortium (ERCC) [107]. This can be used to estimate the relative amount of RNA in each single cell by comparing the number of sequencing reads of the spike-in standard in relation to number of reads per cell. However, there are challenges and uncertainties related to the use of ERCC spike-ins, for example due to technical bias and difficulties in protocol optimization [86]. To develop an easy and flexible method to measure the total amount of polyadenylated RNA, we applied the Smart-seq2 protocol used in paper I. The preamplification part of this protocol will, in theory, amplify all polyadenylated transcripts in each sample and by adding SYBR green I to this

amplification reaction, we were able to quantify the amount of transcripts using qPCR (Figure 8). Melting curve analysis was included as quality control. Standard curve analysis of diluted RNA and sorted cells from cancer cell lines indicated a wide dynamic range, high reproducibility and high sensitivity of the method. These experiments, together with single-cell analysis, demonstrated that we can reliably quantify the total amount of polyadenylated RNA down to single cells and distinguish it from the background noise generated in cell-free controls. Single-cell analysis further showed a four-fold difference in total amount of polyadenylated RNA between single cells from the same proliferating cell line. This variation can, at least to some extent, be explained by variability in cell cycle phase. The expression of polyadenylated RNA in different individual cells followed a log-normal distribution, which is in line with what is normally the case for individual transcripts [84, 108]. Since this protocol solely relies on oligodT priming it will only target polyadenylated transcripts, which will fail to reverse transcribe histone transcripts and many non-coding RNAs that lack polyA tails [109]. However, protocols that rely on other types of priming strategies could potentially be used for assessment of the total RNA amount including that of non-polyadenylated RNA [96-98]. In conclusion, we have developed a method that can be used to measure the total amount of polyadenylated RNA in single cells which may for instance be used to investigate the differences between cell types, cell states or the effect of various external stimuli. This method can also be integrated into our gene expression analysis workflow to be combined with targeted qPCR or RNA sequencing analysis (Figure 8).
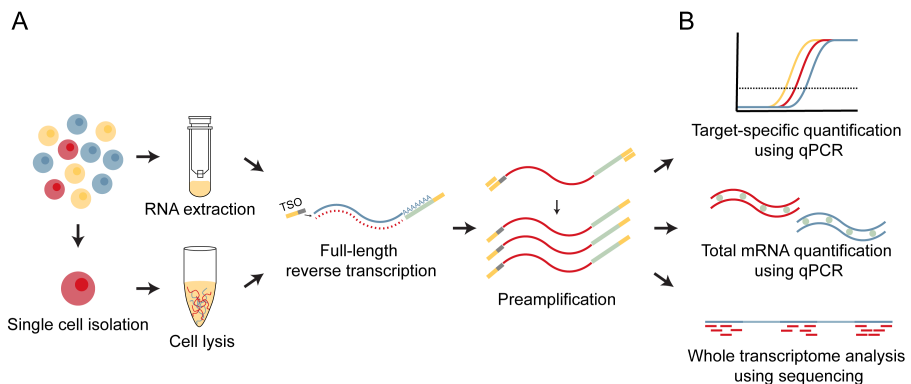


Figure 8. Gene expression analysis workflow. (A) Sample preparation followed by reverse transcription and preamplification according to a full-length global protocol. (B) Different applications based on the same initial protocol that can be used by themselves or in combination. Adapted from paper II.

# EFFECT OF CELL CYCLE PHASE ON GENE EXPRESSION

Cell proliferation is a key part of cancer progression and is therefore of interest when studying cancer cells. Cell cycle deregulation is commonly observed in cancer and

important mediators thereof are proposed as therapeutic targets [110]. As cells pass through the cell cycle, their gene expression profiles will be altered. Therefore, the aim in paper III was to determine the gene expression profiles at different phases of the cell cycle, G1, S and G2/M, using single-cell analysis of an MLS cell line. Myxoid liposarcoma is a slow-growing tumor and we have previously reported that the fusion oncoprotein FUS-DDIT3 is associated with decreased cell proliferation [111, 112]. As it is not clear at what regulatory level FUS-DDIT3 is affecting cell proliferation, we wanted to study if and how the cell cycle in MLS cells was deregulated on the gene expression level. We sorted single MLS cells from different phases of the cell cycle using a DNA-binding dye (Figure 9A) and analyzed the cells with RNA sequencing. Using a pseudo-time line, we ordered the cells within each phase based on their gene expression profile (Figure 9B).
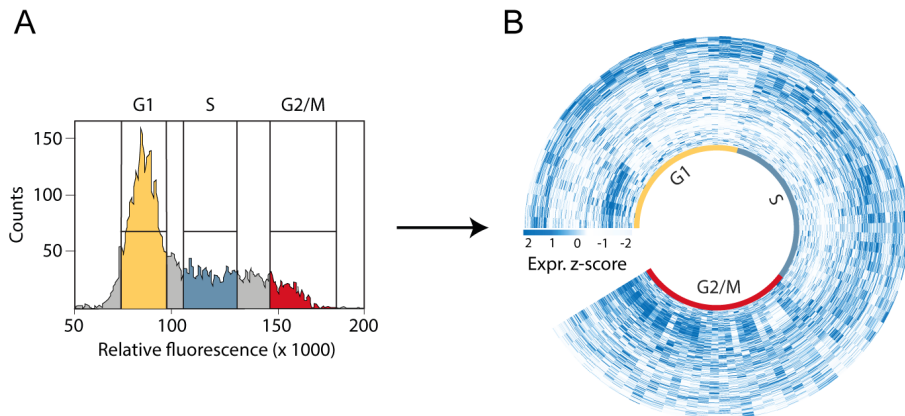


Figure 9. Gene expression variation over the cell cycle. (A) Single cells were collected from different phases of the cell cycle using fluorescence-activated cell sorting based on the intensity of a DNA-binding dye. (B) Normalized expression of 472 genes differentially expressed between any of the collected groups. Each row in the circle represents one gene and each column, perpendicular to the circle, represents one cell. The cells are ordered according to a pseudo-timeline with the cell-cycle phase indicated. Adapted from paper III.

Visualization of the cells using the dimensionality reduction technique t-SNE demonstrated a gradual difference in gene expression from each phase to the next which overlapped with the pseudo-time ordering. Furthermore, the cells were ordered in a circle where the late G2/M cells were similar to the early G1 cells, which is logical since proliferating cells were collected. Of the genes identified as cell-cycle regulated in our study, many could be confirmed to have associations to the cell cycle by comparing them to reported annotations in Gene Ontology [113] or Reactome [114] databases. We could also further validate several gene expression patterns using Cyclebase 3.0 [115]. Additionally, we were able to identify a number of genes differentially expressed between the phases that have not previously been described as cell cycle-regulated. For several of these, we could detect a similar expression pattern in a public dataset of human embryonic

stem cells [116]. The regulated genes could be divided into groups with distinct expression patterns between the different cell cycle phases. These groups could, in turn, be associated with specific gene ontology terms as well as transcription factor- and microRNA binding sites. This provided information about possible mechanisms active in different parts of the cell cycle that can be used as a foundation for future research. Our experimental approach also allowed us to detect two distinct groups of cells in the G1 phase of the cell cycle (Figure 9B). Here, we could identify a number of genes that differed in expression between these two groups of G1 cells that were not otherwise cell-cycle regulated. We propose that these groups are connected to early and late G1 separated by a known restriction point. At this G1 restriction point, cell division is controlled and cells that pass this point are commited to complete the cell cycle. Cells can also exit the cell cycle and enter a quiescent state, G0 [117].

Further analyses of the single-cell gene expression data combined with previously published data on transcript half-life [118] showed that genes with long half-life had a lower cell-to-cell variation compared to more unstable ones. This is expected when considering the notion that genes are transcribed in bursts [119]. The expression of short-lived mRNA transcripts will decrease more between the bursts and thereby induce a larger variation in expression over time which can be translated to a larger variation between different single cells. As discussed in paper II, the total amount of RNA is known to vary depending on several biological factors and, by normalizing the reads of all genes to those of ERCC spike-ins added to each sample, we could observe that the total amount of polyadenylated RNA is increasing throughout the cell cycle from G1 to G2/M, in accordance with the increase in cell size [101, 120]. This result was confirmed by quantification of the preamplified cDNA of each cell.

In MLS cells, we have previously demonstrated with single-cell analysis that FUS-DDIT3 expression is associated with lower cell proliferation [111]. This effect has also been observed when expressing FUS-DDIT3 in cell lines from other tumor types [112]. Still, this study indicates that the cell cycle in proliferating MLS cells is transcriptionally intact as most gene expression patterns could be validated with public data. It has previously been reported that MLS tumors display an abnormal expression pattern of specific cell-cycle controlling proteins indicating a cell cycle deregulation in these tumors [121]. Moreover, it has been shown that MLS tumors contain high proportion of senescent cells, which may explain the slow-growing phenotype [122]. Collectively, this may imply that FUS-DDIT3 controls whether the cells should enter the cell cycle or a senescent state. The selection process occurring during cell culture does not benefit senescent cells in MLS cell lines compared to the situation in tumors, which is supported by increased senescence properties not being detected in MLS cell lines [112]. Therefore, a more comprehensive analysis of primary tumor cells will be needed to determine the effect of FUS-DDIT3 on cell proliferation, for example by single-cell analysis.

In summary, we have identified cell-cycle regulated genes whose expression patterns had previously not been reported as associated with the cell cycle. We have also discovered genes with distinct differences in early and late G1 phase. To reveal the importance of these genes, further mechanistic studies are needed. These studies would also help understand if the genes are related to cancer progression or not. Additionally, we have reported that proliferating MLS cells appear to have a normal cell cycle gene expression profile.

