Norwegian University
of Life Sciences

Master's Thesis 2021 60 ECTS
Faculty of Chemistry, Biotechnology and Food Science

Molecular characterization of single tumor cells in breast cancer

Julie Synøve Myre Monrad
Master of Science in Biotechnology

A thesis submitted for the degree of
Master's Programme in Biotechnology, 60 ETC

Title:

Molecular characterization of single tumor cells in breast cancer

By:

Julie Synøve Myre Monrad

External supervisors:

Inger Riise Bergheim and Hege G. Russnes

Internal supervisor:

Morten Kjos

Department of Cancer Genetics, Institute for Cancer Research,
The Norwegian Radium Hospital, OUH, Oslo, Norway

Faculty of Chemistry, Biotechnology and Food Science,
NMBU – Norwegian University of Life Sciences

June 2021

Acknowledgements

This master thesis was completed as a part of the Master program in Biotechnology at the University of Life Sciences (NMBU). The work presented was carried out at the Department of Cancer Genetics, Institute for Cancer Research at Oslo University Hospital (the Norwegian Radium Hospital) from August 2020 until June 2021.

First, I would like to thank my main supervisor Dr. Hege Russnes, for believing in me and giving me the opportunity to do my thesis in her research group. Thank you for all your good feedback and shared knowledge, as well as inspiring passion for the research. I would also like to thank my co-supervisor Inger Riise Bergheim, for your guidance and help with the laboratory work. I am so grateful of your knowledge and support. My research could not have been done without the support from both of you.

I would like to thank my helpful colleagues and collaborators, Cecilie Bendigtsen Schirmer for your help with CellSearch, Monica Bostad for your help with FACS, Helen Vålerhaugen for your help with ddPCR assay design, Karin Teien Lande and the rest of the Russnes group for help with informative discussions, and Ole Christian Lingjærde and Arne Valebjørg Pladsen for your help with R-plots for result visualization. Thanks to NMBU supervisor for corrections and guidance in the writing process, Morten Kjos.

I would also like to thank the rest of my colleagues at the Department of Cancer Genetics, for making my time here engaging and educational. I cannot wait to learn more from all of you and continue with this exciting work.

Finally, I would like to thank my friends and family, for great support during this particularly special covid-19 year. It is finally time for hugs and fun times together!

Abstract

Breast cancer is the most common cancer among women in Norway, with 3726 new cases and 598 deaths registered in 2019. The five-year relative survival rate has increased and is now at 92%. Breast cancer is one of the few cancer types where patients can experience relapse of the disease many years after the initial diagnosis and treatment. To-day, there are no established diagnostic markers that can predict the risk for late relapse. Breast cancer patients can have tumor cells residing in the bone marrow, so-called micrometastases. Such single disseminating tumor cells (DTCs) can be dormant for years. For breast cancer patients with dormant DTCs in the bone marrow, it is important to understand how they might be activated and form metastases after many years. Information about molecular features of such cells would therefore be of major interest. Sequencing of single cells is a technique that have increased in the last decade, due to better technology and more interest in the heterogeneity and cell-to-cell variation, both in tumors and for rare cell populations. The phenotype of cells is defined by the transcriptome, as the transcription of RNA and translation into proteins defines the activities in a cell. It is therefore of critical importance to have a standard method for RNA isolation from single cells to allow single cell transcription analysis.

Single cell RNA sequencing has mainly been by sequencing of many individual unselected cells in suspension, but rarely of selected individual single cells. A “pipeline” for identification and selection of rare tumor cells by a process that results in RNA of high amount and quality, would be of great value.

The aim of this thesis was to compare the RNA output from four methods with different processes for identification and selection of single tumor cells, as detailed analysis of the RNA demands optimal concentration, fragment lengths and suitability for PCR amplification. A cell line with epithelial tumor cells from breast was used for testing of the methods. The RNA extracted from selected single cells, was tested for amount, quality and amplificability. Further, DTCs from two breast cancer patients were identified and selected by one of the methods and subjected to RNA extraction and quality control.

One of the four methods, the microinjection pipette method, was used as a reference, with unstained and viable cells close to physiological conditions. This method gave RNA from all the isolated single cells, with good quality and amplificability. However, this method is not applicable directly on bone marrow from patient samples, as the rare tumor cells would not be possible to identify and isolate. The second method, identification process by cytospin and

isolation by the micromanipulator, used fixation and staining of the tumor cells by cytokeratin antibodies bound to epithelial tumor cells, and alkaline phosphatase reactions making the tumor cells visible in the microscope. This method had barely measurable concentration of RNA from only a few of the isolated single tumor cells, but not with a good enough quality or amplificability for further analysis. The third method, Fluorescence-activated cell sorting (FACS), identified and sorted single tumor cells based on fluorescent surface antibodies, by a process possible to use on patient samples, separating the rare tumor cells from the normal cells. The RNA output from cells selected by this method was lower than the reference method. Still, RNA from most of the isolated cells had longer RNA fragments indicating good quality, and RNA from 1/3 of the cells was amplificable by expression analysis. The fourth method, a fullblood sample spiked with tumor cells, followed by enrichment of tumor cells by CellSearch and isolation by DEPArray, had the longest total processing time and number of steps. The cells were fixated and stained with antibodies for cytokeratin, and an extraction process between the instruments has potential of losing cells before the isolation by the DEPArray system. This method had slightly higher concentration than the micromanipulator method, but very low, and only a few cells had RNA with longer fragments. In a various degree, 1/5 of the isolated cells had RNA that was amplificable by expression analysis.

DTCs from the patient samples were successfully identified and selected by the FACS method, with a small selection of single cells measured and tested with quality analysis. Even after +/- 20 years of storage, the cells had measurable concentration, fragments of good integrity and amplificable expression of the cDNA, confirming the cell type of the selected cells.

This work has contributed insight into the influence single cell identification and isolation methods for rare tumor cells have on the RNA amount and quality. Further testing of a pipeline for extraction of both DNA and RNA from rare single cells can provide more information about minimal residual disease in metastatic breast cancer patients.

Sammendrag

Brystkreft er den vanligste formen for kreft blant kvinner i Norge, med 3726 nye tilfeller og 598 registrerte dødsfall i 2019. Den fem-års relative overlevelsesraten har økt og er nå på 92%. Brystkreft er en av de få kreft typene hvor pasienter kan oppleve tilbakefall av sykdommen mange år etter første diagnose og behandling. Til dags dato er det ingen gode diagnostiske markører som kan predikere risikoen for slike sene tilbakefall. Brystkreftpasienter kan ha kreft celler i benmargen, såkalte mikrometastaser. Slike spredte, disseminerte kreftceller (DTC) kan være inaktive (dormant) i flere år. For brystkreftpasienter med DTC i benmargen er det viktig å forstå hvordan de kan aktiveres og danne metastaser etter mange år. Informasjon om molekylære trekk ved slike celler vil derfor være av stor interesse. Sekvensering av enkeltceller er en teknikk som har blitt mer brukt de siste tiårene, på grunn av bedre teknologi og en økt interesse i heterogenitet og celle-til-celle variasjon, både i svulster og for sjeldne cellepopulasjoner. Cellers fenotype er definert av transkriptomet, ettersom transkripsjonen av DNA til RNA og translasjonen deretter til proteiner definerer aktivitetene til en celle. Det er derfor kritisk viktig å ha en standardmetode for isolasjon av RNA fra enkeltceller, for å kunne få til sekvensering av enkeltcelle-transkriptomet.

Sekvensering av enkeltcelle-RNA har hovedsakelig vært ved sekvensering av mange individuelle uselekterte celler i suspensjon, men sjelden av utvalgte individuelle enkeltceller. Å ha en «pipeline» for identifikasjon og seleksjon av sjeldne tumorceller, som klarer å gi RNA av høy mengde og kvalitet, vil derfor være viktig.

Målet med denne masteroppgaven var å sammenligne RNA mengden etter bruk av fire metoder med ulike identifikasjons og seleksjons prosesser av enkelte tumorceller. Dette fordi en detaljert sekvenseringsanalyse av RNA krever optimal konsentrasjon, større fragmenter og egnethet for PCR amplifikasjon. En cellelinje med epiteliale tumorceller fra bryst ble brukt til testing av metodene. RNA ekstrahert fra de utvalgte enkeltcellene ble testet for mengde, kvalitet og amplifiserbarhet. Videre ble DTC fra to brystkreftpasienter identifisert og selektert med en av metodene etterfulgt av RNA analyse.

En av de fire metodene, mikroinjeksjonspipette-metoden, ble brukt som referanse, med ufargede, levende celler, da denne var nærmest normale fysiologiske forhold. Denne metoden gav RNA fra alle de isolerte enkeltcellene, med god kvalitet og amplifiserbarhet. Metoden er derimot ikke mulig å bruke direkte på benmarg fra pasientprøver, siden det vil være umulig å identifisere de sjeldne tumorcellene. Den andre metoden, med identifikasjonsprosess ved

cytospin og isolasjon ved hjelp av mikromanipulatoren, bruker fiksering og farging av tumorcellene ved antistoff mot cytokeratin bundet til epiteliale tumorceller, og alkalisk fosfatase-reaksjoner som gjør tumorcellene synlige i mikroskopet. Denne metoden gav såvidt målbare konsentrasjoner av RNA fra kun et fåtall av de isolerte tumorcellene, men ikke med god nok kvalitet eller amplifiserbarhet til videre analyse. Den tredje metoden, fluorescens-aktivert celle sortering (FACS), identifiserte og sorterte enkelte tumorceller basert på fluorescerende overflate antistoff, ved en prosess som er mulig å bruke på pasientprøver, hvor de sjeldne tumorcellene blir adskilt fra de vanlige cellene. RNA mengden fra celler identifisert og isolert med denne metoden var lavere enn referansemetoden. Likevel, RNA fra flesteparten hadde lengre RNA fragmenter som indikerte god kvalitet, og RNA fra 1/3 av cellene var amplifiserbare ved ekspresjonsanalyse. Den fjerde metoden, en fullblodprøve tilsatt tumorceller, etterfulgt av anrikning av tumorcellene med CellSearch og isolering med DEPArray, hadde den lengste prosesseringstiden og flest antall prosesseringssteg. Cellene var fiksert og farget med antistoff mot cytokeratin etterfulgt av en ekstraksjonsprosess mellom instrumentene hvor man potensielt kan miste celler før isolering ved bruk av DEPArray instrumentet. Noen av cellene valgt ut ved denne metoden hadde såvidt høyere konsentrasjon enn mikromanipulator-metoden, og kun et fåtall celler hadde RNA med lengre fragmenter. For 1/5 av cellene var det varierende grad av RNA som var amplifiserbart ved ekspresjonsanalysen.

En rekke DTC fra pasientprøvene ble identifisert og selektert ved FACS metoden, med et lite utvalg av enkeltcellene som ble målt og testet med kvalitetskontroll. Til og med etter +/-20 års oppbevaring i nitrogenfryser, hadde cellene målbar konsentrasjon, fragmenter av god integritet og amplifiserbar ekspresjon av RNA som bekreftet celletypen til de selekterte cellene.

Dette arbeidet har bidratt med innblikk i påvirkningen enkeltcelleidentifisering og seleksjonsmetoder for sjeldne tumorceller har for mengden og kvaliteten til RNA. Videre testing av en «pipeline» for ekstraksjon av både DNA og RNA fra sjeldne enkeltceller kan gi mer informasjon om minimal gjenværende sykdom for brystkreftpasienter med metastatisk sykdom.

Abbreviations

Abbreviations	Description
APC	Allophycocyanin
BRCA	Breast cancer associated gene
CD	Lymphocyte common antigen
cDNA	Complementary DNA
CK	Cytokeratin
CNA	Copy Number Aberrations
CSC	Cancer stem-cell
CTC	Circulating tumor cell
DCIS	Ductal carcinoma *in situ*
DNA	Deoxyribonucleic acid
DTC	Disseminating tumor cell
ECM	Extracellular matrix
EMT	Epithelial mesenchymal transition
EpCAM	Epithelial cell adhesion molecule
ER	Estrogen receptor
FDA	Food and Drug Administration
FGFR	Fibroblast growth factor receptor
GWAS	Genome Wide Association Studies
HER2	Human epidermal growth factor receptor
IDC	Invasive ductal carcinoma
ILC	Invasive lobular carcinoma
IMT	Immunomagnetic bead technique
ISET	Isolation by Size of Epithelial Tumor cells
LCIS	Lobular carcinoma *in situ*
LCM	Laser microdissection capture
MEMS	Micro electro-mechanical system
MET	Mesenchymal epithelial transition
MRD	Minimal residual disease
mRNA	Messenger RNA
NBCG	Norwegian Breast Cancer Group
NGS	Next generation sequencing
PAM	Prediction analysis of microarray
PE	Phycoerythrin
PR	Progesterone receptor
RNA	Ribonucleic acid
ROR	Risk of Recurrence
SD	Standard deviation
SNP	Single nucleotide polymorphism
SOP	Standard operation procedure

TDLU	Terminal ductal lobular unit
TNM	Tumor, node, Metastasis
tRNA	Transfer ribonucleic acid
WGA	Whole Genome Amplification

Table of content

1. Introduction

1.1 The flow of genetic information

Cells produce proteins to execute their biological functions, by using their genetic information in the nucleus, deoxyribonucleic acid (DNA). DNA functions as a template for ribonucleic acid (RNA) synthesis in transcription and for its self-reproduction in DNA synthesis during cell division. RNA is further directed as a template for protein synthesis in translation. The central dogma of molecular biology was defined by Francis Crick, and published in 1958 as the one-way flow of genetic information (as seen in Figure 1), from DNA, to RNA, to protein. [1]. DNA consists of a sugar-phosphate backbone and four nucleic acid bases, adenine (A), cytosine (C), guanine (G) and thymine (T). The bases are bound in pairs of corresponding bases, adenine-thymine, and cytosine-guanine, with hydrogen bonds, creating a double helix of the DNA with complementary strands. The complete set of an organism's DNA is called its genome [2].

Figure 1. The central dogma of molecular biology. Visualization of how the genetic information flows in a cell. Figure obtained from BioRender.com.

1.1.1 Transcription and translation

The first process of the central dogma is transcription or RNA synthesis, when DNA is used as a template to make a single strand of RNA, performed by the protein RNA polymerase along

with multiple other proteins. The creation of the single strand, intermediate messenger RNA (mRNA), occurs in the nucleus. The mRNA is further processed to be ready for translation, by splicing out introns (untranslated sequences), capping of the 5'-end with a methylated guanine base and polyadenylation, adding a poly-A tail on the 3'-end of the mRNA. The mRNA then travels out of the nucleus and into the cytoplasm, aiming to find the macromolecular machines called ribosomes. In the ribosome, the mRNA sequence is translated to an amino acid sequence, with the help of transfer RNA (tRNA) molecules. Specific tRNA's has specific amino acids attached to it. Matching of three following bases on the mRNA molecule, known as a codon, to its three corresponding bases on a tRNA molecule, its *anti*codon, makes sure that the correct amino acid is transferred and added to the growing polypeptide chain [2].

The finished polypeptide folds into a stable structure based on the amino acid sequence, and this is an important part of the central dogma; not only does the nucleotide sequence translate the genes into the proteins' amino acid sequence, but also its three-dimensional structure, which is crucial for the function of the finished protein. The proteins function as the building blocks of the body, and supply with amino acids important for growth and maintenance of cells and tissue. The processes of transcription and translation results in the cells functioning proteins and expression of active genes [2].

1.1.2 Cell division and cell cycle

The continuity of life can be seen by one cell and its duplication and division into two cells, through the cell cycle. All living things reproduce through the cell cycle with its essential mechanisms. Before a cell can divide into two genetically identical daughter cells, it needs to copy its entire genome, the DNA in each chromosome must be replicated to two complete copies. Each daughter cell gets a complete copy of the genome and duplicates of the organelles and macromolecules. For eukaryotes, the cell cycle consists of four phases: the G_1 phase, S phase (DNA *synthesis*), G_2 phase and M phase (*mitosis*). During S phase, the chromosome duplication occurs, whereas most of the other cellular components are duplicated and the cell size is increased throughout the cycle, except during mitosis. During the M phase the duplicated chromosomes are segregated into separate nuclei (mitosis) and the cell divides into two separate cells (cytokinesis). The S and M phases are separated by the two G (gap) phases, where the progression of the cell cycle is regulated by intra- and extracellular signals. With good conditions, cells can delay the progress of G_1 phase, and some can enter a specialized resting

state, called G0 (zero). Some cells can remain in the G0 phase for days, weeks or even years. The cell cycle is a complex process with many control steps for a correct cell division [3].

1.2 Principles of cancer

All the cells of an organism contribute and function together as a complex and regulated system. When a normal cell turns into a cancer cell, it breaks several rules. Two of the properties that defines a cancer cell are: the ability to defy normal cell growth and division, and the ability to invade surrounding tissue [3]. The combination of these properties makes the cancer cell dangerous, as it has a potential to spread and colonize in distant tissue reserved for other cells. When an abnormal cell grows and divides out of control, it results in a defined mass, also called neoplasm or tumor. Tumors that have limited growth potential and do not invade surrounding tissue, are called benign. If the tumor cells have a continued growth, accumulated more alterations and have the ability to invade the surrounding tissue, it is defined as a malignant tumor, i.e. cancer [3]. Cancer is a genetic disease where alterations in the cell's DNA accumulates after cell divisions, called somatic mutations. Mutations are changes or damages in the DNA that disrupts the central dogma, and lead to degradation, modification, or a diversity in the gene expression. DNA mutations can be the loss or duplication of genes, such as copy number alterations (CNA), or larger structural changes in the chromosome. Mutations are caused by both intrinsic factors, such as age, hormonal status and DNA repair defects, and by environmental factors, such as UV-radiation, viral and bacterial infections, and unhealthy lifestyle. The genetic changes can be inherited as germline mutations or by processes that alter the gene activity without changing the DNA sequences, called epigenetic changes. Some examples of epigenetic changes are methylation, acetylation, and chromatin modification [4, 5].

1.2.1 Cancer genomics and the hallmarks of cancer

Douglas Hanahan and Robert Weinberg introduced in 2000 the six biological capabilities that hallmark cancer: sustaining proliferative signalling, evading growth suppressors, activating invasion and metastasis, inducing angiogenesis, and resisting cell death. Genomic instability is underlying these hallmarks and generates genetic diversity with possibility of multiple hallmark functions. A decade later, in 2011 Hanahan and Weinberg included genomic instability and mutation as an enabling characteristic of cancer, along with tumor-promoting inflammation,

and the emerging hallmarks of deregulating cellular energetics and the ability to avoid the immune system. The total ten features as seen in figure 2, includes the large range of cancer characteristics that are observed in malignant cells [6].

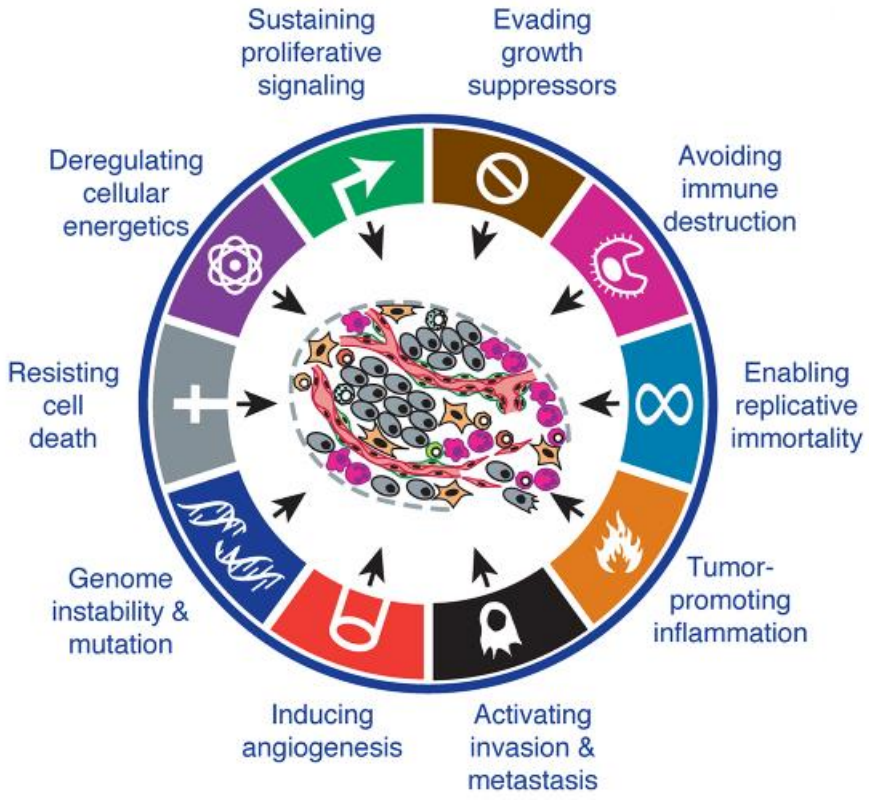

Figure 2. The hallmarks of cancer. The 10 hallmark characteristics of cancer modified from Hanahan and Weinberg, (2011) [6].

The genomic instability in cancer cells can come from defects in the capability of repairing DNA damage or to correct replication errors. All genes with an alteration that contribute to the evolution of tumorigenesis, are called cancer-critical genes. These genes are divided into two main groups. The first group is the proto-oncogenes, which with a gain-of-function mutation can drive a cell toward cancer, usually by stimulating proliferation. The mutant forms of the proto-oncogenes are called oncogenes, which is overactive or over expressed, such as the human epidermal growth factor receptor 2 (*HER2*) oncogene that promotes the growth of cancer cells. The second group are the tumor suppressor genes. With a loss-of-function mutation, these genes can contribute to cancer development by loss of the normal break function the genes have. In normal cells, the proteins of tumor suppressor genes lead the cell to controlled cell

death (apoptosis) in response to stress and DNA damage. The gene *TP53* is an example of a gene encoding a tumor suppressor protein, that is shown to be mutated in 50% of cancer cases [3].

For a normal cell, the cell cycle is under control by several checkpoints (G1, S, G2 and M) during cell division. Cells with one or more defect checkpoints can succeed and divide, and thus form a tumor with distorted proteins, resulting in abnormal properties. Such cells shall ideally be destroyed by the immune system. Several of the hallmarks are connected to genomic instability through inter- and intracellular signalling. Defects in the feedback systems of cell-signalling, may stimulate proliferation and tumor growth. Signals of proliferation and apoptosis regulation in the tumor cells may, in addition, be influenced by the signalling in the surrounding stroma cells [6].

1.2.2 Metastasis

The process of tumor cells leaving the primary tumor, spreading to distant sites to form new tumors, i.e., metastases, is referred to as the metastatic process (Figure 3). Metastatic disease is the main cause of cancer related death. The process consists of several steps, where both intrinsic factors of the tumor cells and the host's immune response will affect metastatic process [7]. For metastasis to occur, some tumor cells must migrate from the primary tumor, invade the surrounding tissue, and enter the blood and/or lymphatic system as circulating tumor cells (CTCs). The CTCs must survive in the circulation until they extravasate and lodge in a distant organ, where they can form metastases. The tumor cells that have reached a distant organ, such as lymph nodes or bone marrow, but have not formed a distinct mass, are referred to as micro metastasis, or disseminated tumor cells (DTCs) [8]. The genetic and epigenetic modifications decide the characteristics of the cancer cells and the surrounding microenvironment. Communication between the cancer cells and the tumor microenvironment, can help the cancer cell to survive stromal challenges, settle and colonize [9].

In 1889 Stephen Paget published his hypothesis about the "Seed and soil", where he described that the spread of a tumor is governed by the interactions between the cancer cells (seed) and the host organ (soil). His theory was that certain tumor cells have specific affinity for niches of certain organs, and that metastasis formed when the seed and soil were compatible. Since his hypothesis, the research on metastasis have continued with aim to fully understand the process [10].

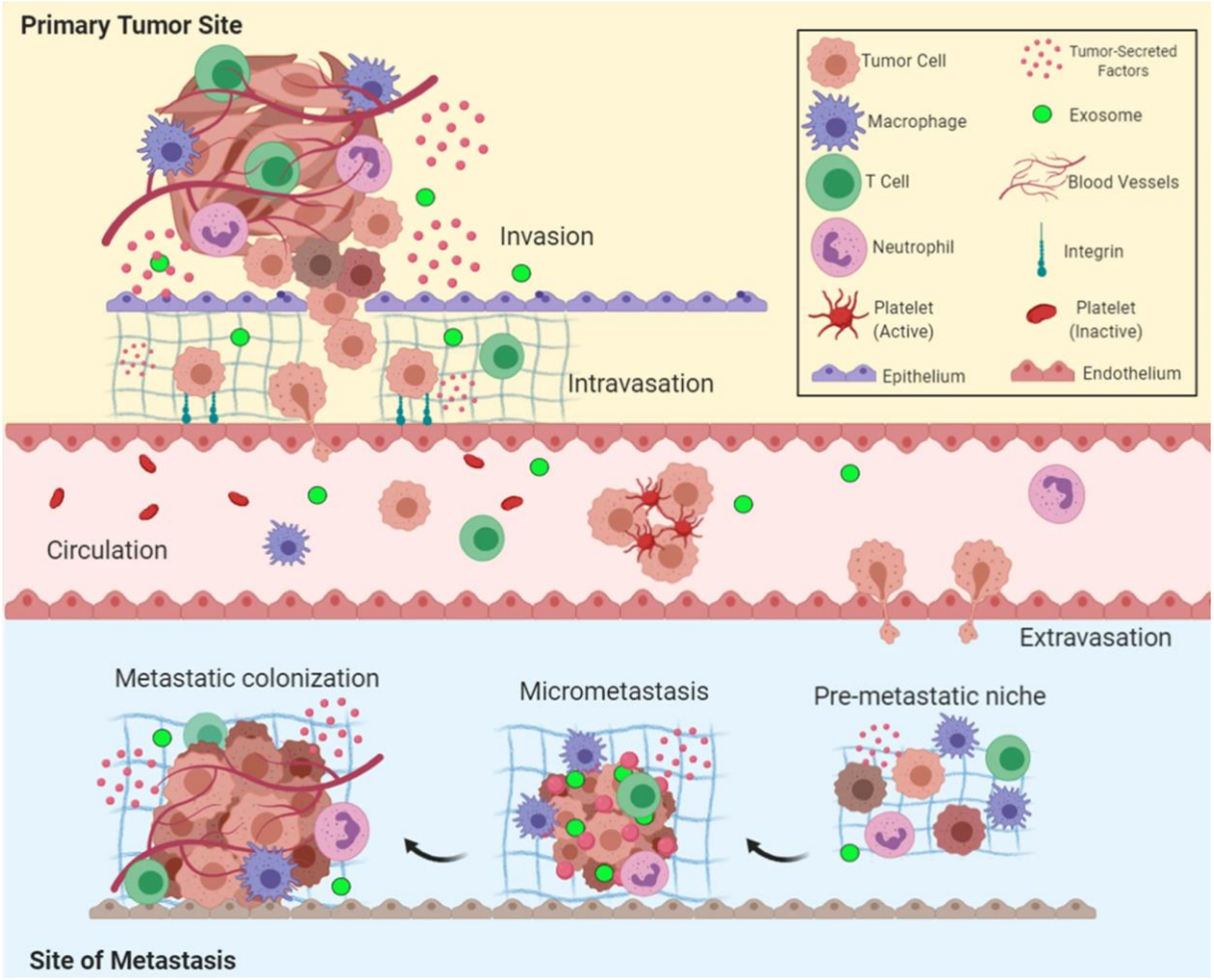

Figure 3. Metastatic cascade overview. The five key steps of metastasis: invasion, intravasation, circulation, extravasation, and colonization. Other cells in the microenvironment and cells from different tissues are included. Obtained from Fares et al. 2020 [9].

1.2.3 Tumor heterogeneity and tumor evolution

For best treatment of cancer, it is important to understand the systemic cancer progression. Klein et al. discussed in 2009 two models of the metastatic cascade; the linear progression model, and the parallel progression model. The linear progression model describes a late dissemination of fully malignant cells, that leave the primary tumor microenvironment and form new metastases in distant organs. The parallel progression model describes an earlier dissemination with acquisition of the malignant phenotype at a distant site, and a maturement of the cell during somatic progression and metastatic growth. The differences in these models lies in how similar or dissimilar the phenotype of the DTCs are to the cells in the primary tumor [7].

The biological heterogeneity of cancer cells in the primary tumor and its metastases, is the main barrier in treatment of metastasis [10]. Every tumor is unique, with a molecular fingerprint unlike any other tumor. Intertumor heterogeneity is the variation between patients that can be recognized in tumors by expressional subtypes, different morphology types, or classes of genomic copy number patterns [11]. During tumor progression, the combination of rapid evolution and increased genomic instability, creates clonal subpopulations [10]. The clonal heterogeneity in human tumors are histopathological diverse, with various degree of proliferation, differentiation, vascularity, inflammation, and invasiveness [6]. This variation within a single tumor, is referred to as intratumor heterogeneity. This has been observed by histopathologists for a long time, who have seen different morphologies and staining behaviours in tumor subpopulations. With new technology of whole genome amplification (WGA) and next generation sequencing (NGS), the intratumor heterogeneity has been defined at a molecular level by genetic variation between tumor subpopulations and among individual malignant cells [11]. The new possibilities in genetic analysis and profiling, with data analysis, creation of a hierarchy of subclones and phylogenetic lineages, can further increase the understanding of tumor evolution and trace the evolution back to the original clone. As demonstrated by Navin et. al. in 2011, robust high-resolution copy number profiles can be obtained by sequencing a single cell, and by examining multiple cells from the same cancers, inferences can be made about the tumor evolution and the spread of cancer [12].

Cellular plasticity is the phenomenon of cells ability to adopt different identities in their phenotype. For cells faced with physiologic and pathologic stress, the cellular plasticity is a mechanism for regeneration or tissue adaptation, but it can also predispose cancerous transformation in the tissue [13]. In cancer cells, the cellular plasticity along with genetic and epigenetic alterations, are mechanisms that promotes diversity and intra-tumor heterogeneity. The intratumor heterogeneity is associated with disease progression and impairment in response to treatment, as plasticity provides the cancer cells' the capacity to shift between differentiating states. The capacity to shift from a state that gives limited tumorigenic potential, to a more undifferentiating or cancer stem-cell like (CSC) state, responsible for long-term tumor growth [14]. Plasticity has shown to be able to convert the cellular phenotype between epithelial cells and cells with mesenchymal traits. The epithelial mesenchymal transition (EMT) is a process of cell adaptation that closely associate with tumor cell proliferation, cancer dormancy and metastasis. The process of EMT have been connected to the initiation of dormancy, as seen in

Figure 4, and backwards mesenchymal epithelial transition (MET) have been connected to reactivation of proliferation, which are fundamental processes for cancer cells to invade and metastasize. Cancer dormancy is a period of cancer progression where residual cancer cells can be resistant to conventional chemo- and radiotherapies and be clinically asymptomatic for a long time. The dormant cells are in a stage of growth arrest, which can occur in the primary tumor formation or after dissemination as DTCs. Many patients relapse years or even decades after radical surgery, as the dormant cells can be activated, the disease can recur with metastases. This necessitates the importance of understanding the whole process of conversion between cell states, in connection to the relationship between EMT, dormancy and metastasis [15].

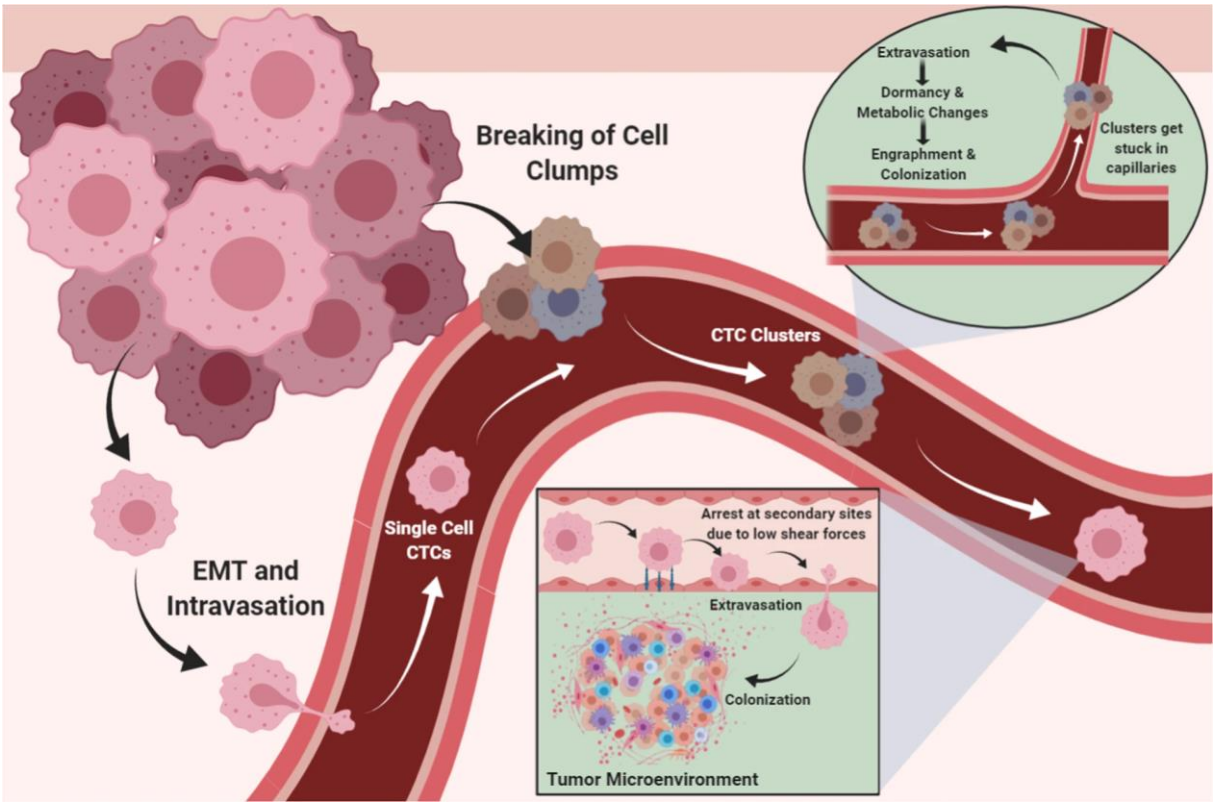

Figure 4. Metastatic process with EMT and CTCs. The CTCs circulate as single CTCs or as CTC clusters, before they get stuck in capillaries and extravasate. The CTCs can change into a dormant state or start colonization. Obtained from Fares et al. 2020 [9].

1.3 Breast cancer

1.3.1 Anatomy of the breast

The breast, medically termed mammary gland, is an organ present in both male and female mammals, but it is only functional in females. The female hormones oestrogen, progesterone and prolactin induce the milk production for breast feeding. The human breast tissue is connected to the ribs by the pectoralis major muscle, but this is not included in the breast anatomy. The breast is made up of a tree-like structure with 15-20 lobules that produce milk, the terminal ductal lobular units (TDLU). These are connected to ducts that transport the milk to the nipples. The anatomy of the breast, as seen in Figure 5, is characterized by lobules and ducts that spread out, with intermingling adipose tissue (fat) and stroma (fibrous tissue, immune cells, vessels and nerves) that make up most of the breast [16]. In the centre of the areolae, the circular dark area of the skin, is the nipple, with the endings of the milk ducts and hundreds of nerves. Stimulation of the nerves in the nipple stimulates the muscles that control the release of milk from the ducts [17].

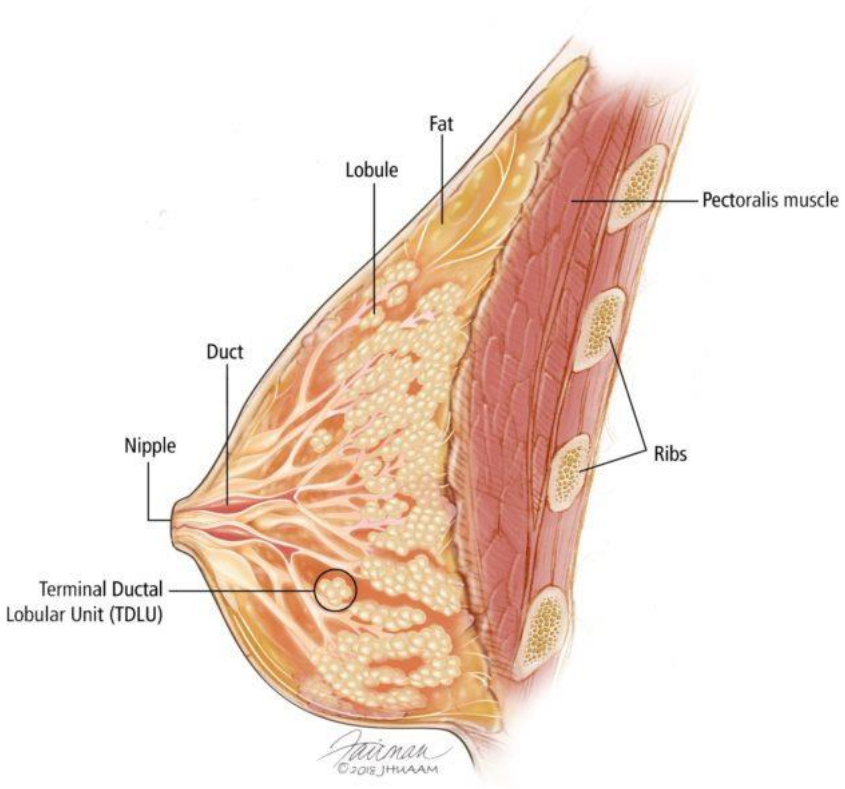

Figure 5. The female breast anatomy. Cross section of a normal human breast. The breast consists of various amounts of stroma, adipose tissue (fat), lobes producing the milk and the ducts that lead the milk to the nipple. Obtained from Johns Hopkins University [7].

The breasts contain blood vessels that circulate throughout the breasts, and lymph vessels that transport excess fluid to the lymphatic system. The lymphatic system also functions as part of the immune system in the body. The lymph vessels connect to lymph nodes under the armpits and in the chest. Epithelial breast cancer originates from cells in the ducts or lobules, probably, and most often from the TDLU areas. Pre-invasive disease has not penetrated the basal membrane (i.e. lobular and ductal carcinoma in situ), but invasive breast cancer and can spread outside the breast via blood and lymph vessels [18]. When breast cancer metastasis via the lymph vessels, it often first involves the tumor-close lymph nodes in the region. These lymph nodes are often referred to as "the sentinel lymph nodes" or simply "the sentinel nodes".

1.3.2 Incidence & epidemiology

Breast cancer is the most common cancer in women in Norway, with 3726 new cases for women were registered in 2019 [19]. Although it is not common, 27 men in Norway were registered with breast cancer in 2019. For women, the incidence rate of breast cancer has doubled since the establishment of the Cancer registry of Norway in 1951. The Norwegian Breast Cancer Screening Program started in 1996 and expanded to be nationwide by 2005. This program invites women between 50 and 69 years to mammography screening, every two years, in order to detect breast cancer at an early stage. Early detection and diagnosis can improve the prognosis, as the cancer can be less advanced. The implementation of the screening program is probably the main reason why an increased incidence of breast cancers among women in Norway is seen; from the period 2010-2014 to 2015-2019, the incidence rate has increased by 7.7% [20]. This may also be connected to better methods for diagnosis, and that an increasing number of women continue mammography screening (outside the national program) when turning 70 years.

Breast cancer has the second highest rank of mortality for women in Norway, with 598 registered deaths in 2019 [21]. The breast cancer mortality has declined in the last years, probably due to a combination of implementation of the screening program, improved diagnostics, and better cancer treatments. The five-year relative survival rate, is the observed survival over a period of five years after time of diagnosis for a patient group, divided by the expected survival of a comparable group in the general population. The groups are compared with respect to key factors affecting the survival, such as gender, age, and calendar year of observation, thus determining the mortality regardless of whether an excess mortality may be

linked to the disease under investigation. For breast cancer in Norway, the five-year relative survival rate has increased the last time-period 2015-2019 compared to 2010-2014, by 1,3%. The graph in Figure 6 represent the last time-period 2015-2019, and shows that the relative survival for females with breast cancer, is at 92% after 5 years. This means that 92% of the women diagnosed with breast cancer are still alive five years after diagnosis. The graph further illustrate that breast cancer patients have a continued reduced survival even after surviving the first 5 years after diagnosis [22].