## IDENTIFYING CANCER STEM CELLS IN MYXOID LIPOSARCOMA

The two main models describing tumor cell heterogeneity are the clonal evolution theory and the CSC model. While a combination of both models probably may be used to describe most tumor types [39, 123], the contribution of each model most likely varies between different tumor entities. Myxoid liposarcoma is genetically stable, hence it is reasonable to believe that there are not so many different mutational clones present and instead, the CSC model may be more influential. Additionally, studies have demonstrated that expression of FUS-DDIT3 in mesenchymal stem cells or progenitor cells lead to formation of tumors in mice resembling myxoid liposarcoma. Therefore, MLS is thought to originate from such an undifferentiated cell type [124, 125]. To study the heterogeneity among MLS cells and to determine what cellular processes are important for different subpopulations, the aim of paper IV [126] was to investigate the presence of a CSC subpopulation in MLS and further characterize these cells. Cancer stem cells are thought to be important for tumor progression and therapy resistance [6, 39], but their presence has not been confirmed in MLS. In order to identify a CSC population, we applied two functional cellular assays commonly used to enrich for CSC properties; non-adherent sphere formation assay and side population assay using Hoechst staining, on three different MLS cell lines. The results indicated subsets of cells with CSC properties in all three MLS cell lines, but to different degree. These three cell lines harbors the same fusion oncogene, *FUS-DDIT3,* but the genomic breakpoint in the *FUS* gene varies. We have previously demonstrated that the expression level of FUS-DDIT3, both at protein and gene expression level, differs between these three cell lines [111]. The cell line with the highest protein expression of FUS-DDIT3 contained most cells with CSC properties. As FUS-DDIT3 expression is varied between individual cells within the same cell line [111], this could indicate a potential connection between FUS-DDIT3 expression and CSC properties in subpopulations of MLS cells.

To further characterize the MLS CSCs, we wanted to study potentially important pathways in this subpopulation. The JAK-STAT signaling pathway is involved in many cellular processes, such as cell proliferation, differentiation, cell migration and apoptosis [127] and has been shown to be related to CSCs [128-130]. Furthermore, a previous study has linked the expression of IL6, a cytokine known to activate the JAK-STAT pathway,

to FUS-DDIT3 [131]. Here, we studied the potential link between JAK-STAT and CSCs in MLS cells. We could show that subsets of MLS cells expressed most JAK-STAT gene members using single-cell gene expression profiling. The observed variation in JAK-STAT signaling between different cells could also be demonstrated on protein level in MLS tissues where active, phosphorylated STAT3 was detected in minor subpopulations, to varying degrees in different samples. Treating MLS cells with a JAK1/2 inhibitor, which blocked phosphorylation of STAT3, showed no effect on general cell proliferation or migration but reduced the number of spheres formed in non-adherent conditions. This suggests that JAK-STAT signaling affects the CSCs but not the proliferating cells. Treatment of MLS cells with chemotherapy, such as doxorubicin, instead target proliferating cells. We could show that doxorubicin treatment increased the number of cells that can form spheres, indicating that the cell population surviving treatment was CSC-enriched. Combining the chemotherapy treatment with the JAK1/2 inhibitor prevented this sphere formation increase suggesting that a combination treatment could serve as an improved therapeutic strategy in MLS, targeting both proliferating and more CSC-like cells (Figure 10). Gene expression data from MLS cells treated with doxorubicin confirmed the association between chemotherapy resistance and JAK-STAT signaling since genes upregulated in treatment-resistant cells showed an overlap with JAK-STAT signaling gene sets. However, other pathways were also enriched suggesting that processes other than only JAK-STAT signaling are involved in regulating the CSCs. This was expected since we observed experimentally that some cells with CSC properties survived JAK1/2 inhibition.
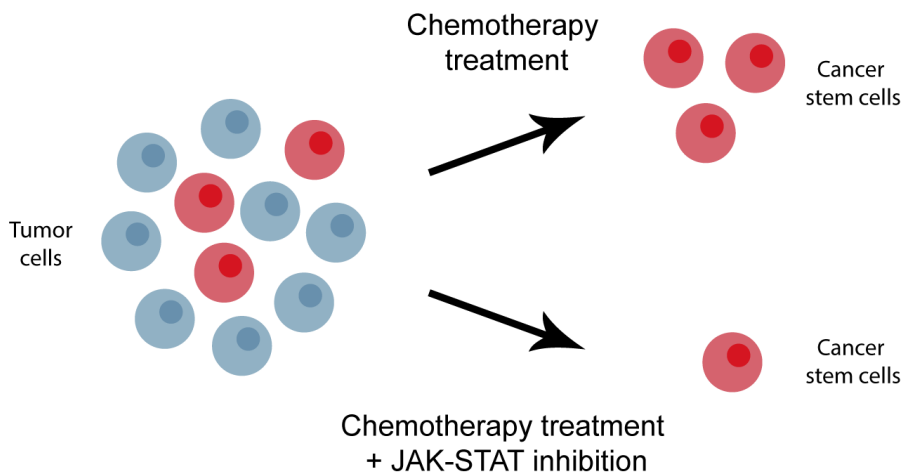


Figure 10. Proposed combination treatment mechanism in MLS. Chemotherapy treatment kills the proliferating cells but the cancer stem cells survive. Combining chemotherapy with JAK-STAT inhibition would also potentially reduce the CSC population.

JAK-STAT signaling also has a role in epigenetic changes [132]. For example, activated STAT3 regulates the expression of methyl transferase DNMT1 [133, 134]. Furthermore, epigenetic regulation is important for maintaining pluripotency and regulate differentiation in embryonic stem cells and has been suggested to have an important role also for CSCs [135, 136]. We wanted to investigate the connection between JAK-STAT signaling and epigenetic regulators in our cells. We could, by immunoprecipitation, show that phosphorylated STAT3 binds to SMARCA4, one of the components of the SWI/SNF chromatin remodeling complex, in all three MLS cell lines. Knock-down of SMARCA4 lead to decreased protein expression of phosphorylated STAT3 as well as decreased sphere formation, suggesting a connection between the epigenetic program and the CSC population in MLS. Additionally, in a separate publication we demonstrated that the majority of FUS-DDIT3 interacts with the SWI/SNF complex, which also was the case for the related fusion oncoprotein EWSR1-FLI1 that is specific for Ewing sarcoma [137]. We could further show that expression of the fusion oncoproteins led to shifts in specific histone marks which could be a functional effect of this binding. The interaction of the fusion oncoproteins to SWI/SNF may be important for the tumor progression in these tumors. Another study has shown a high proportion of superenhancers, clusters of enhancers associated with certain active histone marks, among FUS-DDIT3 target sequences. This further indicates a role for the fusion oncoprotein in generating an aberrant epigenetic landscape in these tumors [138].

In summary, we identified cells with cancer stem cell properties in MLS cell lines and could connect those properties to an active JAK-STAT pathway, something that can be used for developing an improved therapeutic strategy.

## CHARACTERIZATION OF BREAST CANCER STEM CELLS

Breast cancer is a heterogeneous disease with several subtypes. Triple-negative tumors lack specific treatment options and are often aggressive [16], making it valuable with new treatment strategies for these tumors. The presence of CSCs in breast cancer has already been verified [139-141] and characterizing these cells is important both for understanding more about the processes during tumor progression and recurrence, but also to develop novel treatments. Existing markers used to select for breast CSCs include surface marker signature CD44$^+$/CD24$^{-/low}$ [140] and ALDH1 [142]. However, their expression varies depending on breast cancer subtype and the expressions of the different markers do not correlate [143, 144]. There also seems to be subpopulations of cells with CSC properties not defined by these markers [145] and studies investigating their connection to clinical outcome have given variable results [146]. Several pathways have been associated to breast CSCs, including the Wnt [147], Notch [148] and Hedgehog [149] pathways related to self-renewal, the PI3K-mTOR pathway [150] and the JAK2-STAT3 pathway [128]. Several of these pathways have been connected to CSCs in additional tumor types [151]. For

example, we linked the JAK-STAT pathway to MLS CSCs in paper IV. Multiple connections between CSC-related pathways in breast cancer have been reported [152, 153], indicating a complex system for CSC maintenance. Although some attempts of targeting the CSC population in breast cancer have yielded promising initial results [154-158], no therapy is used in the clinic. Altogether, there is a need for further characterization of breast CSCs in order to find improved markers, for example through single-cell analysis. Previous studies have aimed to characterize CSCs by single-cell gene expression analysis without relying on markers to select cell populations, by studying migratory subpopulations [159] or cells able to form spheres [160]. Due to limited amount of studies with varying results, more data and functional validation can give a more complete picture of breast CSC properties.

In paper V, the aim was to further define the CSC population in ER-negative breast cancer by using our single-cell gene expression analysis workflow in combination with functional cellular assays that enables detailed CSC characterization (Figure 11). We used the ER-negative breast cancer cell line MDA-MB-231 and combined the non-adherent mammosphere formation assay with a label-retention assay to enrich for cells with different self-renewing properties and at high or low proliferation rates. This combination has previously been used to isolate human normal mammary stem cells [47]. As control, we included cells grown under normal adherent cell culture conditions. As demonstrated in paper III, the cell cycle phase greatly affects the gene expression profile and we therefore collected the control cells only from the G1 cell cycle phase to minimize cell cycle bias.
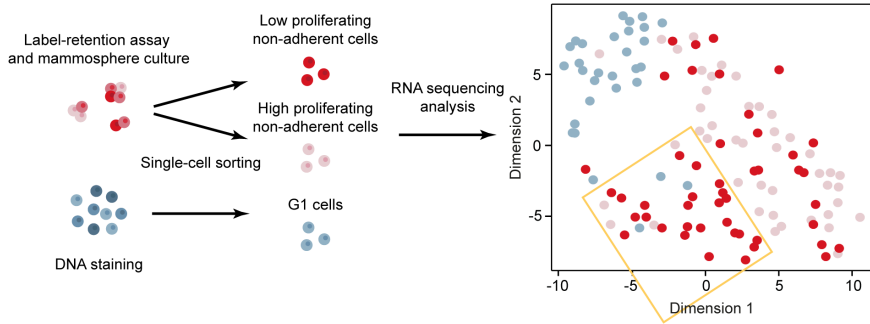


Figure 11. Functional cellular assays combined with single-cell gene expression analysis. Cells were collected from three different subpopulations with various cellular properties based on staining and culture methods. After RNA sequencing, cells were separated based on their gene expression profile into a t-distributed Stochastic Neighbor Embedding (t-SNE) plot. Cells are colored according to their group as defined in the figure. The yellow area indicates one of the clusters identified with *k*-means clustering. This cluster proposedly included cells with CSC properties and was therefore further characterized. Adapted from paper V.

After single-cell RNA sequencing, we clustered the cells into groups based on their gene expression profile independently of sample collection method. The clusters overlapped to large extent with the functional cellular enrichments but not completely. Even though the functional methods used enrich for certain characteristics, they are not completely limited to one specific property. The label retention intensity will be affected by cell size [161] and the mammosphere assay have been shown to select for a larger population than only the CSCs [162], indicating an influence by other properties. Still, gene expression profiling is not enough for a complete phenotypic assessment and thereby a combination of the two may improve identification of cellular subpopulations. As expected, we identified proliferation as an important factor separating the cells. We argued that the cluster of cells enriched for low-proliferating mammosphere-cultured cells (indicated in Figure 11) contained cells with most CSC properties and we further studied this subgroup. We could demonstrate, by normalizing the entire set of genes to spiked-in synthetic RNA, that the total mRNA expression was reduced in this cluster in line with earlier studies showing reduced amount of total RNA in stem cells [163, 164] as well as in quiescent cells [165]. We further identified 14 genes upregulated in the CSC-like cluster, which could be potential biomarkers for the CSC population in ER-negative breast cancer. Indeed, several of them have previously been connected to CSC properties in breast cancer. Three of these genes; *LGALS3*, *MYH9* and *DSTN,* could further be connected to worse relapse-free survival in ER-negative breast cancer, evaluated using a publicly available data base [166], thereby being of extra interest for future studies.

In conclusion, we have through the combined use of functional cellular assays and single-cell gene expression profiling identified potential biomarkers related to CSCs in ER-negative breast cancer. However, to further elucidate the role and potential clinical use of these markers, validation experiments need to be performed. Such experiments could include knock-down or overexpression experiments followed by functional assays using cell lines, similarly to a study of the non-coding RNA *NEAT1* in breast cancer cell lines [167]. Additionally, other more *in vivo*-like model system may yield more relevant information as the microenvironment has an important role for CSC functions [168].

# MICROENVIRONMENTAL EFFECT ON BREAST CANCER CELLS

The microenvironment includes stromal cells, vasculature, extracellular matrix (ECM) and signaling molecules, which greatly affects the properties of tumor cells. In paper VI, we aimed to determine the effect of different tumor microenvironments on cancer cells. We therefore developed a 3D culture system based on tumors that are washed repeatedly with mild detergents to remove all cells and thus generating so called patient-derived scaffolds (PDSs). These PDSs can thereafter be repopulated with cells of interest, such as cancer cell lines (Figure 12). Despite being an *ex vivo* system, PDSs maintain critical properties of the original tumor-specific microenvironments, which can be used to reveal important patient-specific information *in vitro*.
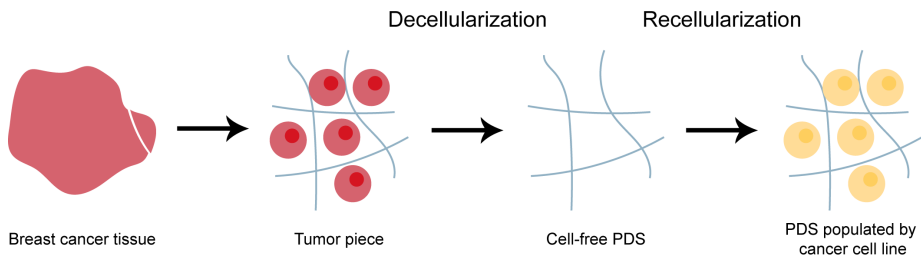


Figure 12. Generation of patient-derived scaffolds (PDSs). A fresh or frozen breast cancer tissue piece is washed with detergents to remove the cells and create PDSs. Cancer cell lines, or other cells of interest, can then be used to repopulate the PDSs.

Scanning electron microscopy analysis of decellularized PDSs, before recellularization, revealed structural differences between individual PDSs but also, in some cases, heterogeneity between specific surfaces of the same PDS. Intertumor heterogeneity was also evaluated using mass spectrometry, which demonstrated the variations in protein composition between 15 decellularized PDSs with different clinical properties. Self-organizing map (SOM) analysis based on the expressed proteins was used to divide the PDSs into two groups, which could be related to grade and expression of proliferation marker Ki67 assessed from the original tumors. This demonstrates that analysis of the protein composition of PDSs alone can potentially provide valuable information that can be linked to clinical information. Additionally, an interaction network was constructed based on the proteins detected in PDSs and their first-order interaction partners, using information from the Human Protein Reference Database [169]. Several PDS proteins with central roles within this network could be connected to the cellular process of exosome secretion, indicating that secretion plays an important role in the microenvironment of PDSs.

Repopulation of PDSs with two breast cancer cell lines, the ER-negative MDA-MB-231 and the ER-positive MCF7, demonstrated infiltrative growth of cells, but also morphological differences between cells within the same cell line. Analyzing the cells

cultured in PDSs compared to normal 2D cultures as a reference point, both at mRNA and protein level, showed increased expression of markers related to pluripotency and EMT whereas markers connected to proliferation and differentiation were decreased as a result of growth in the 3D system. Together with increased mammosphere formation as well as increased tumor growth in immunodeficient mice, these results indicated a CSC-enriched population of cells cultured in PDSs, which has previously been shown in other 3D growth models [170-172]. Furthermore, single-cell gene expression analysis of MCF7 cells cultured in two different PDSs, in comparison to conventional 2D conditions, showed that PDS-cultured cells had become more heterogeneous. The PDS-cultured cells were separated into two subgroups, each including cells from both PDSs, of which one was defined by high expression of differentiation markers whereas the other by high expression of pluripotency and EMT markers. In order to compare the effect of the PDS to *in vivo* conditions, we performed RNA sequencing on cells grown in PDSs compared to both 2D and xenograft-derived cells. Principle component analysis indicated that PDS cultures were more similar to xenograft cultures in relation to 2D growth. Furthermore, among the pathways associated to the upregulated genes in PDS or xenograft, respectively, in relation to 2D, several overlapped including extracellular matrix organization and collagen formation. Collectively, this points towards an adaptation of the cells to a 3D growth environment. Next, we showed that the majority of the significantly enriched pathways associated with the upregulated genes in PDS-cultured cells compared to 2D cells overlapped with pathways which were found based on the mass spectrometry analyses, suggesting a connection between protein composition of PDSs and the induced gene expression profiles of cells cultured within. Additionally, several of the regulated genes were included in the previously constructed protein interaction network further indicating such a connection.

Finally, we could show that high induced expression of EMT-markers *VIM* and *SNAI2* in PDS-cultured cells could be linked to worse relapse-free survival of the patients the tumors were derived from. The VIM protein was also included as one of the central proteins in the network constructed from scaffold proteins, serving as a link between important pathways in the PDSs, induced gene expression changes in cells grown in PDSs and to clinical effects in the patients the PDSs were derived from. This clinical link supports that prognostic information indeed can be extracted from the cell-free tumor microenvironment.