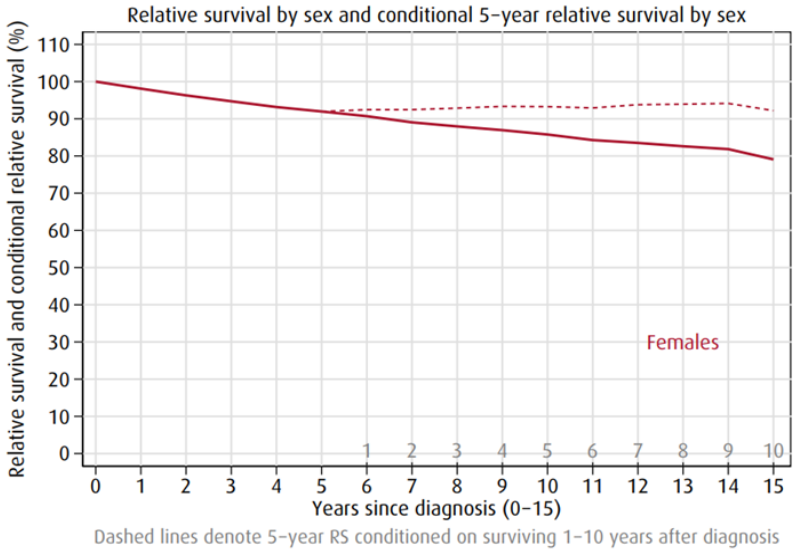

Figure 6. Graphs presenting the relative survival, from the report "Cancer in Norway 2019". The graph shows the relative survival for females from the time of diagnosis, the dotted line shows the conditional 5-year relative survival, conditioned on surviving 1-10 years after diagnosis [20].

1.3.3 Risk factors

There are both environmental and genetic risk factors associated with breast cancer. The environmental risk factors are often connected to hormonal effects, for example during pregnancy and breast feeding, and a high number of menstrual cycles. Women's breast cells will continuously grow and change by the changing levels of the female hormones progesterone and estrogen. Large studies that compare breast cancer occurrence in women that have given birth and breast fed babies, and women that have not, show a far lower risk of breast cancer for women who have undergone pregnancy and breast fed [22]. Other factors include unhealthy diet, low physical activity, healthy weight, alcohol consumption and smoking [23].

The Genome Wide Association studies (GWAS) have identified 93 genes that can be associated with breast cancer [24]. The conventional GWAS studies use panels searching for single nucleotide polymorphisms (SNPs), which is the change in a single base, in the genes associated with breast cancer [24]. SNPs inferring an increased risk for disease will generally have a low frequency in the population because if the mutation causes a lower fitness, natural selection will eliminate it. Only neutral or nearly neutral mutations will be able to accumulate in the genome over generations. For cancer genes, most SNPs are relatively common, but they have a low penetrance in terms of causing disease. Penetrance means the ability of the mutations in the genes to be expressed as a phenotype, causing disease. The balance between penetrance and occurrence in the population describes the relationship of how lethal the genetic mutations are. The SNPs/genes with the highest risk and high penetrance, such as breast cancer genes *BRCA1* and *BRCA2*, have a lower frequency. The SNPs/genes with a low risk and low penetrance, such as fibroblast growth factor receptor 2 (FGFR2) have a higher frequency [24].

Mutations in *BRCA1* and *BRCA2* genes are associated with 5-10% of breast cancers. *BRCA1* and *BRCA2* are breast cancer genes that code for proteins important in chromosome stability, and take part in the repair of DNA double strand breaks [24]. There are always two copies of each of these two genes (alleles), one inherited from each parent. When these genes have certain mutations, causing a disruption of the protein, cells can grow uncontrollably and turn into cancer cells. This only happens if none of the *BRCA1* or *BRCA2* alleles are functional, by mutation of one allele and the other allele of the gene is lost or changed in the cell [25].

1.3.4 Breast cancer initiation and progression

The cancer is believed to have originated as a normal epithelial cell that has undergone multiple genetic and epigenetic alterations that transforms it into a breast cancer cell. Further tumorigenesis is driven by clonal expansion and selection, combined with addition of accumulated genetic changes in the cells [26]. Breast cancer is detected clinically by changes is the breast, as lumps or visual skin changes, or by mammography screening [22].

Studies have shown that the microenvironment in the breast, including the adipose tissue, stroma and the extracellular matrix (ECM) molecules, modulate the tissue specificity of the normal breast, and also influence growth, polarity, survival and invasiveness of breast cancer cells [26]. The epithelial-mesenchymal interactions in the microenvironment are important for the normal development of the mammary gland, but cancer cells can undergo EMT (1.2.3

Tumor heterogeneity and tumor evolution) and acquire the capacity to migrate and invade, separating from the primary tumor and enter the circulation as CTCs. The process of MET then enables the tumor cells to colonize in distant organs [15]. In breast cancer, the invasion starts in the blood vessels or lymphatic system, with metastasis in the regional lymph nodes, and/or in distant organs frequently including bone, liver, lung and brain [27].

1.3.5 Classification and molecular markers

Breast cancer originating from the breast epithelium, is a heterogenous group of tumors, both biologically and molecularly [28]. It is important to distinguish the subtype of each case as they have different prognoses and treatment implications [28]. Breast cancer is classified using many aspects, based on the morphology (i.e., histopathological/microscopic examination) including type and histological grade, by molecular features (protein and gene expression) and by clinical parameters (i.e., stage). Classification is important for diagnosis, prognosis, and prediction of the disease.

1.3.5.1 Histopathological classification

The *type* of breast cancer is defined by its cell morphology, invasiveness, and molecular markers. There are two main types of non-invasive or preinvasive carcinoma in the breasts: ductal carcinoma *in situ* (DCIS) and lobular carcinoma *in situ* (LCIS). LCIS is abnormal cell growth in the lobules, and is less common than DCIS, which is cancer situated in the ducts of the breast that has not invaded into the surrounding tissue [29]. The invasive breast cancer type is cancer that has invaded into the surrounding breast tissue, with the two most common types being *invasive ductal carcinoma* (IDC) and *invasive lobular carcinoma* (ILC), where 70-80% of all breast cancers are invasive ductal carcinomas [30].

Histological grading is a system that can be used on invasive carcinomas with an assessment of how the cancer cells look compared to normal cells. There are several systems used for grading. One of them is the Nottingham Histologic Score system (also called “the Elston-Ellis modification of Scarff-Bloom-Richardson grading system”). A pathologist will study the histology of the cells in a tumor, from a thin slice of formalin fixated paraffin embedded tissue.

The following factors are considered:

1. The amount of gland formation: how differentiating the cancer cells are, recreating glands like normal cells.

2. The nuclear features and degree of pleomorphism: how similar the tumor cells are to normal cells, regarding shape and nuclear size.

3. The mitotic activity: how much the tumor cells are proliferating or dividing.

Each factor is given a score between 1 and 3, and the total score from all the factors result in a tumor *grade*. Grade I tumors have a total score of 3-5, grade II tumors have a total score of 6-7 and grade III tumors have a total score of 8-9 [31].

1.3.5.2 Molecular markers and classification

The molecular markers implemented in clinical practice world-wide, are the estrogen receptor (ER), progesterone receptor (PR), the human epidermal growth factor receptor 2 (HER2) and markers of proliferation. ER and PR are hormone receptors found in breast cancers that respond to hormone signals, by stimulation of cell growth. The cancer is positive for ER and PR if the cancer cells express these receptor proteins. The same goes for HER2, which normally is a receptor that controls the normal growth, division, and repair of healthy breast cells. In 30% of breast cancers the *HER2*-gene is not functioning correctly. Amplification of the gene, resulting in many gene copies, stimulates breast cells to make too many HER2 receptors, which in turn leads to uncontrolled growth and division of the breast cells [32].

The Ki-67 protein in humans, encoded by the *MK167* gene, is a nuclear protein expressed in proliferating cells, and not in resting cells. Therefore, it is used as a marker for proliferation and a status in the cancer cells. Along with ER, PR and HER2, the Ki-67 status can be used as a molecular marker. The biomarkers can be prognostic, predictive or both. Presence of the hormone receptors is a weak prognostic marker, but a strong predictive biomarker, as the patient most likely is predicted to benefit endocrine directed therapy, such as tamoxifen. For patients with tumors with increased HER2 activity (HER2+ tumors), this can be a prognostic and a predictive biomarker, as HER2 expression is associated with poor prognosis and high risk of recurrence (ROR). The patient has increased chance for responding on Anthracycline and Taxane-based chemotherapies and therapies that target the HER2 protein (such as Trastuzumab), but the patient will often not respond well to endocrine-therapies [32]. Breast

cancer classification and treatment recommendations therefore group the tumors by these markers into four categories: 1) ER+ and/or PR+, HER2-, 2) ER+ and/or PR+, HER2+, 3) ER-, PR-, HER2+ and 4) ER-, PR- and HER2- [33].

In 2001 Perou, Sørlie and colleagues studied gene expression profiles, by using microarray technology to measure the transcription levels of genes before and after cancer treatment. Hierarchical clustering of the results identified five intrinsic subtypes of breast cancer, i.e., subsets of tumor with particular patterns of gene activities. The subtypes are called luminal A, luminal B, HER2-enriched, basal-like/triple-negative and normal-like [34]. The subtypes have overlapping features with the groups identified by the biomarkers ER, PR and HER2 is illustrated in Table 1.

Table 1. Expression profiles for the five intrinsic breast cancer subtypes. Each subtype is defined as +/- for estrogen-receptor (ER), progesterone-receptor (PR) and the human epithelial growth factor receptor 2 (HER2) [35].

	ER	PR	HER2
Luminal A	+	+	-
Luminal B	+	+	+/-
Basal-like	-	-	-
HER2-enriched	-	-	+
Normal-like	+	+	-

For the luminal subtypes, markers of luminal epithelial layer are expressed, and they are hormone-receptor positive (ER+ and/or PR+). Luminal A tumors are often low-grade, tend to grow slowly, and have the best prognosis. Luminal B tumors grow slightly faster and tend to have a worse prognosis than Luminal A breast cancers. The basal-like subtype is also called triple-negative, as it is most often negative for estrogen-receptor, progesterone-receptor and HER2. This subtype is more common in younger women and women with the mutated *BRCA1* gene mutation. The HER2-enriched subtype tends to grow faster than the luminal types and have a worse prognosis. However, new targeted therapies against the HER2 protein have shown to be successful. The normal-like subtype is also, ER+, HER2-, but it has a slightly worse prognosis than Luminal A breast cancer [35]. In the later years, a more recent sixth subtype, the Claudin-low subtype/phenotype, has been portrayed. It has been found to be a complex additional phenotype that may permeate breast tumors of various intrinsic subtypes. The

claudin-low phenotype has low genomic instability, proliferation levels and mutational burden, and the immune and stromal cell infiltration levels are high [36].

To improve the intrinsic subtyping for implementation into diagnostics, Parker et al. made a test called Prediction Analysis of Microarray, with hierarchal cluster analysis of gene expression profiling of 50 genes, the "PAM50 Prosigna®" gene signature. The PAM50-subtype classifier and risk model run on the NanoString nCounter Dx Analysis system, classifies breast tumors into the four intrinsic subtypes defined as Lum-A, Lum-B, HER2-enriched and basal-like [37]. The subtype normal-like is not included in this assay. When a patients tumor is tested by the assay, the gene expression is compared to the centroids of the PAM50 genes, and the tumor is assigned a subtype based on the highest correlation. The PAM50 gene signature can be used to predict risk of recurrence, and benefit of hormonal therapy and chemotherapy [38]. The ROR score gives a prediction of the probability of metastasis within the next ten years. In Norway, the PAM50 test was in 2019 concluded to be used in the clinic, for patients with hormone sensitive HER2 negative breast cancer, without metastasis in the lymph nodes. The test is now used to decide which patients should have adjuvant chemotherapy after surgery [39].

1.3.5.3 Clinical staging of breast cancer

To define the cancer *stage*, the pathologist stage measures how advanced the patients tumor is, and if it has spread. For breast cancer, the stages range from stage 0 (pre-invasive disease) to stage IV (metastatic disease). Staging is a prognostic factor used to determine the right treatment. The most used system for determining the stage is the TNM-status (tumor, node, metastasis). The TNM-status is based on the primary tumor size, whether the cancer cells have reached the lymph nodes, and whether the cancer has spread in the body with distant metastasis [40]. The features of the TNM-status are assigned scores called the pathologic T stage (T0-4), N stage (N1-3) and M stage (M0-1), which combined gives a final pathology stage (0-IV).

1.3.6 Diagnosis and treatment of breast cancer

Norwegian Breast Cancer Group (NBCG) is the organization in Norway that defines and updates the guidelines for diagnosis and treatment of breast cancer. The organization also focus on new research and conducting clinical trials. Tumors recognized as cancer are classified based on histological type and grade, estimate of tumor size and stage. Lymph nodes examination for metastasis and the status for the molecular markers ER, PR and HER2 are included in the

national guidelines as of today [33]. The diagnosis will be determining the course of treatment, for each patient individually.

The first step for most breast cancers today, is surgical removal of the primary tumor. This is done either with breast conserving technique or a mastectomy, surgical removal of the whole breast. Sentinel lymph node biopsy is done with or without axillary lymph node dissection. Post-operative radiation is advised after breast reconstructive surgery, if the primary tumor is large (T>50mm), and for women with node positive disease. Adjuvant systemic treatment is used based on several factors, including the use of prognostic and predictive markers, depending on age, size, grade, ER/PR and HER2 status and stage. For instance, for women with hormone receptor positive disease, a five-year adjuvant endocrine treatment is offered. For patients with metastatic disease or relapse with distant metastasis after primary treatment, the systemic treatment is more used. Every patient is advised by an oncologist for the best possible treatment in their case [33].

1.4 Liquid biopsy

Liquid biopsy is a sample of blood, urine, bone marrow or other liquids from patients. It is regarded as a non-invasive method to obtain a range of biological information from a patient, and can be measurements of for instance metabolites, proteins, nucleic acids or cells. For patients that need follow-up over time, the use of liquid biopsy is easier and less invasive than needle biopsy [41]. For cancer patients, the study of tumor disease by identifying small fragments of tumor DNA in a blood sample, can indicate relapse of the disease or can be used to monitor treatment response. This method is mainly used in clinical trials, and can be a step to provide more information about tumor heterogeneity and tumor evolution, as not only one tumor is analysed. This can again be used for further study of cancer initiation and progression [42]. With liquid biopsy, it is possible to use sensitive immunological and molecular procedures to detect single tumor cells or micrometastases in for instance peripheral blood (CTCs) and in organs, such as the bone marrow (DTCs). The presence of such minor depositions of tumor cells are frequently referred to as minimal residual disease (MRD) [43].

1.5 Importance of CTCs and DTCs

1.5.1 CTC and DTCs in breast cancer

For breast cancer, the metastatic route is not always via the lymph nodes, but cancer cells can disseminate directly through the blood (haematogenous dissemination) to distant organs. Approximately 20-30% of breast cancer patients, who do not have axillary lymph-node metastases, eventually develop metastases at distant sites [43]. Bone is one of the common distant organs of metastasis in breast cancer, and bone marrow as a reservoir for DTCs have thus been of major interest. The outcome for breast cancer patients with DTCs present in the bone marrow, have been identified to be less favourable and an increased number of DTCs have shown an increased poor prognosis [44, 45]. To study the malignant potential of the DTCs, their molecular characterization must be investigated. Genetic characterization of single DTCs in the bone marrow, compared to the genetic profiles of the primary tumor, supports the theory of an early event of haematogenous dissemination in tumor progression. This supports the parallel progression model, as the single DTCs often show acquirement of additional genetic defects and different properties than the cells in the primary tumor. But there are also studies supporting the linear progression (also called stepwise) model, based on genotypical and phenotypical diversity and heterogeneous cells within the primary tumor. Therefore, the differences between the primary tumor and DTCs may be caused by tumor evolution of the disseminating cells and/or subclones from a heterogenous primary tumor with different disseminating and metastatic potential [46].

There have only been a few studies comparing the presence of DTCs in bone marrow and of CTCs in peripheral blood at the same time point in breast cancer patients [45, 47]. The studies showed higher frequency of DTCs in bone marrow aspirates than CTCs in blood samples from the same patients. This could be because the bone marrow might provide better conditions for survival of DTCs, compared to CTCs that have a short half-life in the circulation system. Blood analysis of CTCs represent a snapshot of ongoing tumor cell dissemination, whereas the DTCs can survive in the bone marrow for a longer time [48]. For metastatic breast cancer patients, both DTCs and CTCs are shown to be independent prognostic factors for prediction of relapse and overall survival. CTC count before and after treatment can provide information about a patients' response to treatment [45, 49]. The molecular analysis of the DTCs in bone marrow is important because it may help predict the need of additional systemic therapy after successful surgery, so-called adjuvant therapy. Better selection criteria and further study of DTCs can help development of more specific and less toxic treatment for each patient [43].

In contrast to many other cancer types, breast cancer patients can experience metastasis more than 10 years after diagnosis and resection of the primary tumor. Cancer dormancy is characterized as presence of MRD years before a clinically detected metastasis. Single DTCs can enter a nonproliferative quiescent state, which is referred to as tumor-cell dormancy [48]. Little is known about how these dormant DTCs become awake and active. Disturbance of the dormant tumor cells with transition into a dynamic state and cell proliferation and subsequent metastasis, could be due to additional genetic and epigenetic modifications of genes controlling proliferation and apoptosis and influence by the surrounding microenvironment with growth and angiogenic factors [48].

1.5.2 Detection of CTCs and DTCs

There have previously been two main approaches to detect CTCs and DTCs, either immunocytochemical staining or molecular assays. The immunocytochemical detection assays use monoclonal antibodies that bind to tumor associated proteins that are expressed by tumor cells, and not expressed by the normal cells, in particular leukocytes. For epithelial tumors, cytokeratin (CK) and/or EpCAM (Epithelial cell adhesion molecule),) are the most common antibodies used for detection of CTCs/ DTCs [50]. Positive selection of CTCs/DTCs can be combined with negative depletion of hematopoietic cells, with for example the common leukocyte antigen CD45 [51]. The number of CTCs in blood and DTCs in bone marrow can be very low, and it is important to detect MRD down to one tumor cell per million normal cells [43]. Therefore, using an initial enrichment step by positive selection, negative selection, or size-based selection, has increased the efficiency of tumor cell detection.

Techniques for identification of CTCs/DTCs includes size-based selection by membrane filter devices, such as ISET (isolation by size of epithelial tumor cells) or MEMS (micro electro-mechanical system) based microfilter. There are advantages and disadvantages to these enrichment steps, for example by clotting of filters or loss of DTCs that do not express the surface antigen [48]. The most frequently applied techniques are immunomagnetic bead techniques (IMT), with the use of specific antibodies to surface proteins. One of the automatic systems with this detection technique, is the CellSearch® system (Menarini Silicon Biosystems). This has been the most advanced and commercially available technology with approval from the U.S. Food and Drug Administration (FDA) for CTC detection in metastatic breast, colon and prostate cancer [48]. CellSearch uses a system with automatic enrichment by

immunomagnetic beads coated with anti-EpCAM drawing out the cells with EpCAM surface proteins followed by staining with cytokeratin. The system detects the CTCs, but a final isolation step of single tumor cells is still needed after this enrichment. Flow cytometry, such as fluorescence-activated cell sorting (FACS), can also be used when large quantities of cells are available. Single CTCs in suspension are differentiated and quantified based on fluorescent surface antibodies, such as anti-EpCAM, anti-CD45 and DNA staining. Gates in the scatter plots of the cell characteristics, including size, graining and marker expression, are used to identify the CTCs [51]. Disadvantages with the automated FACS system is the need of many CTCs in order to set the gates and its high purity mode results in a high cell loss, which is why it cannot be used for rare cell sorting. Examples of other detection techniques are EPISPOT (epithelial immunospot) with depletion of CD45+ cells, CTC-chip with EpCAM-antibody coupled micro posts and laser scanning cytometry MAINTRAC® with red blood cell lysis [48].

In many studies the quantification of CTCs and DTCs has been the aim, as this has prognostic value [51]. The clinical enumeration and quantitation of CTCs/DTCs is a large and important field but not covered by this thesis.

1.6 Single-cell isolation methods

The interest in single cell sequencing have increased in the last decade, with the possibilities of sequencing a single cells' genome or transcriptome, which can provide information about rare cell population differences, heterogeneity, and evolutionary trajectories. Single cell sequencing of CTCs/DTCs can enable the transcriptional features that would be "diluted" by bulk sequencing [52].

1.6.1 Single cell genome analysis

Sequencing of single cell genomes have been used to study subclones in primary tumors, metastases and to some degree CTCs/DTCs. Whole-genome amplification is necessary to provide the quantify of DNA needed for next-generation sequencing of a single tumor cell [53]. Single cell genomics have provided information about copy-number variations, with whole-arm gain or losses and amplifications or deletions [46].

1.6.2 Single cell transcriptome analysis

There are several protocols for RNA sequencing of single cells, but they have in common the conversion of RNA to the first stand of complementary DNA (cDNA) by reverse transcriptase. Sequencing can be by full transcripts or by sequence tags at the 5' or 3' end. The goal is to capture the cells "original" fragments of RNA and with accuracy, amplify it evenly. The efficiency is influenced by small reaction volumes, functioning enzymes, and amplification can be improved by a smaller number of cycles and inhibition of primer by-products by 'suppression PCR' [54]. High-throughput technologies, such as microarray and RNA-sequencing (RNA-seq) have provided a better understanding of the transcriptomes from complex eukaryotes. Some protocols can sequence different cells' transcriptomes at the same time, by pooling barcoded cDNA for enough starting material for linear amplification by *in vitro* transcription. With individual labels on RNA molecules, the absolute number of original molecules can be counted after amplification [54]. The 10X transcriptomics is an example of pooled sequencing with the transcriptome of the cell population average.