In contrast to many other 3D culture systems, PDSs are unique for each patient, and analyzing cells from homogenous cell lines in heterogeneous PDSs makes it possible to study the effect of various tumor microenvironments. Culturing cells in different PDSs will generate changes in the cellular properties specific to each PDS (Figure 13). Indeed, we could demonstrate that PDSs from different patients affected the cultured cancer cells in dissimilar ways. This makes the PDS model system a potentially valuable tool as a drug screening platform. Drugs can be tested in different PDSs to connect the drug response

to specific breast cancer subtypes, or even patients. Additionally, by studying the PDSs, important biological information about how the environment affect tumor cells can be determined that can lead to the identification of novel drug targets. The PDS system can also be used to study the effect of different microenvironments on other aspects of tumor cell heterogeneity, such as cell states and epigenetic changes. For instance, the PDS system could be an interesting platform for studying cancer stem cells, including validation experiments of the markers identified in paper V. In fact, when analyzing the RNA sequencing data comparing PDS-cultured MDA-MB-231 cells to 2D-cultured cells, 13 of the 14 markers from paper V showed increased expression in the PDS group, of which 7 were significantly two-fold upregulated. Additionally, several of the genes were also significantly upregulated in MCF7 cells cultured in PDSs. This further point towards the fact that the PDS system enriches for CSC properties.
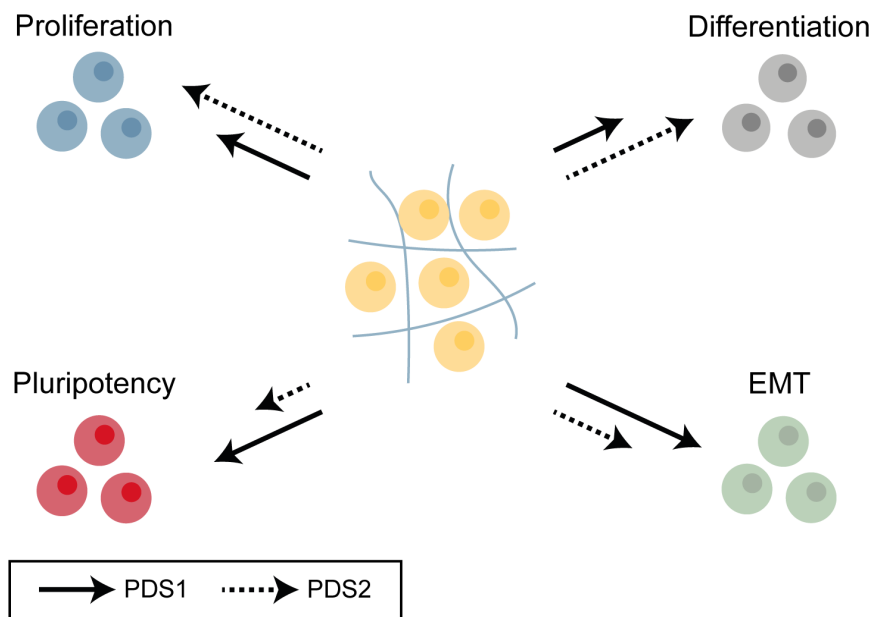


Figure 13. The Patient-derived scaffolds determines the cellular phenotype. When culturing cell lines in different patient-derived scaffolds (PDSs), subpopulations of cells will be formed with various degree of cellular properties such as proliferation, differentiation, pluripotency and EMT. For each PDS, the number of cells within each subpopulation will be altered.

The PDS system, though developed for breast cancer, is not limited to one tumor type but can be evaluated for different solid cancer forms, for example MLS. As MLS is a rare cancer form, there is limited access to patient material as well as relevant model systems. The PDS system could therefore be a welcomed addition to the possibilities of studying MLS cells, as well as similar tumor types. Even though the challenge of accessing tumor

material remains, one tumor can be used to create many PDS pieces and can therefore be used for numerous analyses. Hence, by using several pieces from the same PDS, the microenvironment can be held more constant, while instead comparing different cell lines, treatments or other induced variations among the cells. However, the possibility of comparing many varying tumor microenvironments and relate the induced effect in cell lines to clinical outcome, which we could do in breast cancer with established biobanks and data records, is likely not possible in such a rare tumor type as MLS.

In summary, we have developed a novel culture system based on decellularized PDSs. We could demonstrate that this 3D culture system induces increased heterogeneity, more similar to what is observed *in vivo*, in standardized cancer cell lines. Additionally, molecular properties of the PDSs, as well as of cells cultured in PDSs, could be linked to clinical parameters of the original tumors used to create the PDSs. Further studies could potentially demonstrate the use of this culture system for prognostic or drug screening purposes.

# CONCLUSIONS

We have developed methods for single-cell gene expression analysis. We have studied gene expression changes in different tumor cell subpopulations in relation to various intrinsic and extrinsic sources of variation. We have been able to identify markers and pathways connected to these subpopulations that potentially can be used as biomarkers and future drug targets. Finally, we have developed a novel *in vivo*-like model system.

**Paper I:** We concluded that global preamplification can be used in combination with targeted qPCR analysis enabling a simpler and more flexible workflow, where gene expression profiling using qPCR and RNA sequencing can be combined in single cells.

**Paper II:** We developed a method to measure the total amount of mRNA in single cells which can easily be combined with other transcriptomic analyses. Using this method, we could observe a four-fold variability in the total amount of mRNA between single cells from a cancer cell line.

**Paper III:** We demonstrated how established and previously unreported cell cycle-regulated genes displayed different expression patterns throughout the cell cycle and we also illustrated a distinct difference between early and late G1 phase. Furthermore, we confirmed that stable mRNAs had a reduced variability between single-cells, in line with the transcriptional burst theory.

**Paper IV:** We concluded that myxoid liposarcoma contain a subpopulation of cells with cancer stem cell (CSC) characteristics. Further, we could show that the JAK-STAT signaling pathway control the number of cells with these properties. Hence, a combination of chemotherapy and JAK-STAT inhibition could potentially target both proliferating cells and cells with CSC features, improving the current treatment of myxoid liposarcoma patients.

**Paper V:** Through a combination of functional assays and gene expression profiling, we identified 14 genes upregulated in a subpopulation of breast cancer cells with CSC properties. Some of these potential breast CSC markers were associated to reported stem cell properties and obtained clinical survival data.

**Paper VI:** We demonstrated that cells grown in patient-derived scaffolds (PDSs) displayed increased CSC properties as well as increased heterogeneity, more closely resembling *in vivo* conditions than cells cultured in normal cell culture conditions. Furthermore, we concluded that growth in PDSs lead to different transcriptomic alterations in cells, which in some cases could be linked to clinical information about the original tumor used to create the PDSs, such as recurrence.

# FUTURE PERSPECTIVES

Tumor cell heterogeneity has significant implications for disease progression and treatment response. As illustrated in Figure 2, there are several sources that promote heterogeneity among tumor cells, such as genetic and epigenetic changes as well as the surrounding microenvironment. Characterizing aspects of this variation is therefore important for developing new treatments against different cancer types. Such differences between individual cells can generally not be deciphered when performing analyses at cell population level and instead, single-cell assays are frequently used.

We have developed and refined methods for single-cell gene expression analysis, thus creating a toolbox which may be used to combine sensitive target-specific quantification, high-throughput mRNA sequencing and total mRNA quantification in the same individual cell (Figure 8). By combining these analyses with functional cellular assays and model systems we are able to study several properties of tumor cell heterogeneity, such as cellular states or the microenvironment, and their effect on gene expression. As single-cell RNA sequencing has become more established, high-throughput measurements of other analytes in individual cells are being developed [173], which are needed to study other aspects of cellular variation. Genetic analysis is especially relevant to include in studies of tumor types with large genetic heterogeneity. One example is breast cancer, where there is mutational variability within, as well as between, different subtypes of the disease [14, 15]. Several methods for single-cell genome sequencing have been reported with varying technical challenges regarding, for example, risk of false positives [174]. For the genetically stable MLS, information about the epigenetic landscape is instead of more interest, due to the interactions we recently reported between the fusion oncoprotein FUS-DDIT3 and the SWI/SNF chromatin remodeling complex [175]. Single-cell analysis could potentially be used to connect certain epigenetic changes to FUS-DDIT3 expression or to specific subpopulations, such as CSCs. Epigenetic profiling is complex and various methods are used to study specific epigenetic aspects. At the single-cell level, examples of assays that have been developed include variations of bisulfite sequencing to detect methylation status and ATAC-seq to identify open chromatin regions [176, 177]. Gene expression profiling is widely used as a phenotypic readout [173], but correlations between gene expression and corresponding protein expression are often weak [178, 179]. This we have also observed regarding FUS-DDIT3 expression in MLS [111]. Hence, protein measurements may be desired together with gene expression profiling. Several single-cell protein analysis methods exist, including classical methods such as immunohistochemistry and flow cytometry that are limited in the number of proteins that can be analyzed at the same time. More recent technologies have emerged with increased throughput although most are still antibody-based with challenges regarding specificity and scalability. This might be overcome by recent developments in single-cell mass spectrometry analysis [180].

Regulation of cellular properties occur at both DNA, RNA and protein level. Hence, to get a more complete picture of cellular heterogeneity, combining analyses of multiple analytes in the same single cell is essential, *i.e.* single-cell multiomics. Previously, we have demonstrated the possibility to combine analysis of DNA, RNA and protein levels in the same single cell using qPCR-based methods [181]. Other methods where gene expression analysis has been combined with protein measurements have been reported [182, 183]. Furthermore, several other single-cell-based multiomic approaches have been developed [184], including genome with transcriptome sequencing [185, 186], methylome with transcriptome sequencing [187, 188], and methylome and transcriptome sequencing together with copy number variant detection [189]. A number of companies, including 10X Genomics, Mission Bio and BD Biosciences, have also started to commercialize products to use for multiomics analyses, such as combined assessments of gene expression and protein levels in the same cell. Altogether, these studies show that there are already possibilities for an even more comprehensive workflow where multiple types of analyses can be combined in the same individual cell depending on the specific research question.