Single cell transcriptome analysis can provide the unique cell-to-cell variability that might be "diluted" by bulk sequencing. Several research groups have developed sequencing-based methods for the single cell transcriptome analysis, and the most recent SmartSeq2 protocol by Picelli et. al (2014) [55]. Improvement in reverse transcription (RT), template switching and preamplification increased the cDNA yield from single cells, with a higher sensitivity and less variability [55, 56].

1.7 Importance of isolation of rare single cells

The present technology has opened for analysing a single cells' DNA, RNA, and chromatin state, which again provide information about the genotype and phenotype of the actual cell. Which genes are expressed and what portions of the genome that are active in a cell, defines the single cells phenotype [57]. Characterization of DNA has been successful for DTCs but establishing identification and selection methods compliant with RNA analyses would be of great value. This could be used to determine the phenotype of a DTC and whether it is dormant or not. Knowing more about the activity state of DTCs, combined with the type of genomic alterations, will help us understand the role of DTCs in tumor evolution and how this could result in later metastases. Such knowledge is needed to reveal the value of DTC characterization to monitor MRD in breast cancer patients to be able to add treatment to avoid metastatic disease.

For the selection and isolation of rare single cells, both manual and automatic methods can be used. The clue is picking the right cell, for example the micromanipulation is a precise manual method to target a single cell, but tissues can also be dissociated into cell suspensions [54]. With a cell suspension the automatic enrichment by expressed surface markers and isolation of the rare tumor cells can be done using different strategies, such as laser capture micro dissection, ISET, DEPArray, MagSweeper, Rare cyte or flow cytometry sorting [53, 58].

A previous study has tested how low levels of RNA can be due to degradation and how the degradation could be affected by time and temperature. They found that the RNA could be degraded at different rates at different transcripts, which could lead to a possible bias when measuring the expression levels [59].

For better understanding of the transcriptome of dormant DTCs in the bone marrow of breast cancer patients, the method for selection and isolation of single cells must improve with better RNA quality and quantity. With more sequencing of the genome and transcriptome of these cells, the hope is to reduce the number of relapses of metastatic breast cancer. Although the process of enrichment, cell detection and cell isolation can be viewed as separate steps, they are not independent of each other.

2. Aim of study

The aim of this master thesis was to establish methodology for identification and selection of single tumor cells preserving RNA suitable for sequencing analysis.

Null hypothesis: RNA amount and quality from single tumor cells are the same, regardless of identification and selection method.

Objectives:

1) Select and establish several workflows for identification and selection of epithelial tumor cells
2) Establish single cell RNA conversion, amplification and quality assessment
   The RNA amount
   The RNA quality
   The amplificability of RNA
3) Structured testing of methodology
4) Evaluation of methodology performance

3. Materials

A complete list of the reagents and equipment used in this master thesis is listed in appendix 3.

3.1. Cell lines

Cell line HCC38 (ATCC® CRL-2314™) was used as the main cell line for testing the isolation methods. The cells are from female mammary gland by a 50-year-old patient with primary ductal carcinoma. The receptor expression is ER-, PR- and HER2-, and known gene expression of epithelial glycoprotein 2 (EGP2) and cytokeratin 19 [60].

Cell line HCC2218BL (ATCC® CRL-2363™) was used for preliminary testing of the methods and as a positive control for gene expression of CD45. This cell line is initiated from peripheral blood lymphocytes by transformation with Epstein-Barr virus (EBV), from a 38-year-old female [61]. As a peripheral blood mononucleated cell (PBMC) this cell line should have genes expression CD45 (lymphocyte common antigen). CD45 is expressed on the surface of all leucocytes as a receptor-linked protein tyrosine phosphatase [62].

3.2 Patient samples

Two bone marrow samples from patients in the Neotax study was selected for single cell isolation. Both patients had donated bone marrow for DTC analysis as part of the clinical trial.

The neoadjuvant (Neotax) clinical trial enrolled breast cancer patients from 1997 to 2003 in Norway. The patients were randomly allocated into two treatment arm of paclitaxel or epirubicin, with crossover between treatment arms if the patient had no response [63]. The two selected patients had multiple DTCs detected in bone marrow aspirates as part of the study (820/2 million cells for #1 and 765/2 million cells for #2) and was used to study the applicability of identification and selection by FACS on aspirates preserved for years (suspensions frozen in liquid nitrogen tanks). They patients' identification have been anonymous as a study ID is the only information available (de-identification, code list accessible only for the study principal investigator). Since this master thesis is under the LATE project, the project name for the cells were "LATE".

4. Methods

The laboratory procedures used in this master thesis are described in this chapter. Procedures before and after the sample handling in this master thesis are also briefly described. The flow of the methodology testing is visualised in Figure 7. For the methodology testing, cell lines were used.

Figure 7. Flowchart from cell suspension to the goal of analysing cDNA from single cells.

Four different methods for isolation of single cells were selected for this project:

- Suspension and isolation by microinjection pipette
- Cytospin and isolation by micromanipulator.
- Automatic cell sorting by FACS.
- Enrichment by Cellsearch and automatic cell sorting by DEPArray.

As shown in Figure 8, there are several fundamental differences between these four methods, and each will be described in detail in separate chapters below.

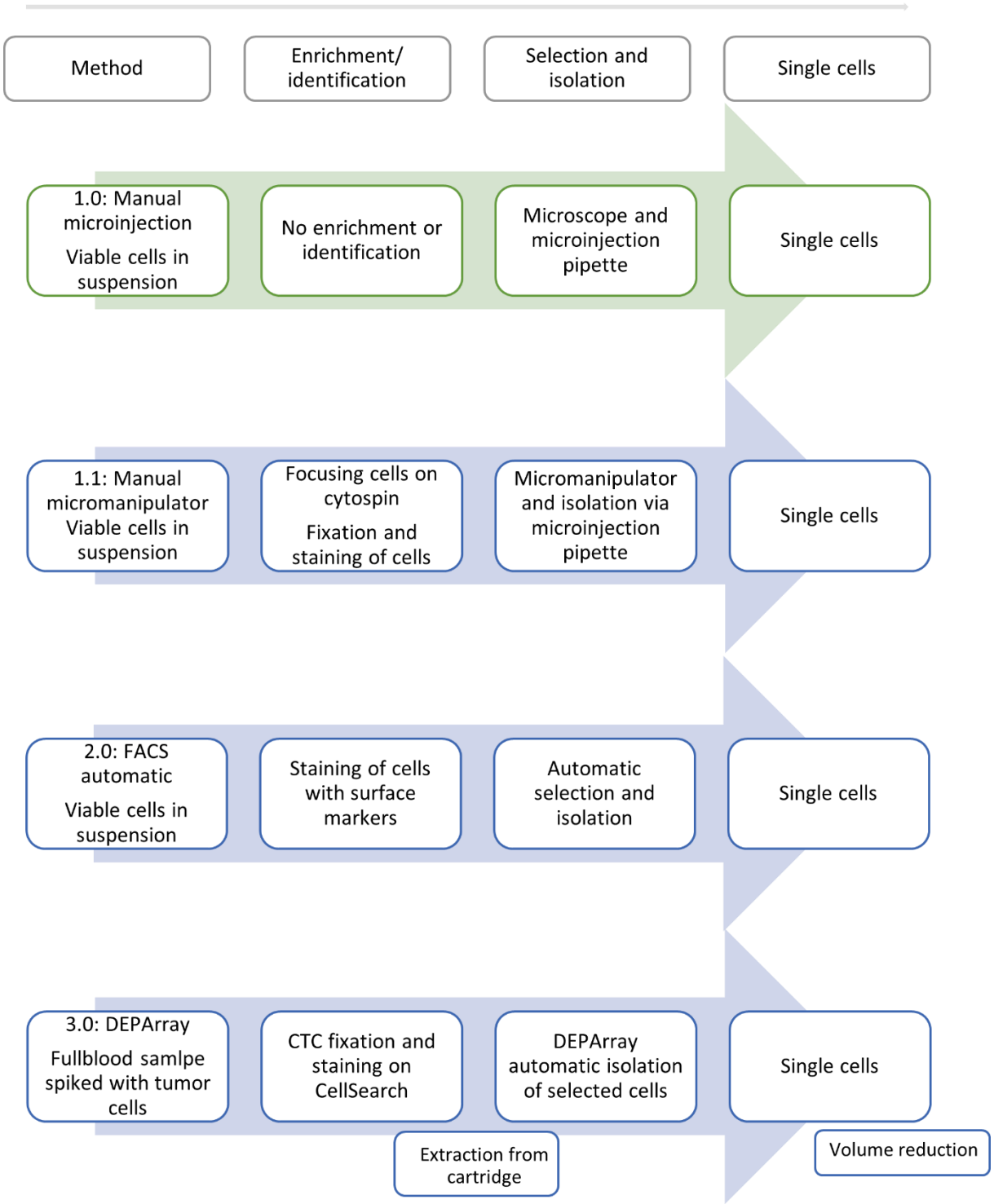

Figure 8. Flowchart of methods for single cell identification and isolation. The manual method with the microinjection pipette is shown in green because this was the reference method. The flow chart shows the identification and selection steps with their differences for the methods. The final step for all the methods leads to the single cells, but for the DEPArray method, the cells need an extra step of volume reduction after isolation.

4.1 Identification of single tumor cells

For a robust comparison of the methods, a high number of single tumor cells must be analysed. A hierarchy ID numbering system was made to avoid confusion or mislabelling. Making the system for naming and numbering in advance made it easier to trace each single cell throughout the project.

Every cell got a unique identifier with four slots for information to easily trace which process each cell belong to:

"COMP_Method number_Cell line_number of cell"

- **COMP**: project name (i.e. "comparison of the methods").
- **Method number**: the four isolation methods were:
  1.0: manual isolation of cells from suspension with microinjection pipette.
  1.1: manual isolation of cells from a glass slide with micromanipulator.
  2.0: automated sorting by FACS, i.e. cells from suspension stained with anti-EpCAM-PE, anti-CD45-APC and Hoechst 33258.
  3.0: DEPArray isolated cells (after enrichment by CellSearch).
- **Cell line**: HCC38.
- **Number of cells**: each cell was numbered, to keep a system and trace each cell during the process.

4.2 Cell culturing

In the process of cell culturing, cells are grown *in vitro* under controlled conditions. Cells have been removed from living tissue and then cultured under artificial conditions [64]. The cell lines used were breast epithelial cell lines HCC38 (ATCC) and peripheral blood lymphocytes line HCC2218BL (ATCC). Cell line HCC38 (ATCC® CRL-2314™) was used as the main cell line for testing the isolation methods. Cell line HCC2218BL (ATCC® CRL-2363™) was used for preliminary testing of the methods and as a positive control for gene expression of CD45. More information about the cell lines can be found in 3.1. Cell lines.

4.2.1 Aseptic technique in the cell lab

Aseptic technique is a set of procedures used to prevent contamination. The technique is used to create a barrier between the sterile cell culture and possible microorganisms in the environment. The strict rules and procedures of aseptic technique must be used continuously when working with cell cultures [65].

Procedure:

Experiments were performed in a sterile biosafety cabinet (Biowizard®) in a lab for cell culture work only. Before and after entering the lab, hands were washed, and a lab coat and clean lab gloves were used. The biosafety cabinet was sterilized with 70% ethanol and the air flow was always kept running during use. The pipette tips and cell culturing flasks were sterile. The equipment was washed with 70% ethanol before and after entering the biosafety cabinet. The hood of reagent bottles and culture flasks were not kept open for longer than necessary [65].

4.2.2 Cell culturing and passaging

During cell culturing, cell passage is a procedure of splitting the cells when they are 70-80% confluent. This is a technique that ensures the cultured cell lines to be growing and alive under the cultured conditions. Using two morphologically different cell lines, the growing conditions had to be assessed individually for each cell line. Cell passaging was necessary to prevent overgrowth, that could lead to reduced ability of cell division and possible mutation within the cell culture.

Procedure:

The cell lines were thawed from liquid nitrogen freezer and cultured in selected cell culture medium and sterile flasks appropriate for the different cell lines (Table 2). The flasks with the cell lines were incubated at 37°C in a 5% CO_2 humidity incubator (Nuaire). The cells were split routinely every 2-3 days or at 70-80% confluency, to prevent overgrowth. 70-80% confluency is when the cells cover roughly 70-80% of the flask bottom [66]. Culture conditions varies between the different cell types, with different flasks and culture medium.

Table 2. Cell culture conditions. The different cell lines used in this master thesis with different tissue of origin, flask type and specific culture medium. RPMI – Gibco Roswell Park Memorial Institute [67]. Foetal bovine serum (FBS).

Cell line	Tissue of origin	Flask type	Culture medium
HCC38	Human female breast epithelial cell line	Corning® flask, T25 Nunc™ EasYFlask™, T75	RPMI-1640 + Glutamax + 10% FBS
HCC2218BL	Human female immortalized lymphoblast cell line	TC-Flask T25 untreated.	RPMI-1640 + Glutamax + 10% FBS

The procedure for splitting of cells vary with respect to the need to use trypsin (Invitrogen) to get the cells to detach from the bottom of the flask. HCC2218BL is non-adherent and grow in suspension in untreated flasks and did not need trypsin. HCC38 are adherent and grown in polystyrene flasks treated with proprietary Nunclon™ Delta surface treatment. These cells adhere to the surface of the flask and trypsin treatment was needed to make them detach. The morphology of the cell lines grown in flasks can be seen in Figure 9. Before adding trypsin, the supernatant was removed, and the cell layer was rinsed with PBS (Gibco). A wash with PBS is necessary to remove dead cells and medium with FBS since FBS inhibit trypsin. 1-1,5 mL trypsin was added, and the flask was set for incubation in an incubator for 4-5 minutes. The cells were then loose in suspension. 6mL fresh culture medium was added to the flask to inactivate the trypsin. Both the adherent and non-adherent cell lines were transferred to 15 mL tubes for centrifugation at 800RPM for 8 minutes. The supernatant was removed without disturbing the cell pellet. New medium was added, and the cell pellet resuspended. The new cell suspension would be counted to ensure a split ratio for good growth. HCC38 was routinely diluted 1:2 in fresh culture medium to a new T75 flask. HCC2218BL was diluted to ensure between 500K and 1.0 x 10^6 cells per ml to new T25 flasks, as this gave the best growth conditions.

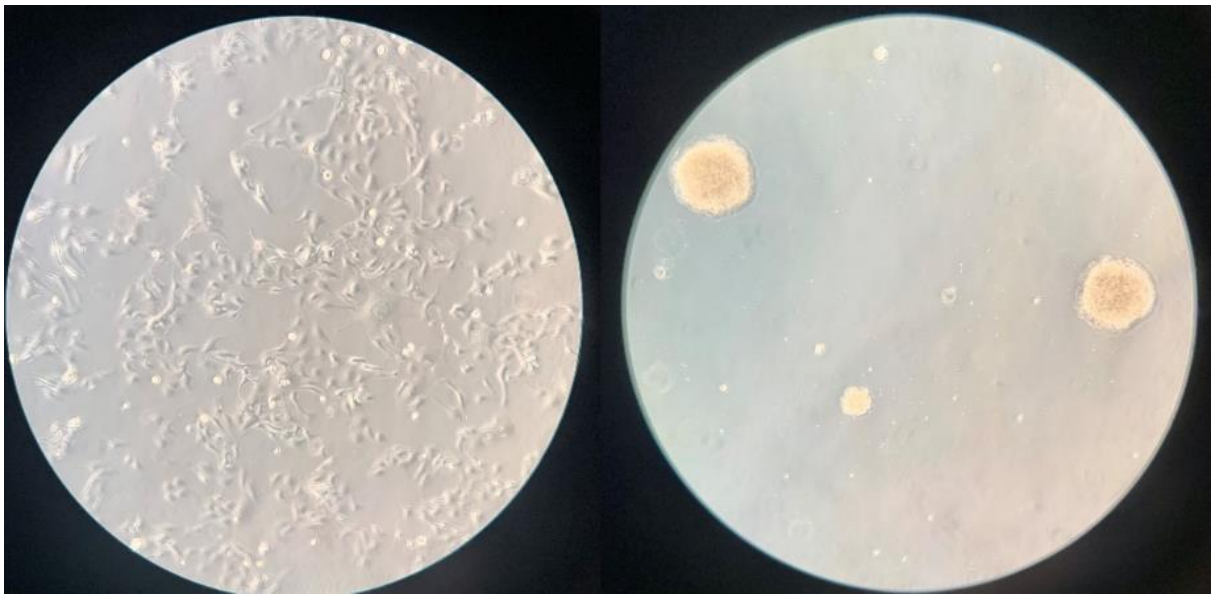

Figure 9. Cell line morphology. From the left HCC38 and HCC2218BL. HCC38 cells are adherent cells that form loosely attached clusters, while HCC2218BL are cells in suspension that grow in multicellular aggregates [60, 61]. Photo taken of the cells via microscope when grown in flasks.

4.2.3 Cell count

The cells were routinely counted with Countess™ (Thermo Fisher Scientific) automated cell counter during cell passaging to check cell count and viability of the cells [68]. Countess™ automated cell counter was also used to count the cells before staining with anti-EpCAM and anti-CD45 in the FACS method.

Procedure:

When the cell pellet was resuspended in fresh culture medium during cell passaging, 10µL cell suspension was mixed with 10µL trypan blue stain 0.4%. 10µL of the mix were disposed by pipet onto a disposable Countess chamber slide. The slide was inserted into the Countess™ II FL instrument and the cell count and viability of the cells were displayed [68].

4.3 Tumor cell identification and single cell isolation

Four methods of tumor cell identification and single cells isolation were tested, two manual methods and two more automated methods. The methods used different identification and selection procedures, which could influence the RNA quality and amount of output.

4.3.1 Microinjection pipette (method 1.0)

For this method, the cells were viable and without any staining when a microinjection pipette (Origio Benelux) was used to isolate single cells. The aim with this method was to keep a rapid processing time with near-to physiological conditions. The cells would be of good quality, with as little interference or manipulation as possible, and as the microinjection pipette method did not include any identification and selection steps, it was used as a reference for the other methods.

Procedure:

For the microinjection pipette, cell suspensions were used. After resuspension in fresh medium, the cells were put on ice until the microinjection set-up was ready. The microinjection pipette and the Leica microscope model TL3000 Ergo was set up in a separate lab, as seen in Figure 10. A prepared solution of 0.5g Polyvinylpyrrolidone (PVP) + 50mL PBS solution was made in advance and mixed to a homogenous solution. A Nunclon™ Delta Surface petri dish was placed on the microscope centre and drops of the PVP + PBS solution of approximately 4µL were placed in a half circle on the petri dish. One drop was used to rinse the capillary tip of the STRIPPER micropipette. The HCC38 cell suspension was resuspended before 1-2µL was added to one of the drops on the petri dish. The cells could be visualized using the microscope. By using the micropipette, individual single cells were aspirated and deposited separately in wells of an 8-well PCR-strip (Eppendorf) containing 10µL RLT buffer (Qiagen). The capillary was rinsed in one of the clean drops, and the drop was inspected to check that the cell was not deposited there. That served as a check that the cell was deposited correctly in the well. To keep a shorter time before the cells isolated were put on dry ice, the 8-well strip was split in two 4-well strips, and as soon as all four wells contained a single cell, it was spun down and frozen on the dry ice.