The studies in this thesis have mainly been performed using established cancer cell lines. Cell lines are easy to handle and uncomplicated to manipulate as well as easily expanded to high amounts. In comparison, *in vivo* experiments are more challenging due to limited tissue availability and high complexity. Still, performing studies using tumor cells yields more relevant information and these two approaches are therefore often combined. For instance, we used primary cell analysis in a previous study to validate findings identified using cell lines [190]. In that study, we demonstrated in cell lines that ER-negative and -positive breast cancer share a common subpopulation of cells with CSC properties and subsequently confirmed these findings using primary tumor cells. Similarly, potential markers or pathways identified with cell line screening, like in papers IV and V, can be further studied using primary tumor cells. Conversely, clinical findings can be experimentally tested using *in vitro* experiments, for example by combining functional assays with different types of downstream analyses. Another interesting alternative for tumor heterogeneity would be to analyze the primary tumor cells as well as the PDS from the same tissue, for example to connect genotype, protein composition of the PDS and effect of the ECM on standardized cell lines This could also include analyses of surrounding stromal cells. In breast cancer, the availability of patient material is far greater compared to MLS. However, the large genetic variation between breast cancer subtypes and individual patients lead to difficulties in generalizing findings. As MLS is genetically more homogenous, there is probably less variation between different tumors. It would be interesting to perform single-cell gene analyses on MLS cells to study the large subpopulation of senescent cells and the role of the fusion oncogene *in vivo*.

To further study tumor cell heterogeneity in relation to the microenvironment, spatial analysis can be applied. Here, the molecular analyses are performed while retaining

information about the position of the cell in a tissue [191, 192]. Recently, various spatially resolved transcriptomic assays have been developed. These differ in terms of sensitivity, spatial resolution and the number of transcripts that can be analyzed simultaneously, as well as if they are targeted or non-targeted [191]. The ability to identify the position of specific subpopulations within a tumor in relation to various parts of the microenvironment, such as blood vessels and neighboring cells, would be invaluable to better understand tumor biology and to design new treatment strategies. As an example, spatial differences in breast cancer has been demonstrated, where regions with varying expression of genes related to prognosis and diagnosis as well as different cancer-related processes such as immune signaling, invasiveness and stemness could be identified [193]. Spatial transcriptomics may be especially interesting in the study of CSCs since there is a specific region within the tumor microenvironment in which CSCs resides, referred to as the CSC niche [194, 195]. Furthermore, spatial analysis may be applied on PDS cultures, potentially determining if the subpopulations of the cultured cells that we identified with single-cell gene expression analysis in paper VI can be connected to specific regions of the PDS.

Further characterization of tumor cell heterogeneity is key to tackle patient variation in treatment response. With technological advancements in multiomics allowing larger number of analyses and cells, the amount of data that can be gained is continuously increasing. Together with the information attained about the microenvironment through novel model systems, such as PDSs, and spatial analyses this opens up for identification of future drug targets.

# ACKNOWLEDGEMENT

There are of course many people I want to thank that have, in one way or another, helped me towards finishing this thesis.

First, my supervisor **Anders**. Thank you for believing in me, for being available and helpful in all parts of my PhD time and for building a great group. My co-supervisor **Pierre**, thank you for bringing me to SCC, for your enthusiasm and for always helping me when I have questions. My co-supervisor **Göran**, thank you for collaborations, and your interest in discussing data. My co-supervisor **Erik**, thank you for your help with the bioinformatic parts of my thesis, and for always welcoming me when I ask for assistance.

I want to thank all current and previous members of Ståhlberg group for creating a wonderful work environment. **Soheila**, you are a great colleague and friend. **Daniel**, no matter how busy you are, you are always willing to help. **Lisa**, I really enjoy working with and spending time with you. **Tom**, so positive and helpful, it was great working with you. **Stefan**, **Manuel** and **Parmida**, you are all so nice and easy to work with. **Gustav**, I appreciate your engagement and many interesting discussions. **Helena**, thanks for your great work with the orders, I know how difficult it can be. **Tobias** and **Junrui**, you are always happy to help. **David**, thanks for pleasant company and for sharing your knowledge. Finally I want to thank the students, especially **Salim**, who struggled with me when learning to analyze single-cell RNA sequencing data.

Furthermore, I want to thank all current and previous members of Åman group for excellent team-work and a lot of fun. Especially **Pernilla**, thank you for everything, you are one of the reasons I have enjoyed my time at SCC so much. **Malin**, I am so glad I got you as a collaborator and office neighbor, thanks for all scientific and private conversation and sorry for always disturbing. **Christoffer**, always nice, positive and helpful. **Christer**, thank you for all help in the beginning and for teaching me a lot in the lab. **Tajana**, for such nice company and collaboration. And all others.

Additionally, thanks to all current and previous members of Landberg group for great collaborations and pleasant social activities, including **Anna**, **Paul**, **Pernilla**, **Emma**, **Mamen, Elena**, **Karoline**, **Sara**, **Andreas**, **Simona**, **Ylva**, **Susann, Nina** and **Pauline**. Especially **Anna** and **Paul**, for our work together with the PDS paper.

I also want to thank all other co-authors. Especially **Joakim**, for all the help with RNA sequencing analysis.

Thanks to everyone at Sahlgrenska Cancer Center for collaborations, fikas and nice chats over the years. To everyone at floor 5 for conversations and help with this and that,

including **Ywonne**, **Marta**, **Mattias**, **Maryam**, **Mohamed**, **Emil, Kristell** and **Christin**. To others that I enjoyed numerous coffee breaks with, including **Gülay**, **Yvonne**, **Linda**, **Tobias**, **Karin L**, **Karin W**, **Carina**, **Ulrika** and **Birgitta**. Thanks to others at SCC that I interacted with in different ways. Especially **Elin F** for your friendliness and answering all my questions regarding thesis writing and dissertation, and **Malin N** for all pleasant FACS days. And many more, there are too many to mention all, but you all mattered.

There are also many people outside of work that I want to thank. First, all my friends. Especially **Lina**, **Linnea** and **Sofia**, I am so happy that I have you. And all other friends from Chalmers, particularly **Elin**, **David** and **Leif** for nice lunches and chats about work as well as private matters.

To my families. My parents, **Martin** and **Maggan**, you have always supported me no matter what I wanted to do, I could not have asked for greater parents. My sister and best friend, **Hanna**, I cannot imagine my life without you. My brother, **Viktor**, you are the best. A wonderful thing about meeting a partner is that you get a second family. **Marie**, **Sven-Åke**, **Niclas**, **Stina**, **Henrik** and **Sara**, thank you for support, help and wonderful company.

Finally, my own family. **Christian**, without you this would not have been possible. Thank you for endless pep talks and support, I love you! **Freja**, you are my sunshine and you make my life better every day, I love you forever!

# REFERENCES

[1] Bray F, et al. *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA Cancer Journal for Clinicians, 2018. 68(6):394-424.

[2] Weinberg RA. *The biology of cancer.* 2nd ed. New York: Garland Science; 2014.

[3] Vogelstein B, et al. *Cancer genome landscapes.* Science, 2013. 340(6127):1546-58.

[4] Hanahan D, Weinberg RA. *Hallmarks of cancer: the next generation.* Cell, 2011. 144(5):646-74.

[5] Nikolaou M, et al. *The challenge of drug resistance in cancer treatment: a current overview.* Clinical & Experimental Metastasis, 2018. 35(4):309-18.

[6] Dean M, Fojo T, Bates S. *Tumour stem cells and drug resistance.* Nat Rev Cancer, 2005. 5(4):275-84.

[7] Sorlie T, et al. *Repeated observation of breast tumor subtypes in independent gene expression data sets.* Proc Natl Acad Sci U S A, 2003. 100.

[8] Senkus E, et al. *Primary breast cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up.* Annals of Oncology, 2013. 24(SUPPL.6):vi7-vi23.

[9] Ståhlberg A, et al. *Normal and functional TP53 in genetically stable myxoid/round cell liposarcoma.* PLoS ONE, 2014. 9(11).

[10] Hofvander J, et al. *Different patterns of clonal evolution among different sarcoma subtypes followed for up to 25 years.* Nature Communications, 2018. 9(1).

[11] Cancerfonden, Socialstyrelsen. *Cancer i siffror 2018.* 2018.

[12] Miller KD, et al. *Cancer treatment and survivorship statistics, 2016.* CA: a cancer journal for clinicians, 2016. 66(4):271-89.

[13] Polyak K. *Heterogeneity in breast cancer.* Journal of Clinical Investigation, 2011. 121(10):3786-8.

[14] Lehmann BD, 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.

[15] Koboldt DC, et al. *Comprehensive molecular portraits of human breast tumours.* Nature, 2012. 490(7418):61-70.

[16] Badve S, et al. *Basal-like and triple-negative breast cancers: a critical review with an emphasis on the implications for pathologists and oncologists.* Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc, 2011. 24(2):157-67.

[17] Fletcher CDM, et al. *WHO Classification of Tumours of Soft Tissue and Bone.* 4th ed. Lyon: IARC Press; 2013.

[18] Panagopoulos I, et al. *Fusion of the EWS and CHOP genes in myxoid liposarcoma.* Oncogene, 1996. 12(3):489-94.

[19] Crozat A, et al. *Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma.* Nature, 1993. 363(6430):640-4.

[20] Åman P. *Fusion oncogenes in tumor development.* Seminars in Cancer Biology, 2005. 15(3):236-43.

[21] Lindén M, et al. *FET family fusion oncoproteins target the SWI/SNF chromatin remodeling complex.* EMBO Reports, 2019.

[22] Hofvander J, et al. *Different patterns of clonal evolution among different sarcoma subtypes followed for up to 25 years.* Nature communications, 2018. 9(1):3662-.

[23] Moreau LC, et al. *Myxoid\round cell liposarcoma (MRCLS) revisited: An analysis of 418 primarily managed cases.* Annals of Surgical Oncology, 2012. 19(4):1081-8.

[24] Hoffman A, et al. *Localized and metastatic myxoid/round cell liposarcoma: clinical and molecular observations.* Cancer, 2013. 119(10):1868-77.

[25] Kubista M, Dreyer-Lamm J, Ståhlberg A. *The secrets of the cell.* Molecular aspects of medicine, 2018. 59:1-4.

[26] Bedard PL, et al. *Tumour heterogeneity in the clinic.* Nature, 2013. 501(7467):355-64.

[27] Sun X-x, Yu Q. *Intra-tumor heterogeneity of cancer cells and its implications for cancer treatment.* Acta pharmacologica Sinica, 2015. 36(10):1219-27.

[28] Caiado F, Silva-Santos B, Norell H. *Intra-tumour heterogeneity - going beyond genetics.* The FEBS journal, 2016. 283(12):2245-58.

[29] Kadoch C, et al. *Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy.* Nature Genetics, 2013. 45(6):592-601.

[30] Jones PA, Issa JPJ, Baylin S. *Targeting the cancer epigenome for therapy.* Nature Reviews Genetics, 2016. 17(10):630-41.

[31] Whitfield ML, et al. *Identification of genes periodically expressed in the human cell cycle and their expression in tumors.* Mol Biol Cell, 2002. 13(6):1977-2000.

[32] Raj A, et al. *Stochastic mRNA Synthesis in Mammalian Cells.* PLOS Biology, 2006. 4(10):e309.

[33] Gatenby RA, Gillies RJ. *A microenvironmental model of carcinogenesis.* Nature Reviews Cancer, 2008. 8(1):56-61.

[34] Lu P, Weaver VM, Werb Z. *The extracellular matrix: A dynamic niche in cancer progression.* The Journal of Cell Biology, 2012. 196(4):395-406.

[35] Park CC, Bissell MJ, Barcellos-Hoff MH. *The influence of the microenvironment on the malignant phenotype.* Molecular Medicine Today, 2000. 6(8):324-9.

[36] Nowell PC. *The clonal evolution of tumor cell populations.* Science, 1976. 194(4260):23-8.

[37] Reya T, et al. *Stem cells, cancer, and cancer stem cells.* Nature, 2001. 414(6859):105-11.

[38] Kreso A, Dick JE. *Evolution of the cancer stem cell model.* Cell Stem Cell, 2014. 14(3):275-91.

[39] Beck B, Blanpain C. *Unravelling cancer stem cell potential.* Nature Reviews Cancer, 2013. 13(10):727-38.

[40] Lyle S, Moore N. *Quiescent, slow-cycling stem cell populations in cancer: A review of the evidence and discussion of significance.* Journal of Oncology, 2011.

[41] Marjanovic ND, Weinberg RA, Chaffer CL. *Cell Plasticity and Heterogeneity in Cancer.* Clinical Chemistry, 2013. 59(1):168.

[42] Chaffer CL, et al. *Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state.* Proceedings of the National Academy of Sciences, 2011. 108(19):7950.

[43] Mani SA, et al. *The epithelial-mesenchymal transition generates cells with properties of stem cells*. Cell, 2008. 133(4):704-15.

[44] Lapidot T, et al. *A cell initiating human acute myeloid leukaemia after transplantation into SCID mice*. Nature, 1994. 367(6464):645-8.

[45] Al-Hajj M, et al. *Prospective identification of tumorigenic breast cancer cells*. Proc Natl Acad Sci U S A, 2003. 100.

[46] Dontu G, et al. *In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells*. Genes and Development, 2003. 17(10):1253-70.

[47] Pece S, et al. *Biological and Molecular Heterogeneity of Breast Cancers Correlates with Their Cancer Stem Cell Content*. Cell, 2010. 140(1):62-73.

[48] Ho MM, et al. *Side Population in Human Lung Cancer Cell Lines and Tumors Is Enriched with Stem-like Cancer Cells*. Cancer Research, 2007. 67(10):4827.

[49] Richichi C, et al. *Marker-independent Method for Isolating Slow-Dividing Cancer Stem Cells in Human Glioblastoma*. Neoplasia (New York, NY), 2013. 15(7):840-7.

[50] Rainusso N, et al. *Identification and gene expression profiling of tumor-initiating cells isolated from human osteosarcoma cell lines in an orthotopic mouse model*. Cancer Biology and Therapy, 2011. 12(4):278-87.

[51] Patrawala L, et al. *Side population is enriched in tumorigenic, stem-like cancer cells, whereas ABCG2+ and ABCG2- cancer cells are similarly tumorigenic*. Cancer Research, 2005. 65(14):6207-19.

[52] Wu C, Alman BA. *Side population cells in human cancers*. Cancer Letters, 2008. 268(1):1-9.

[53] Quail DF, Joyce JA. *Microenvironmental regulation of tumor progression and metastasis*. Nature Medicine, 2013. 19(11):1423-37.

[54] Grivennikov SI, Greten FR, Karin M. *Immunity, Inflammation, and Cancer*. Cell, 2010. 140(6):883-99.

[55] Wilding JL, Bodmer WF. *Cancer Cell Lines for Drug Discovery and Development*. Cancer Research, 2014. 74(9):2377.

[56] Day C-P, Merlino G, Van Dyke T. *Preclinical mouse cancer models: a maze of opportunities and challenges*. Cell, 2015. 163(1):39-53.

[57] Fang Y, Eglen RM. *Three-Dimensional Cell Cultures in Drug Discovery and Development*. SLAS Discov, 2017. 22(5):456-72.

[58] Langhans SA. *Three-Dimensional in Vitro Cell Culture Models in Drug Discovery and Drug Repositioning*. Frontiers in Pharmacology, 2018. 9(6).

[59] Kretzschmar K, Clevers H. *Organoids: Modeling Development and the Stem Cell Niche in a Dish*. Dev Cell, 2016. 38(6):590-600.

[60] Lancaster MA, Knoblich JA. *Organogenesis in a dish: Modeling development and disease using organoid technologies*. Science, 2014. 345(6194):1247125.

[61] Sachs N, et al. *A Living Biobank of Breast Cancer Organoids Captures Disease Heterogeneity*. Cell, 2018. 172(1):373-86.e10.

[62] Kleinman HK, Martin GR. *Matrigel: Basement membrane matrix with biological activity*. Seminars in Cancer Biology, 2005. 15(5):378-86.

[63] Liu G, et al. *Human breast cancer decellularized scaffolds promote epithelial-to-mesenchymal transitions and stemness of breast cancer cells in vitro*. J Cell Physiol, 2019. 234(6):9447-56.

[64] Dunne LW, et al. *Human decellularized adipose tissue scaffold as a model for breast cancer cell growth and drug treatments.* Biomaterials, 2014. 35(18):4940-9.

[65] Pinto ML, et al. *Decellularized human colorectal cancer matrices polarize macrophages towards an anti-inflammatory phenotype promoting cancer cell invasion via CCL18.* Biomaterials, 2017. 124:211-24.

[66] Grün D, et al. *Single-cell messenger RNA sequencing reveals rare intestinal cell types.* Nature, 2015. 525(7568):251-5.

[67] Zeisel A, et al. *Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq.* Science, 2015. 347(6226):1138-42.

[68] Treutlein B, et al. *Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq.* Nature, 2016. 534(7607):391-5.

[69] Rizvi AH, et al. *Single-cell topological RNA-seq analysis reveals insights into cellular differentiation and development.* Nature biotechnology, 2017. 35(6):551-60.

[70] Zhang Y, et al. *Single-cell transcriptome analysis reveals tumor immune microenvironment heterogenicity and granulocytes enrichment in colorectal cancer liver metastases.* Cancer Letters, 2019.

[71] Azizi E, et al. *Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment.* Cell, 2018. 174(5):1293-308.e36.

[72] Miyamoto DT, et al. *RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance.* Science, 2015. 349(6254):1351-6.

[73] Shaffer SM, et al. *Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance.* Nature, 2017. 546(7658):431-5.

[74] Eberwine J, et al. *Analysis of gene expression in single live neurons.* Proceedings of the National Academy of Sciences, 1992. 89(7):3010.

[75] Tang F, et al. *mRNA-Seq whole-transcriptome analysis of a single cell.* Nature methods, 2009. 6(5):377-82.

[76] Hedlund E, Deng Q. *Single-cell RNA sequencing: Technical advancements and biological applications.* Molecular aspects of medicine, 2018. 59:36-46.

[77] Macosko EZ, et al. *Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets.* Cell, 2015. 161(5):1202-14.

[78] Zheng GX, et al. *Massively parallel digital transcriptional profiling of single cells.* Nat Commun, 2017. 8:14049.

[79] Ziegenhain C, et al. *Comparative Analysis of Single-Cell RNA Sequencing Methods.* Molecular Cell, 2017. 65(4):631-43.e4.

[80] Picelli S, et al. *Full-length RNA-seq from single cells using Smart-seq2.* Nat Protocols, 2014. 9(1):171-81.

[81] Kolodziejczyk AA, Lönnberg T. *Global and targeted approaches to single-cell transcriptome characterization.* Briefings in functional genomics, 2018. 17(4):209-19.

[82] Ståhlberg A, Kubista M. *Technical aspects and recommendations for single-cell qPCR.* Molecular aspects of medicine, 2018. 59:28-35.

[83] Ståhlberg A, Kubista M, Åman P. *Single-cell gene-expression profiling and its potential diagnostic applications.* Expert Review of Molecular Diagnostics, 2011. 11(7):735-40.

[84] Ståhlberg A, Kubista M. *The workflow of single-cell expression profiling using quantitative real-time PCR*. Expert Review of Molecular Diagnostics, 2014. 14(3):323-31.