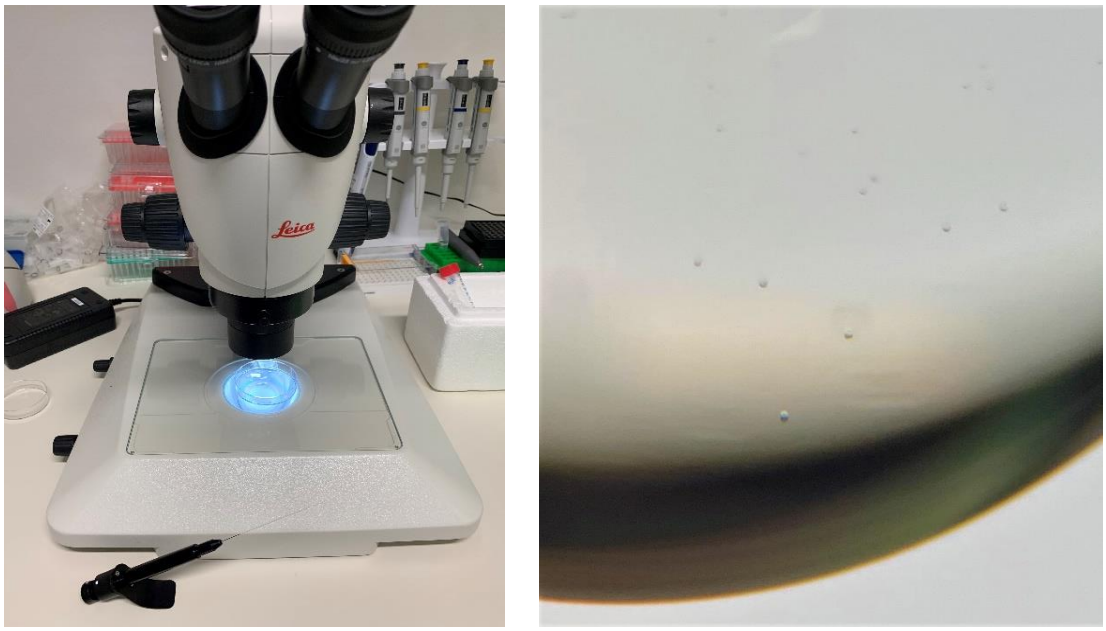

Figure 10. Method 1.0 with Microinjection-pipette. Left: example of set-up for the microinjection method, with the microinjection pipette, the stereo microscope, and the petri dish with the cells in drops of PBS + PVP. **Right:** Seen in the microscope, a droplet with cells. The use of contrast helps visualizing the unstained cells.

4.3.2 Micromanipulation (method 1.1)

For this method non-viable single cells were picked by a glass capillary and a microscope, micromanipulator (Eppendorf). Here, cells in suspension were fixed on glass slides, by a process called cytospin. The samples were fixated and stained by cytokeratin antibodies binding to epithelial cells by alkaline phosphatase reactions. The epithelial tumor cells could then be visualized in the microscope and isolated by the micromanipulator.

Procedure:

The cytospins were prepared by Section for experimental pathology, department of pathology OUS, as follows: A cell suspension from HCC38 was diluted to $1x10^6$ cells/ml PBS with 1% FBS. The Superfrost Plus object slides and filters was set in the cytospin holder (brand) and 0,5ml cell suspension ($5x10^5$ cells/slide) was added to the well. The slides were centrifuged in a cytocentrifuge (Hettich Universal) for 3 minutes at 120g. Most of the supernatant was removed and the slides were centrifuged for 1 minute with 1100g. The remaining fluid was then absorbed by the filter. The cytospins were airdried and wrapped in aluminium foil and frozen in -80℃ for storage, as seen in Figure 11.

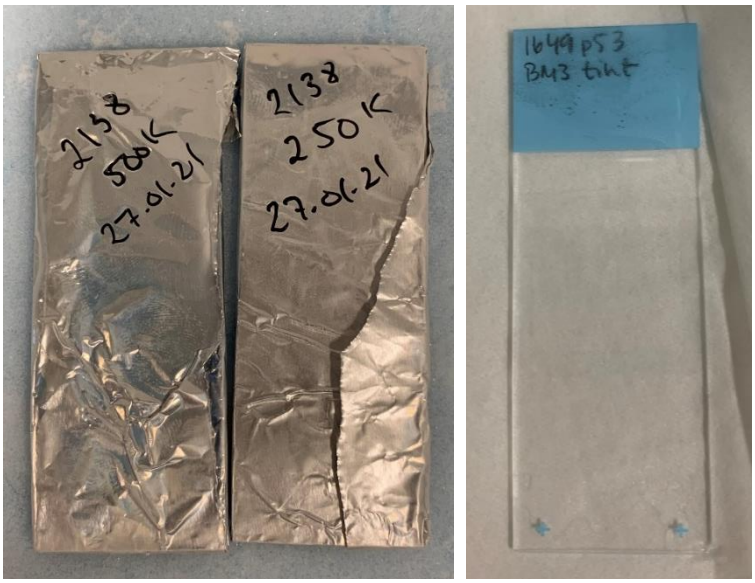

Figure 11. Wrapped and unwrapped cytospins for illustration. Patient samples.

Staining protocol (developed by Section for experimental pathology, dept. of pathology): The cytospins were thawed for 15-20 minutes at room temperature before start, then fixated in acetone for 10 minutes, before a PAP-pen was used to draw a barrier around the drop of cells. A 1:100 dilution of pancreatin AE-1/AE-3 antibody (Nordic BioSite) and PBS/w 1% BSA was made and 100µL of the 1:100 antibody solution was added to each cytospin and incubated in a humid chamber for 30 minutes. A washing buffer with 1 part 0.05M Tris-HCl-buffer (pH 7.4) and 4 parts 0.9% NaCl was then used to wash the cytospins twice for five minutes, new buffer for the second wash. Then a 1:500 dilution of 1 µL Streptavidin Alkaline Streptavidine Phosphatase (S-ALP, Vector Laboratories) + 499 µL of 10mM Hepes (0,15M NaCl + 0,1mg BSA/mL) was made and 100 µL was added to each cytospin and incubated in a humid chamber for 30 minutes. The washing procedure was repeated. The substrate was made in a fume hood with a 1:10 ratio of Levamisole Endogenous Alkaline Phosphatase Inhibitor (Dako) and BCIP/NBT (Dako). The cytospins were incubated with 30µL of the substrate in a humid chamber for five minutes. The last step was another wash procedure with distilled water for five minutes twice. The slides were then stored in PBS at 4°C.

The antibody AE-1/AE-3 binds to the cytokeratin AE-1/AE-3, which is in the cytoskeleton in the cytoplasm of epithelial cells. Then Levamisole endogenous alkaline phosphatase inhibitor and the ready to use substrate BCIP/NBT, will react with the bound antibodies and create a dark purple/blue colour. The time of this process decides how dark the cell staining should be, in this

protocol only five minutes in order to not over-stain the cells [69]. The cells should then be isolated by micromanipulation as soon after the staining process as possible to ensure good quality of the cells. In this master thesis the isolation was done the same day and within the following day.

The micromanipulator used was the Transferman NK2 that consists of a control board (joystick), module unit and a power supply along with a microscope (Zeiss Axiovert 40 C) with a digital microscope camera (AxioCam ICc1 rev.3), for images of the cells (Figure 12). When the cytospin was on the microscope, the cells were covered in a layer of PVP + PBS solution, to ensure they did not dry out, and to have liquid for cell transfer. To transfer the cells into the capillary, a manual oil piston pump (CellTram Oil/Vario Pump) was used to get suction. The glass capillaries were 40µm bevelled with a diagonal opening to scrape the cells off the cytospin and then use the piston pump to vacuum them with the PVP + PBS into the capillary. The cell would then be deposited in a droplet of PVP + PBS on a separate petri dish. The petri dish was then moved to the stereo microscope and the STRIPPER micropipette was used to transfer the single cell to a strip with 10µL RLT buffer. The cells in this method were stained and easier to observe in the microscope, compared to the method with the microinjection pipette and unstained cells in suspension.

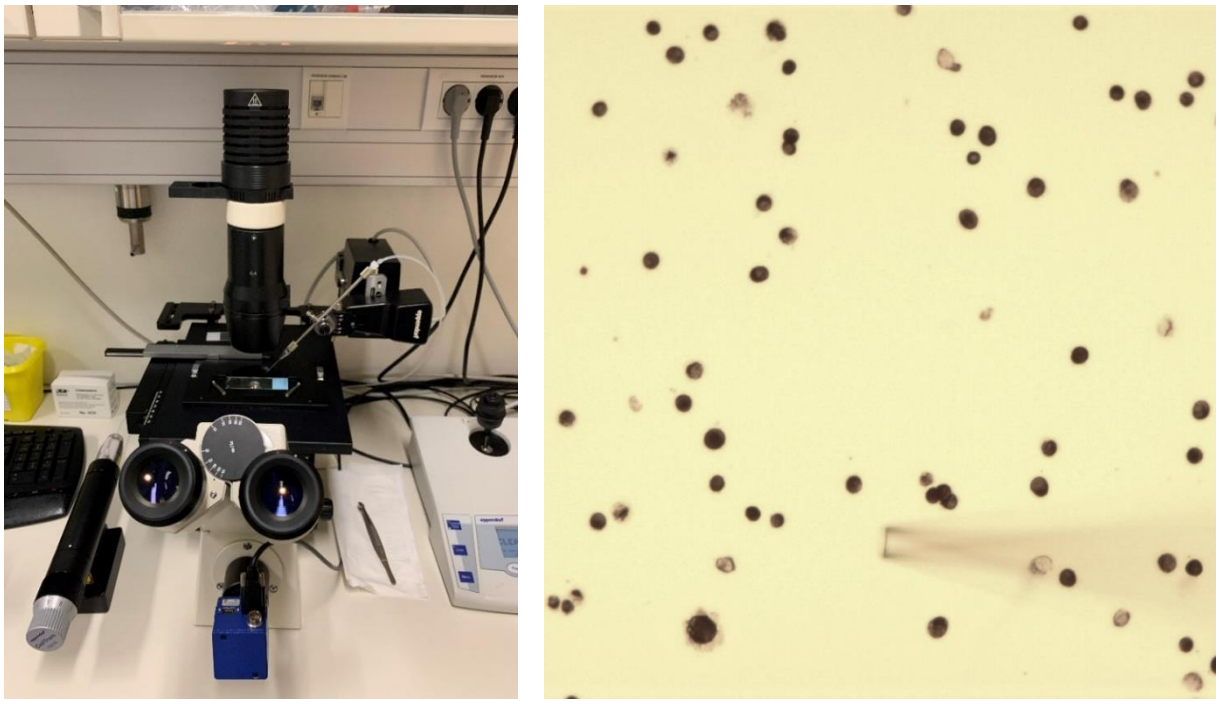

Figure 12. Method 1.1 with micromanipulator. Left: Micromanipulator set-up with a cytospin. **Right:** Stained HCC38 cells as seen in the microscope, with a shadow of the capillary.

4.3.3 Fluorescence-activated cell sorting (FACS, method 2.0)

The fluorescence-activated cell sorting (FACS) Aria II instrument was used to automatically identify and select single-cells from suspension to a 96-well plate (Thermo Fisher Scientific). All FACS experiments were carried out at the Flow cytometry core facility, Dept. of core facilities, Institute for Cancer Research. The FACS Aria II uses a specialized flow cytometry that can sort out single cells from a heterogenous suspension. The instrument uses fluorescence to separate the cells in suspension by their antibodies specific for the cell surface protein, the cell size, and the graining of the cell. In this master thesis the anti-CD45-APC (MACS), anti-Ep-CAM-PE (Santa Cruz Biotechnology) antibodies, and Hoechst 33258 (Thermo Fisher Scientific) were used to stain the cells in the suspension. The anti-CD45 would bind to the surface markers on white blood cells, and the anti-EpCAM would bind to surface markers on the tumor cells. Hoechst 33258 is a blue, fluorescent nucleic acid stain that bind to dsDNA, and was used as a live/dead marker, to stain the dead cells. For APC the FACS Aria II uses a red laser with a 670/14nm bandpass filter and for PE a yellow laser with 585/15nm bandpass filter and 570nm long pass dichroic filter [70]. This method was tested by several steps to ensure that the staining of the cells and the sorting would not degrade the quality of the tumor cells. The staining was first tested with the Countess instrument, to ensure that the cells were binding to the antibody staining.

Procedure:

Suspension of cells from the HCC38 cell line was stained with ani-CD45-APC, anti-EpCAM-PE and Hoechst 33258 before sorting with FACS. A suspension of a mix of HCC38 and HCC2218BL was only used preliminary to test the staining and sorting.

Staining protocol with antibodies: The cell suspension containing 1 million cells was transferred to a 5mL polystyrene round-bottom tube with cell-strainer cap (FALCON®), hereafter referred to as FACS tube. The cells were centrifuged at 540G for 5 minutes using a swing out – bucket centrifuge. The supernatant was removed, and the cell pellet was resuspended in 1mL 4°C PBS + 0,5% BSA. The cells were centrifuged at 540G for 5 minutes. The supernatant was removed, and the cell pellet was resuspended in 300µL cold PBS + 0,5% BSA. The cell suspension was filtrated through the cell-strainer cap to ensure a single cell suspension. 2µL anti-EpCAM-PE and 5µL anti-CD45-APC per 1 million cells was added to the cell suspension and set for 15 minutes incubation at 4°C in the dark. After incubation 1ml cold PBS + 0,5% BSA was added before centrifugation at 540G for 5 minutes. The supernatant was removed, and the cell pellet

was resuspended in 1mL 4℃ RMPI 1640 culture medium. 2µL of the Hoechst 33258 stain was added at the flow cytometry core facility and the cells were kept on ice until sorting.

The sorting with the FACS Aria II instrument was done by the flow cytometry core facility as follows: a suspension of unstained cells of HCC38 was first used as a control to set a threshold for the voltage from the detector to rule out background noise. The same settings could then be used on the stained cells. As seen in Figure 13, the cell suspension was sent through a narrow flow stream that separates the cells and pass them through a fluorescence measuring station, where the fluorescent character of each cell is measured. An electrical charging point and vibration breaks the cells into droplets. Each droplet containing a cell is charged based on their fluorescent character, and when the stream of droplets are passing through an electrostatic deflection system, the cells of the desired characteristics are separated from the stream and deported into a well on a 96 well plate [71].

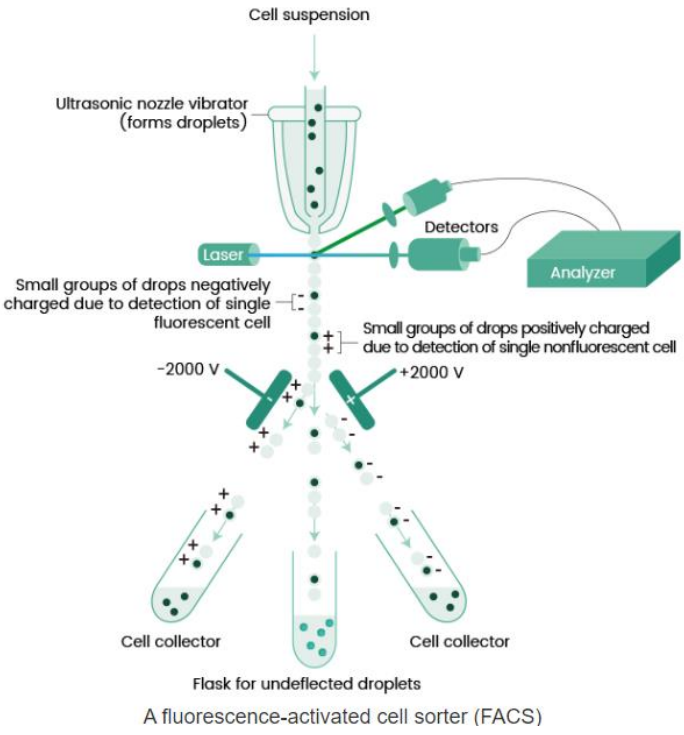

Figure 13. Visualization of FACS cell sorting. Visualization of how the cells were sent through the laser and detector, before charging of the cell and sorting into a tube (cell collector) or a 96-well plate in this method. Figure obtained from SinoBiological [71].

With this method the stained tumor cells could be sorted into wells on a 96 well plate containing 10µL RLT buffer. The plates were centrifuged right away, to make sure that the cells were located at the bottom of the wells in the buffer, and not on the walls of the well. The plates were then put in the -80̊ C freezer as quickly as possible, to ensure good quality of the DNA and RNA. The sorting of cell line HCC38 during preliminary testing, is visualized in Figure 14.

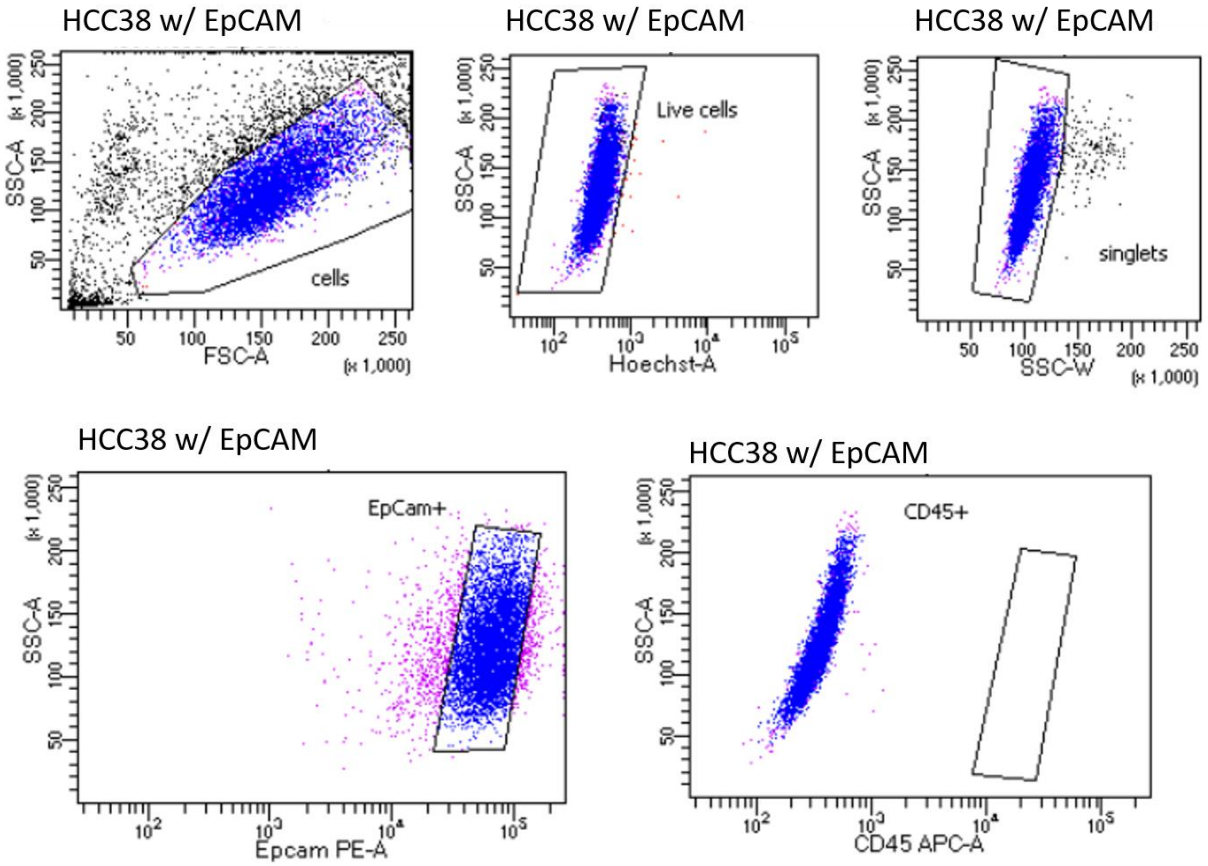

Figure 14. HCC38 sorting on FACS. An example of preliminary testing with a suspension of HCC38 stained with anti-EpCAM and anti-CD45 sorted by the FACS instrument. The top left scatterplot shows forward scatter and side scatter, the top middle scatterplot shows the Hoechst staining, and the top right scatterplot show the size of the cells to discriminate the doublets. In the example above, the Hoechst staining was not added, but the Hoechst plot still differentiates the live cells due to their autofluorescence. A gating has been set to sort out single live cells, and the bottom two scatter plots show the marker expression, where the cells that are positive for EpCAM (bottom left) are gated.

4.3.4 DEPArray (method 3.0)

The method with the DEPArray selection and isolation use the CellSearch® system (Menarini Silicon Biosystems) for identification and enrichment of CTCs followed by DEPArray for selection and isolation of single cells. The tumor cells from the HCC38 cell line were spiked into a fullblood sample and not in suspension like the other methods.

4.3.4.1 CellSearch enrichment

The CellSearch system is an instrument that can identify and isolate circulating tumor cells (CTC) from a blood sample. The system detects CTCs of epithelial origin based on surface markers with positive EpCAM and negative CD45, in addition to cytokeratin 8+, 18+ or 19+ [72]. The CellSearch system was used as the enrichment step in the method with CellSearch and DEPArray (Menarini Silicon Biosystems). As the system is optimized for CTCs and not DTCs, fullblood sample was needed for enrichment of the tumor cells.