[85] Luecken MD, Theis FJ. *Current best practices in single-cell RNA-seq analysis: a tutorial*. Molecular systems biology, 2019. 15(6):e8746.

[86] Vallejos CA, et al. *Normalizing single-cell RNA sequencing data: challenges and opportunities*. Nature methods, 2017. 14(6):565-71.

[87] Brennecke P, et al. *Accounting for technical noise in single-cell RNA-seq experiments*. Nat Meth, 2013. 10(11):1093-5.

[88] Soneson C, Robinson MD. *Bias, robustness and scalability in single-cell differential expression analysis*. Nature methods, 2018. 15(4):255-61.

[89] Saelens W, et al. *A comparison of single-cell trajectory inference methods*. Nature biotechnology, 2019. 37(5):547-54.

[90] Guo G, et al. *Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire*. Cell Stem Cell, 2013. 13(4):492-505.

[91] Andersson D, et al. *Properties of targeted preamplification in DNA and cDNA quantification*. Expert Review of Molecular Diagnostics. 0(0):1-16.

[92] Dalerba P, et al. *Single-cell dissection of transcriptional heterogeneity in human colon tumors*. Nature biotechnology, 2011. 29(12):1120-7.

[93] Kurimoto K, et al. *An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis*. Nucleic Acids Res, 2006. 34(5):e42.

[94] Islam S, et al. *Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq*. Genome research, 2011. 21(7):1160-7.

[95] Ståhlberg A, Bengtsson M. *Single-cell gene expression profiling using reverse transcription quantitative real-time PCR*. Methods, 2010. 50(4):282-8.

[96] Sheng K, et al. *Effective detection of variation in single-cell transcriptomes using MATQ-seq*. Nature methods, 2017. 14:267.

[97] Fan X, et al. *Single-cell RNA-seq transcriptome analysis of linear and circular RNAs in mouse preimplantation embryos*. Genome Biology, 2015. 16(1):148.

[98] Verboom K, et al. *SMARTer single cell total RNA sequencing*. Nucleic Acids Research, 2019. 47(16):e93-e.

[99] Hayashi T, et al. *Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs*. Nature Communications, 2018. 9(1):619.

[100] Marguerat S, Bähler J. *Coordinating genome expression with cell size*. Trends in Genetics, 2012. 28(11):560-5.

[101] Mitchison JM. *Growth During the Cell Cycle*. International Review of Cytology: Academic Press; 2003. p. 165-258.

[102] Dolatabadi S, et al. *Cell cycle and cell size dependent gene expression reveals distinct subpopulations at single-cell level*. Frontiers in genetics, 2017. 8(JAN).

[103] Hu Z, et al. *Nucleosome loss leads to global transcriptional up-regulation and genomic instability during yeast aging*. Genes Dev, 2014. 28(4):396-408.

[104] Lin CY, et al. *Transcriptional amplification in tumor cells with elevated c-Myc*. Cell, 2012. 151(1):56-67.

[105] Nie Z, et al. *c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells*. Cell, 2012. 151(1):68-79.

[106] Li Y, et al. *Global transcriptional and translational repression in human-embryonic-stem-cell-derived Rett syndrome neurons*. Cell Stem Cell, 2013. 13(4):446-58.

[107] Jiang L, et al. *Synthetic spike-in standards for RNA-seq experiments*. Genome research, 2011. 21(9):1543-51.

[108] Wills QF, et al. *Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments*. Nature biotechnology, 2013. 31:748.

[109] Yang L, et al. *Genomewide characterization of non-polyadenylated RNAs*. Genome Biology, 2011. 12(2).

[110] Malumbres M, Barbacid M. *Cell cycle, CDKs and cancer: a changing paradigm*. Nature Reviews Cancer, 2009. 9(3):153-66.

[111] Åman P, et al. *Regulatory mechanisms, expression levels and proliferation effects of the FUS-DDIT3 fusion oncogene in liposarcoma*. The Journal of pathology, 2016. 238(5):689-99.

[112] Engström K, et al. *The Myxoid/Round Cell Liposarcoma Fusion Oncogene FUS-DDIT3 and the Normal DDIT3 Induce a Liposarcoma Phenotype in Transfected Human Fibrosarcoma Cells*. The American Journal of Pathology, 2006. 168(5):1642-53.

[113] Ashburner M, et al. *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium*. Nat Genet, 2000. 25(1):25-9.

[114] Croft D, et al. *Reactome: a database of reactions, pathways and biological processes*. Nucleic Acids Res, 2011. 39(Database issue):D691-7.

[115] Santos A, Wernersson R, Jensen LJ. *Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes*. Nucleic Acids Res, 2015. 43(Database issue):D1140-4.

[116] Leng N, et al. *Oscope identifies oscillatory genes in unsynchronized single-cell RNA-seq experiments*. Nature methods, 2015. 12(10):947-50.

[117] Morgan DO. *The Cell Cycle: Principles of Control*: OUP/New Science Press; 2007.

[118] Iwamoto F, et al. *Transcription-dependent nucleolar cap localization and possible nuclear function of DExH RNA helicase RHAU*. Exp Cell Res, 2008. 314(6):1378-91.

[119] Raj A, van Oudenaarden A. *Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences*. Cell, 2008. 135(2):216-26.

[120] Zetterberg A, Killander D. *Quantitative cytophotometric and autoradiographic studies on the rate of protein synthesis during interphase in mouse fibroblasts in vitro*. Experimental Cell Research, 1965. 40(1):1-11.

[121] Olofsson A, et al. *Abnormal expression of cell cycle regulators in FUS-CHOP carrying liposarcomas*. Int J Oncol, 2004. 25(5):1349-55.

[122] Kåbjörn Gustafsson C, et al. *Cell senescence in myxoid/round cell liposarcoma*. Sarcoma, 2014. 2014:208786-.

[123] Kreso A, Dick JE. *Evolution of the cancer stem cell model*. Cell Stem Cell, 2014. 14(3):275-91.

[124] Riggi N, et al. *Expression of the FUS-CHOP fusion protein in primary mesenchymal progenitor cells gives rise to a model of myxoid liposarcoma*. Cancer Res, 2006. 66(14):7016-23.

[125] Charytonowicz E, et al. *PPARgamma agonists enhance ET-743-induced adipogenic differentiation in a transgenic mouse model of myxoid round cell liposarcoma.* J Clin Invest, 2012. 122(3):886-98.

[126] Dolatabadi S, et al. *JAK–STAT signalling controls cancer stem cell properties including chemotherapy resistance in myxoid liposarcoma.* International Journal of Cancer, 2019. 145(2):435-49.

[127] Rawlings JS, Rosler KM, Harrison DA. *The JAK/STAT signaling pathway.* Journal of Cell Science, 2004. 117(8):1281-3.

[128] Marotta LLC, et al. *The JAK2/STAT3 signaling pathway is required for growth of CD44 +CD24- stem cell-like breast cancer cells in human tumors.* Journal of Clinical Investigation, 2011. 121(7):2723-35.

[129] Villalva C, et al. *STAT3 is essential for the maintenance of neurosphere-initiating tumor cells in patients with glioblastomas: a potential for targeted therapy?* Int J Cancer, 2011. 128(4):826-38.

[130] Wan S, et al. *Tumor-associated macrophages produce interleukin 6 and signal via STAT3 to promote expansion of human hepatocellular carcinoma stem cells.* Gastroenterology, 2014. 147(6):1393-404.

[131] Göransson M, et al. *Myxoid liposarcoma FUS-DDIT3 fusion oncogene induces C/EBP β-mediated interleukin 6 expression.* International Journal of Cancer, 2005. 115(4):556-60.

[132] Yu H, et al. *Revisiting STAT3 signalling in cancer: new and unexpected biological functions.* Nature Reviews Cancer, 2014. 14:736.

[133] Ambrogio C, et al. *NPM-ALK oncogenic tyrosine kinase controls T-cell identity by transcriptional regulation and epigenetic silencing in lymphoma cells.* Cancer Res, 2009. 69(22):8611-9.

[134] Wu J, et al. *Kaposi's sarcoma-associated herpesvirus (KSHV) vIL-6 promotes cell proliferation and migration by upregulating DNMT1 via STAT3 activation.* PLoS One, 2014. 9(3):e93478.

[135] Sharma S, Kelly TK, Jones PA. *Epigenetics in cancer.* Carcinogenesis, 2010. 31(1):27-36.

[136] Sparmann A, van Lohuizen M. *Polycomb silencers control cell fate, development and cancer.* Nature Reviews Cancer, 2006. 6(11):846-56.

[137] Lindén M, et al. *FET family fusion oncoproteins target the SWI/SNF chromatin remodeling complex.* EMBO reports, 2019. 20(5):e45766.

[138] Chen Y, et al. *Bromodomain and extraterminal proteins foster the core transcriptional regulatory programs and confer vulnerability in liposarcoma.* Nature Communications, 2019. 10(1):1353.

[139] Lawson DA, et al. *Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells.* Nature, 2015. 526(7571):131-5.

[140] Al-Hajj M, et al. *Prospective identification of tumorigenic breast cancer cells.* Proceedings of the National Academy of Sciences of the United States of America, 2003. 100(7):3983-8.

[141] Fillmore CM, Kuperwasser C. *Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy.* Breast Cancer Research, 2008. 10(2).

[142] Ginestier C, et al. *ALDH1 Is a Marker of Normal and Malignant Human Mammary Stem Cells and a Predictor of Poor Clinical Outcome.* Cell Stem Cell, 2007. 1(5):555-67.

[143] Ricardo S, et al. *Breast cancer stem cell markers CD44, CD24 and ALDH1: expression distribution within intrinsic molecular subtype.* Journal of Clinical Pathology, 2011. 64(11):937.

[144] Li W, et al. *Unraveling the roles of CD44/CD24 and ALDH1 as cancer stem cell markers in tumorigenesis and metastasis.* Scientific reports, 2017. 7(1):13856.

[145] Hwang-Verslues WW, et al. *Multiple lineages of human breast cancer stem/progenitor cells identified by profiling with stem cell markers.* PLoS One, 2009. 4(12):e8377.

[146] Collina F, et al. *Prognostic Value of Cancer Stem Cells Markers in Triple-Negative Breast Cancer.* BioMed research international, 2015. 2015:158682.

[147] Khramtsov AI, et al. *Wnt/β-catenin pathway activation is enriched in basal-like breast cancers and predicts poor outcome.* American Journal of Pathology, 2010. 176(6):2911-20.

[148] Harrison H, et al. *Regulation of breast cancer stem cell activity by signaling through the Notch4 receptor.* Cancer Res, 2010. 70(2):709-18.

[149] Liu S, et al. *Hedgehog signaling and Bmi-1 regulate self-renewal of normal and malignant human mammary stem cells.* Cancer Res, 2006. 66(12):6063-71.

[150] Zhou J, et al. *Activation of the PTEN/mTOR/STAT3 pathway in breast cancer stem-like cells is required for viability and maintenance.* Proceedings of the National Academy of Sciences of the United States of America, 2007. 104(41):16158-63.

[151] Zhou BBS, et al. *Tumour-initiating cells: Challenges and opportunities for anticancer drug discovery.* Nature Reviews Drug Discovery, 2009. 8(10):806-23.

[152] Korkaya H, et al. *Regulation of mammary stem/progenitor cells by PTEN/Akt/beta-catenin signaling.* PLoS Biol, 2009. 7(6):e1000121.

[153] Ramaswamy B, et al. *Hedgehog signaling is a novel therapeutic target in tamoxifen-resistant breast cancer aberrantly activated by PI3K/AKT pathway.* Cancer Research, 2012. 72(19):5048-59.

[154] Gurney A, et al. *Wnt pathway inhibition via the targeting of Frizzled receptors results in decreased growth and tumorigenicity of human tumors.* Proceedings of the National Academy of Sciences of the United States of America, 2012. 109(29):11717-22.

[155] Gupta PB, et al. *Identification of Selective Inhibitors of Cancer Stem Cells by High-Throughput Screening.* Cell, 2009. 138(4):645-59.

[156] Ginestier C, et al. *CXCR1 blockade selectively targets human breast cancer stem cells in vitro and in xenografts.* Journal of Clinical Investigation, 2010. 120(2):485-97.

[157] Hirsch HA, et al. *Metformin selectively targets cancer stem cells, and acts together with chemotherapy to block tumor growth and prolong remission.* Cancer Research, 2009. 69(19):7507-11.

[158] Hu Y, et al. *Sabutoclax, pan-active BCL-2 protein family antagonist, overcomes drug resistance and eliminates cancer stem cells in breast cancer.* Cancer Lett, 2018. 423:47-59.

[159] Chen YC, et al. *Single-cell RNA-sequencing of migratory breast cancer cells: Discovering genes associated with cancer metastasis.* Analyst, 2019. 144(24):7296-309.

[160] Tong M, et al. *Transcriptomic but not genomic variability confers phenotype of breast cancer stem cells*. Cancer Communications, 2018. 38(1):56.

[161] Dolatabadi S, et al. *Cell Cycle and Cell Size Dependent Gene Expression Reveals Distinct Subpopulations at Single-Cell Level*. Frontiers in genetics, 2017. 8:1.

[162] Pastrana E, Silva-Vargas V, Doetsch F. *Eyes wide open: a critical review of sphere-formation as an assay for stem cells*. Cell Stem Cell, 2011. 8(5):486-98.

[163] Sanchez Carlos G, et al. *Regulation of Ribosome Biogenesis and Protein Synthesis Controls Germline Stem Cell Differentiation*. Cell Stem Cell, 2016. 18(2):276-90.

[164] Sampath P, et al. *A Hierarchical Network Controls Protein Translation during Murine Embryonic Stem Cell Self-Renewal and Differentiation*. Cell Stem Cell, 2008. 2(5):448-60.

[165] Marguerat S, et al. *Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells*. Cell, 2012. 151(3):671-83.

[166] Gyorffy B, et al. *An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients*. Breast Cancer Res Treat, 2010. 123(3):725-31.

[167] Li X, et al. *The lncRNA NEAT1 facilitates cell growth and invasion via the miR-211/HMGA2 axis in breast cancer*. International Journal of Biological Macromolecules, 2017. 105:346-53.

[168] Brown Y, Hua S, Tanwar PS. *Extracellular matrix-mediated regulation of cancer stem cells and chemoresistance*. Int J Biochem Cell Biol, 2019. 109:90-104.

[169] Keshava Prasad TS, et al. *Human Protein Reference Database--2009 update*. Nucleic Acids Res, 2009. 37(Database issue):D767-72.

[170] Balachander GM, et al. *Enhanced Metastatic Potential in a 3D Tissue Scaffold toward a Comprehensive in Vitro Model for Breast Cancer Metastasis*. ACS Appl Mater Interfaces, 2015. 7(50):27810-22.

[171] Chen L, et al. *The enhancement of cancer stem cell properties of MCF-7 cells in 3D collagen scaffolds for modeling of cancer and anti-cancer drugs*. Biomaterials, 2012. 33(5):1437-44.

[172] Sims-Mourtada J, et al. *Enrichment of breast cancer stem-like cells by growth on electrospun polycaprolactone-chitosan nanofiber scaffolds*. Int J Nanomedicine, 2014. 9:995-1003.

[173] Linnarsson S, Teichmann SA. *Single-cell genomics: coming of age*. Genome Biology, 2016. 17(1):97.

[174] Gawad C, Koh W, Quake SR. *Single-cell genome sequencing: current state of the science*. Nature Reviews Genetics, 2016. 17(3):175-88.

[175] Lindén M, et al. *FET family fusion oncoproteins target the SWI/SNF chromatin remodeling complex*. EMBO Rep, 2019. 20(5).

[176] Schwartzman O, Tanay A. *Single-cell epigenomics: Techniques and emerging applications*. Nature Reviews Genetics, 2015. 16(12):716-26.

[177] Clark SJ, et al. *Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity*. Genome Biology, 2016. 17(1):72.

[178] Edfors F, et al. *Gene-specific correlation of RNA and protein levels in human cells and tissues*. Molecular systems biology, 2016. 12(10):883-.

[179] Vogel C, Marcotte EM. *Insights into the regulation of protein abundance from proteomic and transcriptomic analyses.* Nature Reviews Genetics, 2012. 13(4):227-32.

[180] Levy E, Slavov N. *Single cell protein analysis for systems biology.* Essays in biochemistry, 2018. 62(4):595-605.

[181] Svec D, et al. *Identification of inhibitors regulating cell proliferation and FUS-DDIT3 expression in myxoid liposarcoma using combined DNA, mRNA, and protein analyses.* Laboratory Investigation, 2018. 98(7):957-67.

[182] Darmanis S, et al. *Simultaneous Multiplexed Measurement of RNA and Proteins in Single Cells.* Cell Reports, 2016. 14(2):380-9.

[183] Stoeckius M, et al. *Simultaneous epitope and transcriptome measurement in single cells.* Nature methods, 2017. 14(9):865-8.

[184] Macaulay IC, Ponting CP, Voet T. *Single-Cell Multiomics: Multiple Measurements from Single Cells.* Trends in Genetics, 2017. 33(2):155-68.

[185] Macaulay IC, et al. *G&T-seq: parallel sequencing of single-cell genomes and transcriptomes.* Nature methods, 2015. 12(6):519-22.

[186] Dey SS, et al. *Integrated genome and transcriptome sequencing of the same cell.* Nature biotechnology, 2015. 33(3):285-9.

[187] Angermueller C, et al. *Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity.* Nature methods, 2016. 13(3):229-32.

[188] Hu Y, et al. *Simultaneous profiling of transcriptome and DNA methylome from a single cell.* Genome Biol, 2016. 17:88.

[189] Hou Y, et al. *Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas.* Cell Research, 2016. 26(3):304-19.

[190] Akrap N, et al. *Identification of Distinct Breast Cancer Stem Cell Populations Based on Single-Cell Analyses of Functionally Enriched Stem and Progenitor Pools.* Stem Cell Reports, 2016. 6(1):121-36.

[191] Mignardi M, et al. *Bridging Histology and Bioinformatics - Computational Analysis of Spatially Resolved Transcriptomics.* Proceedings of the IEEE, 2017. 105(3):530-41.

[192] Crosetto N, Bienko M, Van Oudenaarden A. *Spatially resolved transcriptomics and beyond.* Nature Reviews Genetics, 2015. 16(1):57-66.

[193] Svedlund J, et al. *Generation of in situ sequencing based OncoMaps to spatially resolve gene expression profiles of diagnostic and prognostic markers in breast cancer.* EBioMedicine, 2019. 48:212-23.

[194] Ye J, et al. *The cancer stem cell niche: cross talk between cancer stem cells and their microenvironment.* Tumor Biology, 2014. 35(5):3945-51.

[195] Lu H, et al. *A breast cancer stem cell niche supported by juxtacrine signalling from monocytes and macrophages.* Nature Cell Biology, 2014. 16(11):1105-17.