Procedure:

A fullblood sample (in a CellSave tube) from a healthy donor was spiked with HCC38 cells to test how specific the CellSearch system could identify and isolate the tumor cells. With the use of the CellSearch CTC kit, the system uses immunomagnetic beads and fluorescence imaging technology for analysis and selection of tumor cells of choice [72]. This is a method of positive immunomagnetic bead selection of epithelial cells with the help of ferrofluids (nanoparticles with a magnetic core and a coat of antibodies against the EpCAM antigen). After the immunomagnetic selection, antibodies marked with fluorescence are added; anti-CK-PE (epithelial cells), anti-CD45-APC (leucocytes) and the DNA in the nucleus is stained with the blue fluorescent 4′,6-diamidino-2-phenylindole (DAPI). The instrument can select the tumor cells based on morphology and correct phenotype; CK+, DAPI+ and CD45-. In this thesis the CellSearch system was run by personnel at Section for experimental pathology, Dept. of pathology, OUS. The enriched sample (from the fullblood sample) were deposited into a CellSearch cartridge and further processed before the DEPArray™ system (see 4.3.4.2 DEPArray Single Cell Isolation).

4.3.4.2 DEPArray Single Cell Isolation

DEPArray™ technology system is an instrument that use a non-uniform electric field to exert forces on neutral, polarizable particles, such as cells suspended in a liquid. The principle of

electro kinetic called dielectrophoretic (DEP), trap cells in positions in an array and with DEP the cells can be moved by a change in the electric field patterns. With a combination of high quality image-based cell selection, the DEPArray™ system can manipulate and recover individual cells of interest [73].

Procedure:

To prepare the cells in the cartridge from CellSearch an extraction protocol for fixed cells from Menarini Silicon Biosystems was used [74]. First a gel loading tip was coated with a solution of PBS + 2% BSA and set aside in a 1,5 ml Eppendorf tube. The top of the CellSearch cartridge was carefully removed with the back end of a 100-1000µL tip. The coated tip was then used to carefully resuspend the cell suspension five times before transferring 200µL of the cell suspension to the sample tube, a 1,5ml Eppendorf Protein Lo-Bind tube. This was repeated until all the cell suspension was transferred. The coated tip was put back to the store tube. 325µL SB115 buffer was added at the top of the cartridge with new tips. The coated tip was used to mix the buffer gently in the cartridge five times to wash the inside and transfer the whole volume to the sample tube. This step was repeated once. The sample tube was centrifuged in a swinging bucket centrifuge (Eppendorf 5810R) at 1000G for 5 minutes. Then the supernatant was gently aspirated with a new 1000µL tip, until there was approximately 50µL left in the sample tube and discarded. Using a new 100-1000µL tip, 1000µL SB115 buffer from the flask were added to the top of the sample tube, so that the buffer ran down alongside the walls of the tube without mixing. The sample tube was centrifuged again at 1000G for 5 minutes in the swinging bucket. Carefully removed and discarded 900µL of the supernatant. The pipet was set to 90µL to remove more of the supernatant. Using new tips for each of these steps. The pipet was set to 12µL, and the exact remaining volume was measured. Using the same tip for this step and for transferring the sample to the DEPArray cartridge. The volume should be exactly 12µL, if the volume is <12µL, adjust with SB115 buffer. If the volume is >12µL, centrifuge at 1000G for 5 minutes and remove excess volume. Adjust to 12µL. This protocol is accurate, in order to extract all the cells in the cartridge. 2500 µL degassed SB115 buffer were added to the DEPArray cartridge, this to ensure fewer small bubbles. Then the 12µL sample were added and the cartridge was inserted in the DEPArray [75].

After instrument start up, the CTC-RUO program for fixed cells was selected. Chip scan settings with APC (anti-CD45), FITC (core colour), PE (anti-cytokeratin), DAPI and brightfield were chosen. The enrichment by CellSearch was not completely specific, and some leukocytes were also sorted out with the tumor cells. After sample loading the sample scan was started,

with images shown continuously. This was paused and by choosing selected images that contained the CTCs and leucocytes, the optimal light contrast, focus time and background were selected for all filters. This made the images clearer and easier to differentiate whole cells from fragmented cells. The new values were then used to scan the whole cartridge. In the next step, the cell selection was performed. First the cells of different characteristics were divided into sub-populations. Looking at scatter plot and data visualisation, gates were set around the populations and tables were created for each sub-population. The sub-populations were given names, such as "PE+, APC-" for the HCC38 tumor cells. Viewing images of each cell, as seen in Figure 15, cells were selected based on the strength of the PE and APC staining [76]. MicroAMP® Reaction Tubes with Cap (0.2mL) were inserted in a tray, for as many cells that would be extracted.

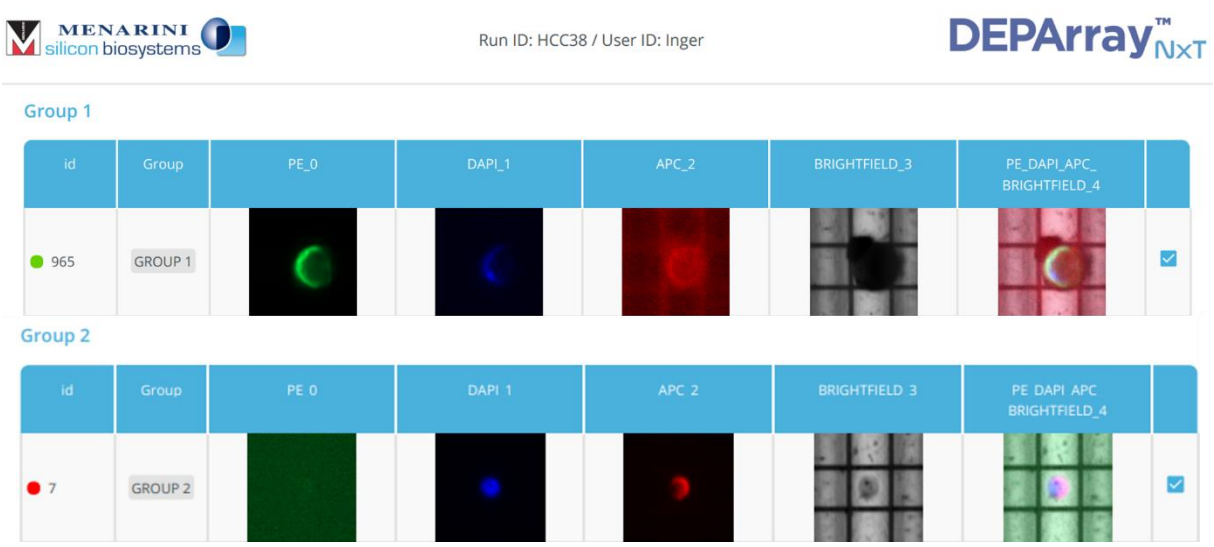

Figure 15. Cell selection on DEPArray. One single cell from each group is shown as example, Group 1 with HCC38 cells and Group 2 with leukocytes. PE (anti-cytokeratin), DAPI (nuclear staining), APC (anti-CD45), and brightfield, shown first separate and then as a complete image. The tumor cell had a clear green PE staining around the cell, a central blue DAPI staining, and no clear red APC staining, mostly background APC. The brightfield shows something around the cell, but the cell itself looks intact and whole. The leukocyte has no PE staining, a clear blue DAPI stain and a strong red APC staining. The brightfield shows that the leukocyte cell lies separate and is smaller than the tumor cell.

After collection of selected singe cells in the MicroAMP® Reaction Tubes from the DEPArray, the VRNxT™ - Volume Reduction Instrument was used to reduce the output volume. The tubes with the isolated cells were first centrifuged at 1000G for 3 minutes to ensure that the cell was at the bottom of the tube. Then 100µL PBS was added to each tube and centrifuged again at

1000G for 10 minutes. The sample tubes were then placed in a holder and the tubes were opened so that the dedicated cap could be put on. The “DEPArray OUT” program was chosen on the instrument, and the holder with the samples was inserted. The VRNxT™ uses centrifugal forces to collect and trap the excess fluid in the dedicated cap [77]. After the volume reduction the holder was taken out of the instrument and the cap was removed. Each sample was removed carefully, and the lid was closed [78]. 10µL RLT buffer was added to all the samples, and they were stored in a -80° C freezer.

4.4 cDNA preparation

The RNA of the lysed single cells was converted to cDNA before further analysis. This conversion was done by reverse transcription, which is when reverse transcription enzymes copy an RNA template into double-stranded DNA called cDNA. With short primers complementary to the 3’end of the RNA, the first stand cDNA can be used directly as a template for the polymerase chain reaction (PCR). PCR is a method to amplify many copies of cDNA [2]. For this process, the SmartSeq2 protocol was used. The SmartSeq2 protocol is designed to create cDNA of high quality from single cells [55]. With this protocol it is possible to study cell-to-cell variation in the transcriptome, even with low levels of RNA from the cell. Prior to the SmartSeq2 protocol, the cells have been deposited in a lysis buffer (RLT) that lyse the cell, blocks the degradation of RNA, and stabilizes it with a ribonuclease inhibitor. With the SmartSeq2 protocol the low amount of RNA from lysed single cells was reverse transcribed to cDNA and amplified. The cDNA can be sequenced and studied to understand the cells transcriptome and heterogeneity between single cells in different environments[79].

Procedure:

The SmartSeq2-protocol was performed in a UV-sterilized hood and the surfaces was sprayed with RNase before start, to prevent possible degradation of RNA.

1. **Clean up and removal of the RLT buffer with RNA Clean beads (Agencourt).**

This first step is not included in the SmartSeq2 protocol by Picelli et.al. 2014 [55]. It was added as a first step to clean up the sample and remove the RLT lysis buffer, as the RLT buffer will degrade the enzymes used in the Smartseq2 protocol. With a bead: sample ratio of 0.8:1, 8µL RNA Clean beads was added to the 10µL sample. In the first rounds of testing, 2,3µL RLT buffer was used, but for the manual cell isolation methods an increase in the volume to depose the cells in made the procedure easier. With the use of low-elution magnets, the sample was washed twice with 100µL 80% ethanol to remove the RLT buffer. After the last wash all the ethanol residue was removed, and the beads air dried for five minutes. The samples were then eluted in 2,5µL nuclease free water and incubated off the magnet. After five minutes incubation, the strips was transferred to the magnet, and incubated five minutes allowing the beads to settle, and the solution was clear. With the sample strip on the magnet, 2,3µL of the sample was transferred to a new strip without disturbing the beads.

2. **Addition of free dNTPs and oligo(dT).**

Incubation with free 10mM dNTP Mix and 10µM oligo(dT)- tailed oligonucleotides with a universal 5'-anchor sequence. A mix of 1µL free dNTPs and 1µL oligo(dT) to each well, the samples were incubated in a BioER LifeECO ThermoCycler with a 105°C heated lid, at 72°C for 3 minutes and immediately put back on ice.

10µM oligo(dT) was made by a dilution from a 100µM oligo(dT); 10µL of 100µM oligo(dT) + 90µL nuclease free water.

For more than one sample, a mix of dNTP Mix and oligo-dT was made, with 10% extra.

Table 3. dNTP Mix + Oligo mix.

dNTP + Oligo mix	1 well (µL)
10 mM dNTP Mix	1
10 µM oligo (dT)	1

3. Preparation of a reverse transcription mix.

The mix for the reverse transcription was prepared in a 1,5mL Eppendorf tube, with 10% extra of each reagent when preparing multiple samples.

Table 4. Contents for the reverse transcription mix.

RT mix	1 well (µl)
SuperScript II first strand buffer 5x	2
DTT 100mM	0,5
Betaine 5M	2
$MgCl_2$ 1M	0,06
Template Switching Oligo 100µM	0,1
SUPERase Inhibitor 200U/µl	0,25
Nuclease free water	0,5
Total	5,7

5,7µl of the RT mix was added to each well and mixed gently by vortex and spun down. The final volume for each sample was 10µl.

4. Reverse transcription reaction.

With a heated lid at 105°C, the samples were incubated with the following program, on a BioER LifeECO ThermoCycler.

Table 5. Thermocycler program for the reverse transcription reaction.

Cycles	Temperature (°C)	Time (minutes)
1	42	90
10	50	2
	42	2
1	70	15
1	4	Hold

5. PCR amplification.

The mix for the PCR amplification was prepared in a 1,5mL Eppendorf tube, with 10% extra of each reagent when preparing multiple samples.

Table 6. PCR mix preparation.

PCR	1 well (µl)
KAPA Hifi	12,5
IS PCR primers	0,25
Water	2,25
Total	15

15µl of the PCR mix was added to each well, and the samples were vortexed and spun down with a total volume of 25µl. With a heated lid at 105°C, the strips were incubated with the following program, on a BioER LifeECO ThermoCycler.

Table 7. Thermocycler program for PCR amplification.

Cycles	Temperature (°C)	Time
1	98	3 min
24	98	20 sec
	67	15 sec
	72	6 min
1	72	6 min
1	4	Hold

The PCR product could be stored at 4°C if the purification is subsequent or for longer at -20°C or -80°C.

7. cDNA purification.

For this step AMPure XP beads (Beckman Coulter) are used. First, the beads were equilibrated to room temperature for 15 minutes before use, and vigorously mixed. At the same time the strips with the amplified samples were spun down at 1000G for 1 minute at 4°C. 0,8:1 bead:sample ratio was used, 20µl beads was added to the 25µl samples. The samples were mixed thoroughly by pipetting up and down to a homogenous solution and incubated at room temperature for five minutes. The strips were then transferred to a low-elution magnet for five minutes allowing the beads to settle until the solution was clear. The supernatant was removed

carefully without disturbing the beads. With the strips on the magnet, the samples were washed twice with 100µl of 80% ethanol. After the second wash, all the residual ethanol was removed from the wells. The AMPure beads were then dried for five minutes to allow for all ethanol remains to evaporate. The strips were removed from the magnet, added 20µl of nuclease free water, and mixed well by pipetting up and down. The samples were incubated off the magnet for two minutes. Then the strips were returned to the magnet allowing the beads to settle for five minutes until the solution was clear. 18µl of supernatant was carefully transferred to a new strip and put in a -20°C freezer. The samples were then ready for the quality control steps.

4.5 Quality Controls

In this work, the choice of quality controls was important to be able to compare the methods. The quality controls used to measure the cDNA concentration, cDNA fragment size and the integrity of the cDNA from the samples.

4.5.1 Qubit

The Qubit® Fluorometer 4.0 was used to quantify the amount of cDNA present in the samples after the SmartSeq protocol. The fluorescent Qubit reagent binds specifically to the target of interest, which is cDNA. This control step provides a precise and accurate measurement for samples of low concentration, with high sensitivity [80]. The Qubit measurement of the cDNA was used as the first control step.

Procedure:

For this procedure the Qubit™ 1X dsDNA HS Assay Kits were used on the Qubit® Fluorometer 4.0. For each set-up, new standards were performed for a new calibration. The reagent and the standards had to equilibrate for 30 minutes to ensure room temperature before use. One Qubit® Assay tube was used for each standard and sample. For the two standards, the following volumes were added to the tube: 190µL Qubit™ Working solution + 10µL standard. For the samples the following volumes were added: 199µL Qubit™ Working solution + 1µL sample. The tubes were vortexed for 2-3 seconds and incubated for 2 minutes at room temperature. The program for dsDNA was selected on the Qubit® Fluorometer 4.0 and the two standards were read first [81]. The input volume of 1µL for the samples was selected before reading the

samples. The Qubit® Fluorometer 4.0 then calculated the original sample concentration in ng/µL. For the samples that were "too low" after the first measurement, a new Qubit tube with 195µL Qubit™ Working solution and 5µL sample was measured. This would also calculate the smallest concentrations if present in the sample.

4.5.2 Bioanalyzer

The Agilent 2100 Bioanalyzer system was used for high-resolution automatic electrophoresis, with a measurement of fragment length, quantity, and integrity of the fragments in the samples. With small sample concentrations this control step can separate primer-dimer fragments from the wanted fragment lengths. The analysis is based on traditional gel electrophoresis principles and have been transformed into electrophoresis on a chip. The DNA molecules are separated by size with the voltage gradient in the chip. Detection is based on laser-induced fluorescence detection (LIF) and the data is transformed into gel-like bands and electropherograms (peaks). A ladder is used as a standard with components of known sizes and migration time. Quantitation of the DNA is done by comparing the area under the upper marker peak with the area under the sample peaks [82].

All the samples were run on Bioanalyzer, and the samples with a higher concentration were diluted down to 2 ng/µL. The first runs demonstrated that too high concentrations would be out of measurable range. The high samples were not rerun, but the rest of the samples from the same method was diluted to demonstrate how the fragment lengths would be with this method.

Procedure:

The high sensitivity DNA kit (Agilent) was used. The reagents were equilibrated to room temperature for 30 minutes before use. A gel-dye mix was first made, with 15µL high sensitivity DNA dye concentrate added to a high sensitivity DNA gel matrix vial. The solution was mixed thoroughly by vortexing and spun down before it was transferred to a spin filter. The solution was centrifuged at 2240G for 15 minutes. A new High sensitivity DNA chip was set on the chip priming station and 9µL of the gel-dye mix was added to the well marked with a G (white on black). The plunger positioned at 1mL was pressed down until it was held by the clip in the priming station. After exactly 60 seconds the clip was released and after 5 seconds, it was slowly pulled back to 1mL position. The chip priming station was opened and 9µL of the gel-dye mix was added to the wells marked with a dark G. The pipette was at all times held at the bottom of the wells. 5µL of the marker was added to all the remaining wells. 1µL of the high sensitivity

DNA ladder was added to the well marked for the ladder. 1μL of the samples were added to the wells, with alteration between a negative control and RNA control [83]. The samples with a higher concentration given by Qubit, was diluted down to approximately 2ng/μL. When all the samples were added to the chip, the chip was set in a horizontally vortex for 1 minute at 2400rpm. The Agilent 2100 Bioanalyzer was turned on and the chip with the samples was inserted. A connected computer with the Bioanalyzer 2100 Expert Software was used. Sample ID was stated for each chip and the run was started within 5 minutes of chip preparation [83]. Before and after each run, the electrodes were cleaned with a cleaning chip filled with 350 μL deionized analysis-grade water.

4.6 Molecular analyses of RNA from single tumor cells

When the single cell RNA was amplified and converted to cDNA, further analysis could be performed. As an initial analysis, to assess the amplicability of the cDNA, the gene expression was analysed by digital PCR. In addition, a more detailed analysis of genome-wide RNA expression by 10X Genomics methodology, was performed for cells identified and selected by the FACS method.

4.6.1 Digital Droplet PCR

The digital droplet PCR (ddPCR) analysis was used to determine the cell type for the single cells isolated, by designed assays for the genes *KRT8*, *KRT18* and *CD45*. *KRT8* and *KRT18* are genes coding for two types of cytokeratin proteins expressed in epithelial cells, and *CD45* is a gene coding for a protein specific for normal white blood cells. These were chosen both to check whether or not the cDNA was amplifiable by PCR but also as expression of these proteins will separate the tumor cells from the leukocytes and serve as a control step to conclude a single cell's phenotype.

Digital droplet PCR is a highly sensitive technology that can identify and quantify small amounts of a target molecule without the use of standard curves or reference genes [84]. In this master thesis the QX200 digital droplet system (Bio-Rad) was used. The QX200 Droplet Generator creates between 10 000-20 000 water-oil emulsion droplets of each sample, and with the use of a VeritiDx thermo cycler (Thermo Fisher Scientific), end-point PCR was performed. The droplets function as test tubes were the PCR reaction takes place. Each droplet contain

primers, fluorescent probes (TaqMan) with FAM, HEX or VIC, specifically designed for the assay [84]. After PCR amplification the QX200 Droplet Reader creates a single flow of droplets, and the fluorescent signal of each droplet is measured. If the target gene(s) are amplified in a droplet, the probe will bind and give a fluorescent signal. The number of positive and negative droplets are used to calculate the concentration of the target molecules in each sample [85].

4.6.1.1 Digital Droplet PCR Assay Design

To perform ddPCR, an assay tailor-made for the desired genes must first be designed, with a forward primer, a reverse primer, and a probe. This was done with help from Dept. of pathology. Ensembl.com was used to find the sequence of the exons for each of the genes. At least one primer per gene should span an exon-exon junction, to avoid amplification of DNA. The program PrimerExpress (Thermo Fisher Scientific) was used to design possible forward and reverse primers, and probes for the selected exons. This program also calculated the annealing temperatures that was used in the PCR process. See materials section for details about primer-probe assay sequences.

For primer design the following guidelines from Bio-Rad were followed:

- The designed primers have 50-60% GC content.
- The annealing temperature T_m should be 50-65℃.
- Avoid secondary structures, such as primer-dimers by adjusting the primer locations to be outside the target sequence secondary structure.
- Avoid repeats longer than 3 bases of Gs or Cs.
- Choose primers with Gs and Cs at the 3’ nucleotide of primers when possible.
- Check that the forward and reverse primer sequences does not have 3’ complementarity to avoid primer-dimers.

For probe design the following guidelines from Bio-Rad were followed:

- The probe sequence must be between the two primers of the amplicon, but the primer cannot overlap the probe.
- The T_m should be 3-10℃ higher than the primer T_m.

- The sequence chosen should be within the target that has a 30-80% GC-content, and the probe must be designed to anneal to the strand that has more Gs than Cs, so the probe contains more Cs than Gs.
- The probe should be less than 30 nucleotides long because the distance between fluorophore and quencher affects the baseline signal intensity.
- Probes should not have a G nucleotide at the 5' end because this quenches the fluorescence signal even after hydrolysis.

Basic Local Alignment Search Tool (BLAST) was used to ensure primer specificity, by checking that PCR primers only matched the intended target genes, KRT8, KRT18 and CD45 [85].

All primers and probes for the digital droplet PCR quality control set-up was ordered from Integrated DNA Technologies (IDT), with sequences as shown in Table 8.

Table 8. Primers and probe sequences for the KRT8, KRT18 and CD45 genes.

KRT8	**Sequence**
Forward primer ex3-4	5'- CTC GAA GCA ACA TGG ACA ACA -3'
Reverse primer ex3-4	5'- ACT TCT CCT GGC CCA GAG TCT –3'
Probe ex3-4	5'-/56-FAM/AGA GCT ACA /ZEN/ TCA ACA ACC TTA GGC GGC A/3IABkFQ/ -3'
KRT18	**Sequence**
Forward primer ex2-3	5'- CCG CAT CGT TCT GCA GAT T-3'
Reverse primer ex2-3	5'-GGC CAG CTCTGT CTC ATA CTT GA–3'
Probe ex2-3	5'-/56-FAM/AAT GCC CGT /ZEN/ CTT GCT GAT GAC TT/3IABkFQ/ -3'
CD45	**Sequence**
Forward primer ex5-6	5'- CTC CGC CGC CAA TGC -3'
Reverse primer ex5-6	5'- TCC TGG GAC ATC TGA GAT AGC A-3'
Probe ex5-6	5'-/5HEX/AAA CTC AAC /ZEN/ CCT ACC CCA GGC AGC A/3IABkFQ/ -3'

4.6.1.2 Optimisation process

To test the designed assays the ddPCR was first run as a single-plex, which means each gene assay was tested separately. When the assays have been tested, the goal is to multiplex the assays, to use less reagents and less volume of each sample. The steps in the optimisation process are visualized in Figure 16.

To test the assays positive controls for the genes was necessary and therefore HCC38 and HCC2218BL DNA and RNA was extracted using the manufacturers protocol for Allprep DNA/RNA/miRNA Universal kit automated on the QIAcube (Qiagen). The RNA was synthesized to cDNA with the superscript IV VILO mastermix (Thermo Fisher Scientific) and a two-step RT-qPCR. The DNA and cDNA from both cell lines were used as positive controls to verify that the assays were working. Nuclease free water was used as a no template control (NTC) to ensure that there is no cross contamination.

After some testing of the assays the results were not optimal for all three assays. Therefore, a temperature gradient test was performed, to find the optimal annealing temperature. This temperature would have the greatest amount of signal separation (between positive and negative) and the lowest amount of droplet rain. Rain is when droplets are in between the positive and negative signals, in a way that makes them look like rain from the positive "cloud" of droplets. 60°C was found to be the optimal annealing temperature for all three assays [86].

When the single-plex assays were adequate, the assays were tested for multiplex. The amount of input of cDNA was also tested, with input from 0,25-1ng. The input should be at a level that gives a clear separation of negative and positive droplets, as the software uses a Poisson distribution to determine the initial copy number of the target DNA molecule in the input reaction as copies/ μL in the sample [84]. Further testing on the amount of input was done, with 0,06ng, 0,125ng and 0,25ng. The ideal input was determined to be approximately 0,125ng. This amount was used when testing the samples. Testing of multiplex analysis of the assays gave best results with a multiplex of KRT18 and CD45. The KRT8 assay was excluded from the multiplex, as it had more rain and intermediate fluorescent signal from the droplets and had a less clear separation between the negative and positive droplets. Multiplexing the assays would save both reagents, sample volume and time.

To test if the single cell RNA is contaminated with DNA that could affect the assay, a DNA specific test was performed. An assay designed specifically for DNA was obtained from the department of Pathology and tested with both DNA and RNA from cell lines and RNA from single cells.

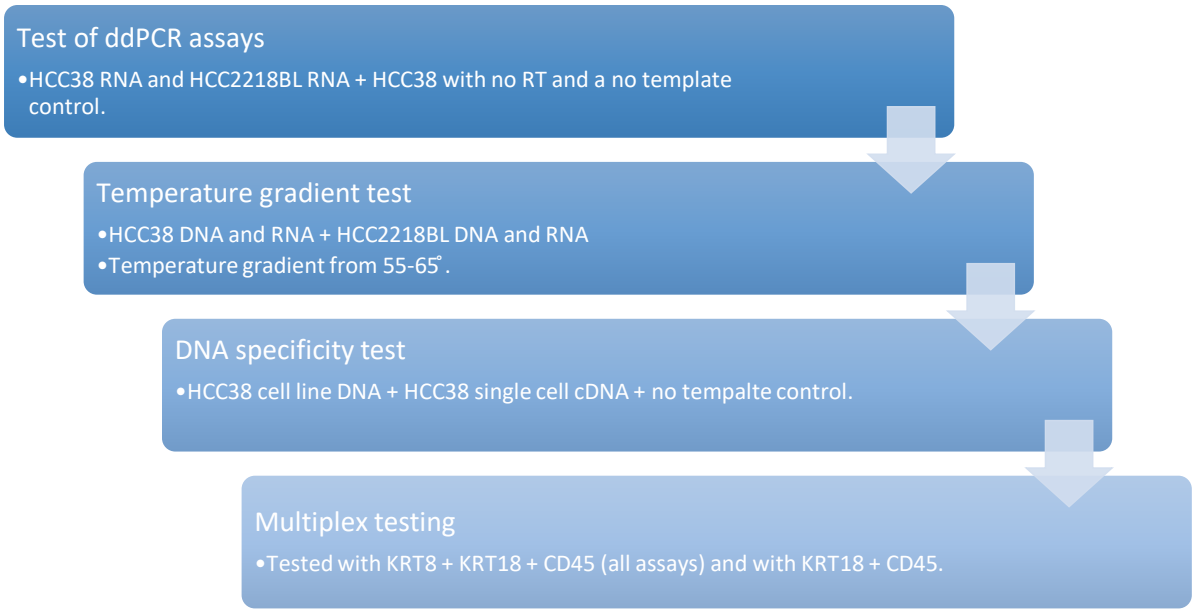

Figure 16. Steps in ddPCR optimisation process. Several tests were performed to find the optimal temperatures, to ensure DNA specificity, and to test multiplexing of the assays. This figure shows the stepwise process of assay optimization. HCC38 and HCC2218BL DNA and RNA extracted by QIAcube.

4.6.1.3 Digital Droplet PCR analysis

After the optimization process and testing of multiplexing the assays, the samples from each method were analysed.

Procedure:

The ddPCR protocol was Based on protocol from Bio-Rad [85].

First a primer-probe mix was made for each of the genes separately, with 900/900/250nM ratio. The concentration of both forward and reverse primer was 100µM and 10µM for the probe. For 20 reactions, a primer-probe mix was made by 9µL forward primer + 9µL reverse primer + 25µL probe.

A mastermix of supermix, primer-probe mix, and nuclease free water was made for each gene assay in a DNA LoBind 1,5mL tube (Eppendorf). For preparation of mastermix for several samples, an addition of 20% volume was used, to have enough for all the samples. 22,8µL of the mastermix was added to an 8-well PCR strip and 1,2µL of each sample was added to the respective wells in the strip.

Table 9: Set-up for mastermix and sample preparation. The mastermix was first made in a separate DNA LoBind tube, and then added in an 8-well PCR strip along with the samples.

Reagent	**Volume per reaction/ well, µL**	**Volume per well (+ 20%), µL**	**8 wells, µL**
2x ddPCR Supermix for probes (no dUTP)	10	12	96
20x target (FAM and/or (HEX) primer/probe*	1	1,2	9,6
Sample	1	1,2	9,6
Nuclease free water	8	9,6	76,8
Total volume	20	24	192

* 900nM of forward and reverse primer and 250nM of the probe.

For droplet generation a new DG8 Cartridge (Bio-Rad) with 8 wells was used for each column of samples. The cartridge was set in the cartridge holder and 20µL of each sample was added to the middle row of wells, with the pipette at an angle. 70µL Droplet Generation Oil (Bio-Rad) was added by slow pipetting at an angle to the wells of the bottom-row. A gasket was fastened over the cartridge by the taps on the side of the cartridge holder. The cartridge was inserted into the QX200 Droplet Generator, and the generation of droplets started automatically when closing the lid.

After droplet generation, 40µL of the droplet emulsions in the top row wells were transferred to a semi-skirted 96-well plate (Bio-Rad). The pipetting of the droplets should be done slowly, for 5 seconds up and 5 seconds down, to ensure that all droplets were transferred to the plate. When all the samples were transferred to the 96-well plate, the plate was sealed with foil by a PX1 Plate sealer, for 5 seconds at 180°C. Then a VeritiDx Thermal cycler (Thermo Fisher Scientific) was used for the amplification of the target sequences.

Samples from each method was analysed together in a plate set-up, as shown in Appendix 1. Plate set-up for digital droplet PCR.

Table 10: PCR program for amplification of target sequence.

Cycling step	Temperature°C	Time	Number of cycles	Ramp rate
Enzyme activation	95	10 min.	1	2°C/sec
Denaturing	94	30 sec.	40	
Annealing/extension	60	1 min.		
Enzyme deactivation	98	10 min.	1	
Hold	4	∞	1	1°C/sec.

The finished PCR reaction was set to 4°C over-night to stabilize the droplets and give a higher number of valid droplets in the reading.

The QX200 Droplet Reader (Bio-Rad) was used with QuantaSoft software program on a connected computer. The 96-well plate was inserted into the QX200 Droplet Reader without removing the foil seal. In QuantaSoft the wells that should be analysed were defined with FAM/HEX fluorophore, ddPCR 2x Supermix (no UTP) and sample ID. The genes *KRT8* and *KRT18* was read in channel 1 (FAM) and *CD45* was read in channel 2 (HEX). In the single-plex testing only one channel at a time was selected, for the respective gene assays, but in the multiplex testing, both channels were selected for the same well.

4.6.2 Single cell gene expression by 10X Genomics

The 10X Genomics analysis was performed to address how the RNA was affected by the identification and selection procedures in the FACS method, in comparison to unstained, viable single cells in suspension. The analysis could give a better understanding whether the RNA from all genes (i.e., globally) were lost, or if some genes were more affected than others. The FACS sorting was performed by the Flow cytometry core facility, Dept. of core facilities, Institute for Cancer Research and the 10X Genomics were performed by the Genomics Core facility, Norwegian Radium Hospital.

Procedure:

The 10X Genomics has a set-up for four samples at a time. In this project, two parallels of unstained, viable cells in suspension and two parallels of stained cells sorted through the FACS instrument were tested. The requirements were 32µL of each sample with 500-700K cells/µL

in a PBS + 0,5% BSA solution. The sample preparations were done by the student and the sorting by FACS was done by the flow cytometry core facility.

The cells were split routinely as described in 4.2 Cell lines, and in this experiment, they were resuspended in PBS with 0,5% BSA instead of cell culture medium. The cells from one flask were used for FACS staining and sorting, with the same procedure as described in 4.3.3. with anti-EpCAM, anti-CD45 and Hoechst 33258 staining. 268 500 cells were sorted into a FACS 5mL tube with 1mL PBS with BSA. The Countess™ automated cell counter was used to count the cells in dilution down to a suspension of 500-700K cells/µL. The two parallels of 32µL were deposed in a 1,5mL DNA LoBind tube. The cells from another flask were used for the two parallels of unstained, viable cells. These cells were only split and resuspended in the PBS with BSA before counting on the Countess cell counter, and dilution.

The following samples were delivered to the Genomics Core Facility for 10x samples prep and subsequent sequencing:

1: HCC38 unstained, viable with 677 cells/µL.

2: HCC38 unstained, viable with 677 cells/µL.

3: HCC38 stained and sorted by FACS with 540 cells/µL.

4: HCC38 stained and sorted by FACS with 540 cells/µL.

4.7 Patient samples

Two patient samples were included in this master thesis, as part of a large collaborative project where DTCs are to be subjected to whole genome and whole transcriptome (G&T) sequencing at the Wellcome Sanger Institute in Cambridge. Both primary tumor and later metastases from each of the patients were available, and the aim of the project is comparison of DNA and RNA alterations from primary tumor, DTCs and metastatic tumor tissue to infer the two patients' evolutionary trajectories of their disease.

One vial of bone marrow aspirates from each of the two samples were thawed and sorted by FACS as previously described. In addition, DTCs were isolated from cytospins by the micromanipulation method.

Procedure:

The samples were processed one at a time for the whole process of thawing, identification, and sorting, manually and automatically. Thawing of the patient samples were done in a lab for cell culture work only. Prior to the patient samples, a test sample, i.e., a patient sample from a patient that was omitted from the trial, was used to test how the sample performed after 20 years in the freezer. This sample was used to test the workflow and to test the staining protocol of the cytospins. The staining of the cytospins was done as described for the micromanipulator 1.1 method.

Protocol for thawing samples by the section for experimental pathology, department of pathology OUS: Before the cell vials were thawed, 15mL RPMI culture medium was heated up to 37°C in a 50mL Sarstedt tube in a water bath (VWR). 300µL DNase 1 (2mg/mL) was added to the culture medium when it was warm. The cells were thawed directly in the water bath, for no longer than 1 minute. 10% DMSO (that is added to the cells before freezing) is harmful for the cells and must be diluted as soon as the cells are thawed. The cells were transferred to the warm culture medium immediately after thawing, and the capsule was rinsed with medium to transfer all the cells possible. The cell suspension was left to incubate for 15 minutes in room temperature. The cells were then centrifuged at 300G for 10 minutes. The supernatant was removed, and the cell pellet was resuspended in PBS + 1% FBS. The pellet was at times difficult to resuspend, possibly due to many years of storage. The filter on FACS tubes were used to filter the suspension. The cells were counted on Countess. For the cytospins, 3mL with 3 x 10^6 cells were used. The rest of the cell suspension was used for FACS.

Cytospins were prepared by the Department of Pathology and kept in -80°C freezer. The staining of the cytospins was done with the same protocol that was used for cell line HCC38. Three cytospins was thawed and stained at a time, two with 500K cells and one with 250K cells, when processing both patient samples. The micromanipulator was used to isolate the tumor cells. The patient samples were much different from the cell lines in this case, and an experienced pathologist helped to identify the tumor cells.

Staining protocol before FACS sorting was the same as the protocol used for the FACS method with cells from cell line HCC38. 2 µL anti-EpCAM-PE and 5µL anti-CD45-APC was used for both samples. FACS was again performed at the department of core facilities, the Flow cytometry core facility, Institute for Cancer Research. As for the cell line sorting, an unstained sample was used to help set the gating around the tumor cells in the patient samples. After the

first run of the unstained sample, Hoechst 33258 (live/dead marker) was added to the stained cell suspension. The pictures in Figure 17 and Figure 18 show how the cells were gated for the patient samples with staining and live/dead marker. The sorting of the patient samples shows more cells and larger populations than for the cell line.

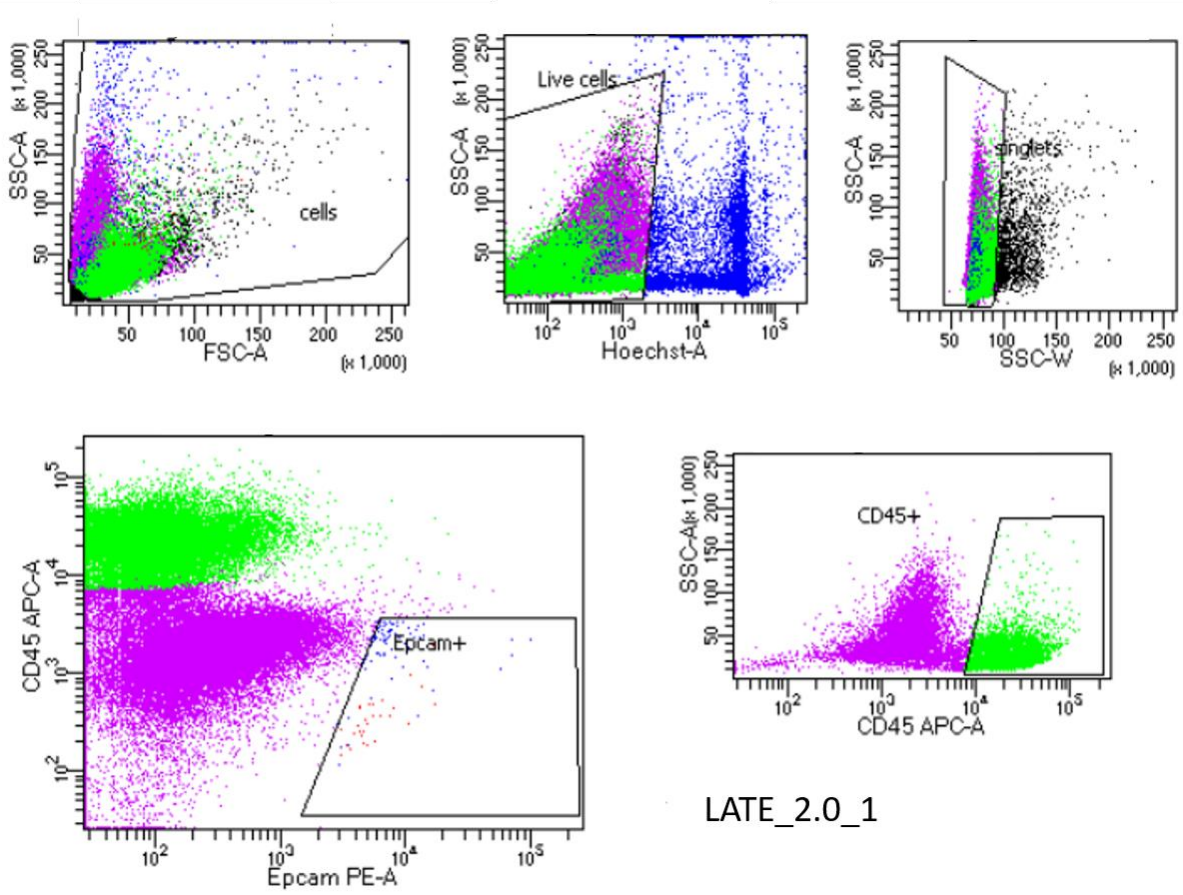

Figure 17: Gating for patient sample #1 with staining and added Hoechst 33258. The top left scatter plot shows the forward scatter versus side scatter, and separates the cells based on size and granularity (complexity). The top middle scatter plot shows the dead cells stained in blue, and a gating set around the live cells. The top right scatter plot shows the size of the cells and discriminates the doublets, with a gating around the single cells. The bottom right shows the leukocytes with CD45+ expression, and the rest is most likely debris. The bottom left scatter plot show the marker expression, with the rare tumor cells with EpCAM+ gated in the lower right quadrant. This shows how few tumor cells there was compared to normal cells in the bone marrow.

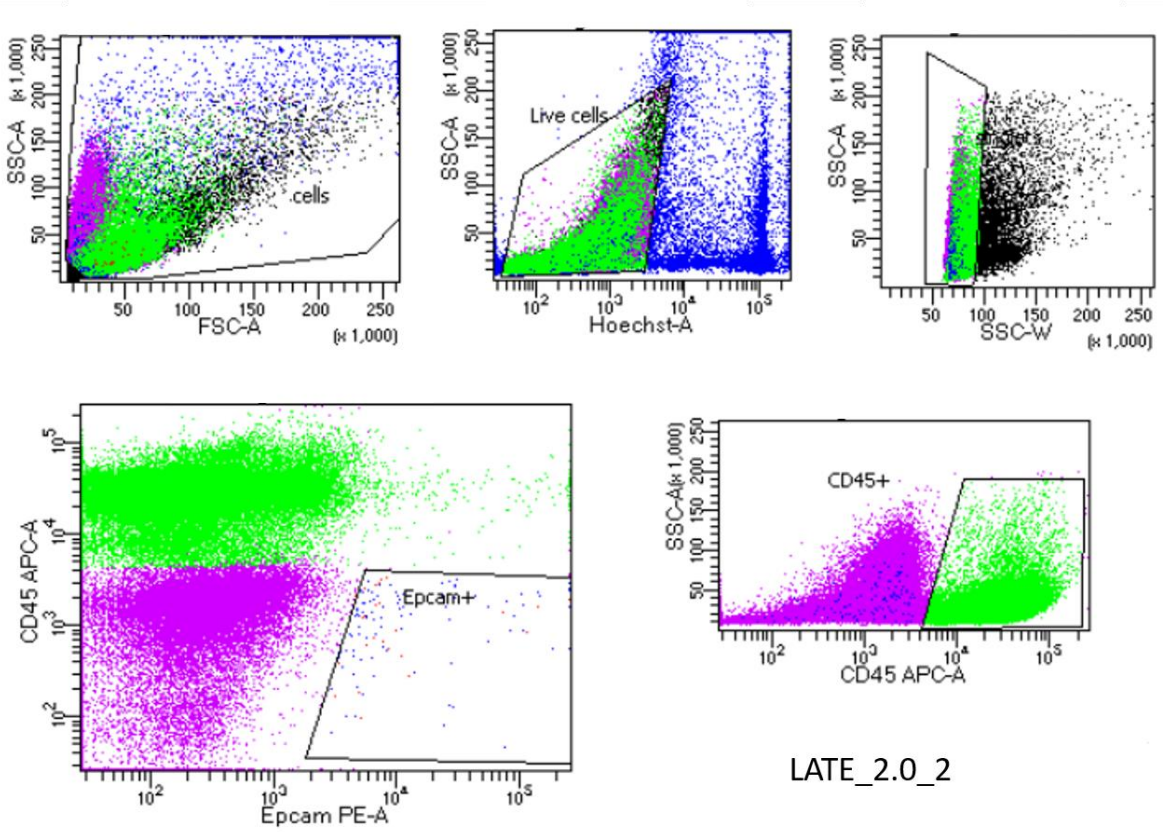

Figure 18: Gating for patient sample #2 with staining and added Hoechst 33258. The top left scatter plot shows the forward scatter versus side scatter, and separates the cells based on size and granularity (complexity). The top middle scatter plot shows the dead cells stained in blue, and a gating set around the live cells. The top right scatter plot shows the size of the cells and discriminates the doublets, with a gating around the single cells. The bottom right shows the leukocytes with CD45+ expression, and the rest is most likely debris. The bottom left scatter plot show the marker expression, with the rare tumor cells with EpCAM+ gated in the lower right quadrant. This shows how few tumor cells there was compared to normal cells in the bone marrow.

Both samples were sorted into 3 half-filled 96-well plates, column A-D. The plates were only filled in half of the wells to prevent single cells on the wall of the wells to dry out because the sorting time was long. After sorting of each plate, it was spun down and put in -80°C freezer as quickly as possible. When all plates were sorted, as many tumor cells as possible were sorted out into a 1,5 DNA LoBind tube with RPMI 1640 medium. For sample #2, 239 cells were sorted and for sample #1, 700 tumor cells were sorted. For sample #1 1 million CD45+ cells were also sorted, and these cells were frozen as a cell pellet at -80°C.

The tumor cells that were sorted in the culture medium was manually isolated by the 1.0 method, with the microinjection pipette the same day. The cells were picked into 8-well PCR strips with buffer RLT and put on dry ice as soon as the strip was full. For sample #2, 26 cells were picked and for sample #1, 40 cells were picked.

To test if the single cells that were identified and sorted were tumor cells, four of #1 single cells and three #2 single cells that were identified and selected by FACS, and then manually isolated by the microinjection pipette, were converted to cDNA by the Smartseq2 protocol. The concentration of the cDNA was measured by Qubit, the fragment quality was measured by Bioanalyzer, and to analyse the cell origin, the some of the cells with the highest concentrations were analysed with the ddPCR assay.

5. Results

The results consist of measurements and analysis of the 30 isolated HCC38 single cells' cDNA from each of the four methods. The results from each single cell are presented with cDNA concentration, fragment size distribution and measured gene expression.

The results from the 10X Genomics analysis were not ready in time to be included in this thesis.

For the patient samples, only one of them have been sent to the Wellcome Sanger Institute, but the results are not yet ready, as a result of a delay due to the COVID-pandemic and Brexit process. Some single cells were kept in Oslo and tested, to see if the identification and selection had sorted tumor cells (results presented below).

5.1 Concentrations of cDNA obtained from single cells

After all the identification and selection processes in the different methods, as described in

4.3 Tumor cell identification and single cell isolation, conversion of RNA and amplification of the cDNA, the concentrations were measured. The concentration of cDNA in the samples was measured on Qubit using the HS DNA assay. The results are shown in Table 11 in ng/µL. Some of the samples had concentrations below the detection limit, which means close to zero. These values were set to 0.

Table 11. Concentrations in ng/µL, measured with DNA HS assay on Qubit. The results from each method are separated into rows, with cell identification number and concentration under Qubit (ng/µL).

Method 1.0 Microinjection pipette	
Cell ID #	Qubit
COMP_1.0_HCC38_1	25,6
COMP_1.0_HCC38_2	8,5
COMP_1.0_HCC38_3	5,76
COMP_1.0_HCC38_4	9,22
COMP_1.0_HCC38_5	3,58
COMP_1.0_HCC38_6	19,6
COMP_1.0_HCC38_7	8,28
COMP_1.0_HCC38_8	8
COMP_1.0_HCC38_9	6,34
COMP_1.0_HCC38_10	4,42
COMP_1.0_HCC38_11	2,34
COMP_1.0_HCC38_12	4,88
COMP_1.0_HCC38_13	4,92
COMP_1.0_HCC38_14	9,78
COMP_1.0_HCC38_15	4,18
COMP_1.0_HCC38_16	5,8
COMP_1.0_HCC38_17	3,2
COMP_1.0_HCC38_18	5,66
COMP_1.0_HCC38_19	3,36
COMP_1.0_HCC38_20	3,72
COMP_1.0_HCC38_21	11,3
COMP_1.0_HCC38_22	22
COMP_1.0_HCC38_23	6,6
COMP_1.0_HCC38_24	7,44
COMP_1.0_HCC38_25	5,12
COMP_1.0_HCC38_26	11,7
COMP_1.0_HCC38_27	6,34
COMP_1.0_HCC38_28	9,4
COMP_1.0_HCC38_29	8,4
COMP_1.0_HCC38_30	10,3

Method 1.1 Micromanipulator	
Cell ID #	Qubit
COMP_1.1_HCC38_1	0,00
COMP_1.1_HCC38_2	0,00
COMP_1.1_HCC38_3	0,00
COMP_1.1_HCC38_4	0,00
COMP_1.1_HCC38_5	0,00
COMP_1.1_HCC38_6	0,00
COMP_1.1_HCC38_7	0,00
COMP_1.1_HCC38_8	0,00
COMP_1.1_HCC38_9	0,00
COMP_1.1_HCC38_10	0,10
COMP_1.1_HCC38_11	0,03
COMP_1.1_HCC38_12	0,00
COMP_1.1_HCC38_13	0,00
COMP_1.1_HCC38_14	0,00
COMP_1.1_HCC38_15	0,02
COMP_1.1_HCC38_16	0,02
COMP_1.1_HCC38_17	0,03
COMP_1.1_HCC38_18	0,00
COMP_1.1_HCC38_19	0,00
COMP_1.1_HCC38_20	0,00
COMP_1.1_HCC38_21	0,18
COMP_1.1_HCC38_22	0,14
COMP_1.1_HCC38_23	0,00
COMP_1.1_HCC38_24	0,00
COMP_1.1_HCC38_25	0,00
COMP_1.1_HCC38_26	0,00
COMP_1.1_HCC38_27	0,00
COMP_1.1_HCC38_28	0,00
COMP_1.1_HCC38_29	0,00
COMP_1.1_HCC38_30	0,00

Method 2.0 FACS	
Cell ID #	Qubit
COMP_2.0_HCC38_1	0,52
COMP_2.0_HCC38_2	0,52
COMP_2.0_HCC38_3	0,70
COMP_2.0_HCC38_4	2,04
COMP_2.0_HCC38_5	4,64
COMP_2.0_HCC38_6	1,95
COMP_2.0_HCC38_7	0,84
COMP_2.0_HCC38_8	3,48
COMP_2.0_HCC38_9	0,60
COMP_2.0_HCC38_10	0,96
COMP_2.0_HCC38_11	1,03
COMP_2.0_HCC38_12	4,4
COMP_2.0_HCC38_13	0,61
COMP_2.0_HCC38_14	0,81
COMP_2.0_HCC38_15	0,87
COMP_2.0_HCC38_16	1,44
COMP_2.0_HCC38_17	3,78
COMP_2.0_HCC38_18	0,71
COMP_2.0_HCC38_19	3,3
COMP_2.0_HCC38_20	2,3
COMP_2.0_HCC38_21	2,26
COMP_2.0_HCC38_22	0,85
COMP_2.0_HCC38_23	1,32
COMP_2.0_HCC38_24	0,00
COMP_2.0_HCC38_25	0,00
COMP_2.0_HCC38_26	0,79
COMP_2.0_HCC38_27	1,06
COMP_2.0_HCC38_28	0,00
COMP_2.0_HCC38_29	1,31
COMP_2.0_HCC38_30	1,41

Method 3.0 DEPArray	
Cell ID #	Qubit
COMP_3.0_HCC38_1	0,76
COMP_3.0_HCC38_2	0,55
COMP_3.0_HCC38_3	0,86
COMP_3.0_HCC38_4	0,76
COMP_3.0_HCC38_5	0,58
COMP_3.0_HCC38_6	1,09
COMP_3.0_HCC38_7	0,94
COMP_3.0_HCC38_8	2,34
COMP_3.0_HCC38_9	0,53
COMP_3.0_HCC38_10	0,57
COMP_3.0_HCC38_11	0,77
COMP_3.0_HCC38_12	0,78
COMP_3.0_HCC38_13	0,73
COMP_3.0_HCC38_14	0,96
COMP_3.0_HCC38_15	0,5
COMP_3.0_HCC38_16	0,8
COMP_3.0_HCC38_17	0,64
COMP_3.0_HCC38_18	0,64
COMP_3.0_HCC38_19	0,67
COMP_3.0_HCC38_20	0,74
COMP_3.0_HCC38_21	0,38
COMP_3.0_HCC38_22	0,32
COMP_3.0_HCC38_23	0,31
COMP_3.0_HCC38_24	0,43
COMP_3.0_HCC38_25	0,48
COMP_3.0_HCC38_26	0,51
COMP_3.0_HCC38_27	0,41
COMP_3.0_HCC38_28	0,43
COMP_3.0_HCC38_29	0,59
COMP_3.0_HCC38_30	0,39

The mean concentration within each method with the standard deviation, is shown in Table 12.

Table 12. Mean concentration (ng/µL) and standard deviation for each method.

Microinjection	Micromanipulator	FACS	DEPArray
8,19 ($^{+/-}$SD: 5,47)	0,02 ($^{+/-}$SD: 0,04)	1,48 ($^{+/-}$SD: 1,27)	0,68 ($^{+/-}$SD: 0,37)

The mean gives a quick view of how the concentration is for the methods, but to better visualize the distribution of the concentrations, boxplot and dot-plots were made as seen in Figure 19. This shows how the concentrations are compared to each other and how the variation is within each method.

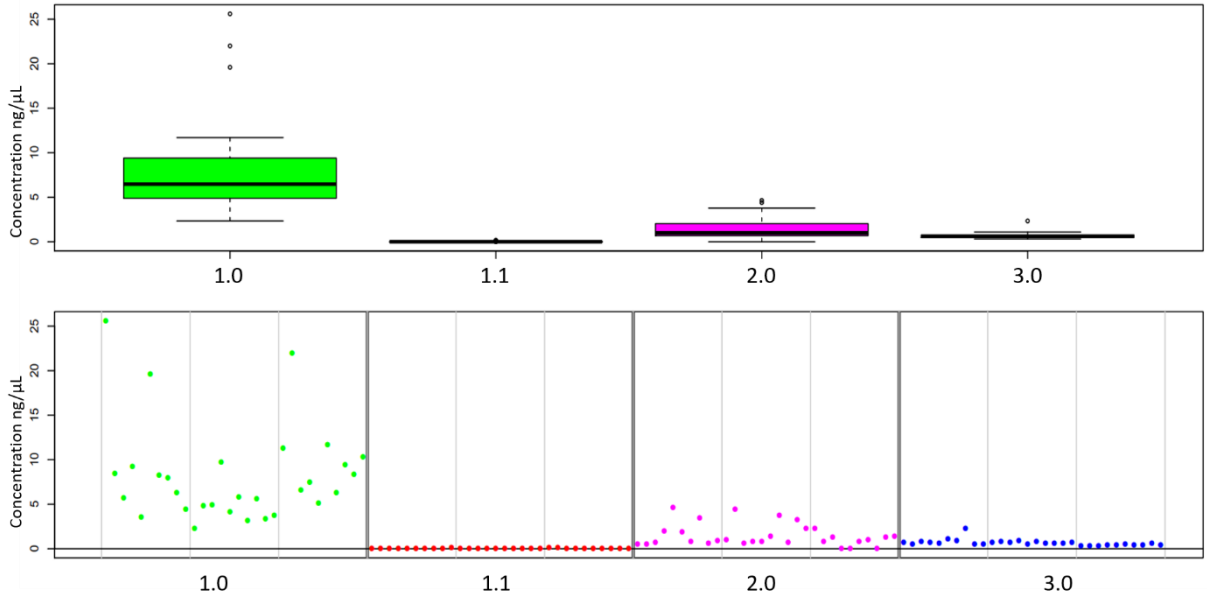

Figure 19. Plot of concentration distribution. The method number is written along the x-axis and the concentration is along the y-axis in ng/µL. Each method has a colour, microinjection method (1.0) is green, micromanipulator (1.1) is red, FACS (2.0) is pink and DEPArray (3.0) is blue. From the top, boxplot of the concentration distribution for each method. The dot plot shows each single cell concentration as a dot and the light grey lines within each method divides the ten single cells converted and amplified to cDNA in the same set-up.

The results visualized by the boxplot shows that the variation within the reference method is larger than the three other methods. This can be because the concentrations are higher with this method, and that three cells had particularly high concentration (three outliers above 15ng/µL). The FACS results also display cell-to-cell variation with concentrations almost up to 5ng/µL. The cells from the micromanipulator and DEPArray methods had either no measurable cDNA or very low concentrations, but the DEPArray boxplot shows that the mean is slightly higher as some samples had measurable concentrations.

5.2 Quality analysis of single cell cDNA

Fragment size and number of fragments of the cDNA from the samples was measured by Bioanalyzer. The gel electrophoresis separates the fragments by size and creates a curve based on number of fragment unit and the length of the fragments. When the RNA is of higher quality,

the fragments will be longer, and the number of fragments will increase with higher amount of RNA. Wide, high curve are a result of RNA good integrity and long fragments, while the curves on the lower end of the scale represent more degraded RNA with shorter fragments. For each run a size ladder was used as a standard, and a negative control or RNA control was analyzed along with the samples, from the same set-up in the SmartSeq2 protocol.

To show the results from the Bioanalyzer analysis, five selected examples from each method are presented in Figure 20, with their corresponding size ladders. The identification number for the samples is written above each graph.

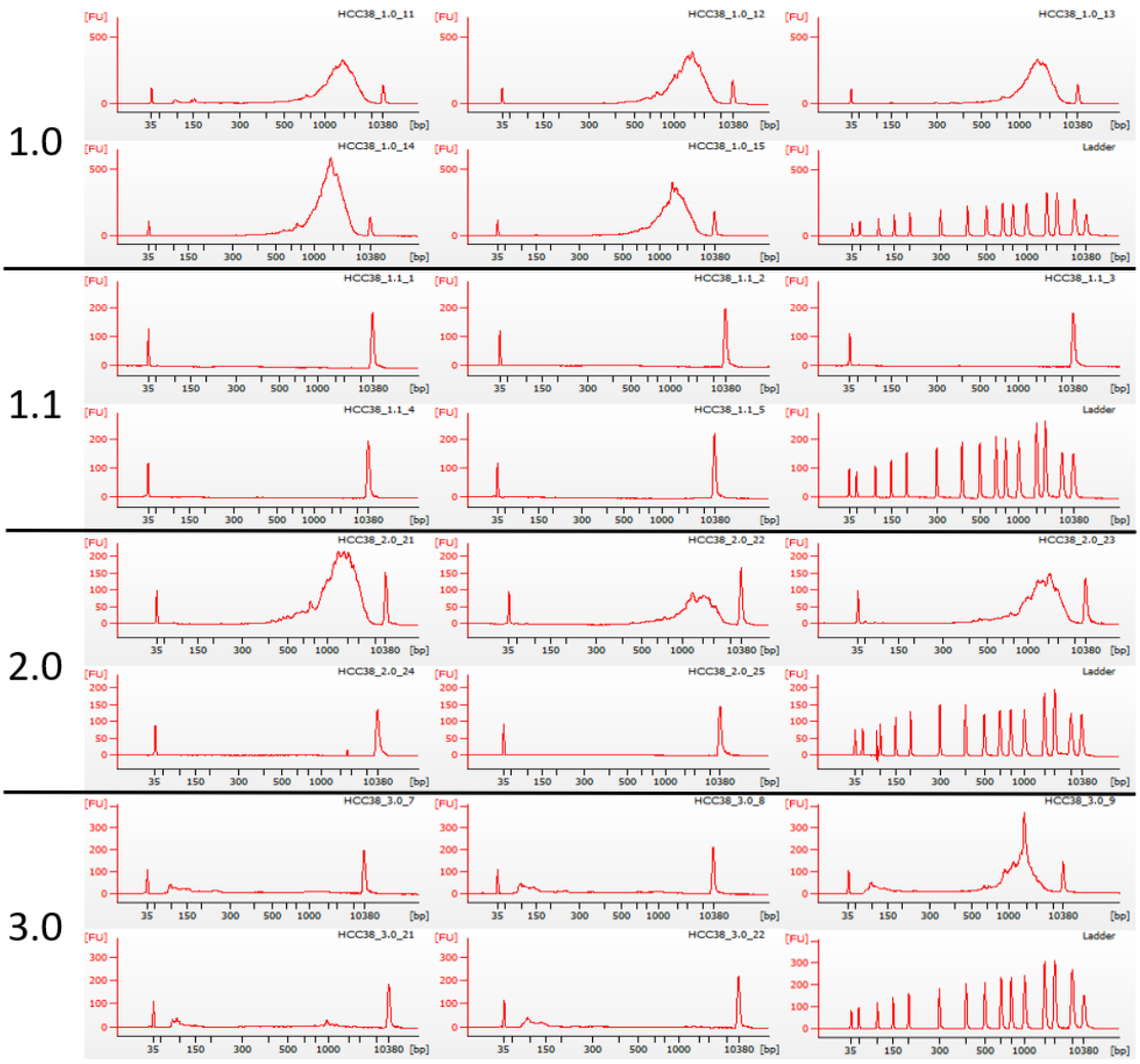

Figure 20. Example of Bioanalyzer results from selected samples from all methods with corresponding ladders. 1.0=microinjection method, 1.1=micromanipulator method, 2.0=FACS method and 3.0=DEPArray method. The X-axis represents the length of the fragments with number of basepairs (bp) and the Y-axis represent the number of fragment units (FU). The difference in the Y-axis with number of fragment units, show how the size from the different methods can be compared, which also makes the ladders smaller when the size range is higher.

Samples with high concentration (above 3 ng/µL) have been diluted down to around 2 ng/µL. Sample HCC38_3.0_8 and HCC38_3.0_9 was opposite, because of wrong pipetting in the well in the set-up.

The results show that samples from the microinjection method have a high and wide curve, with fragments above 1000 basepairs in length and up to 500 fragment units. This is an indication of good quality and integrity of the cDNA. Some of the samples from the FACS (2.0) method have the highest peaks compared to the other methods, with up to 200 fragment units and fragment length above 1000 basepairs. Most of the DEPArray (3.0) samples only have a smaller peak around 100 basepairs, but sample HCC38_3.0_9 (HCC38_3.0_8), have a high peak above 1000 basepairs. This corresponds to the concentration measured by Qubit. All the samples from the micromanipulator (1.1) method have flat curves, with no peaks, indicating no measurable fragments.

5.3 Gene expression analysis using ddPCR

ddPCR analysis, as described in 4.6.1 Digital Droplet PCR, was then used as a positive/negative test, to investigate if the samples had amplifiable cDNA and whether they expressed the genes *KRT18* or *CD45*. As mentioned previously, *KRT18* is a gene expressed by epithelial cells and *CD45* is a gene expressed by white blood cells. *KRT18* was tested in channel 1, with a signal from the FAM fluorophore if the cells were positive for the *KRT18* gene. *CD45* was tested in channel 2, with a signal from the HEX fluorophore if the cells were positive for the *CD45* gene.

The thousand droplets with their reaction mix and individual PCR reaction was counted and detection of whether the PCR amplification had taken place (positive) or not (negative). The more amplified copies of the target sequence, the stronger the amplitude of the fluorescent signal. A clear differentiation between the positive and negative droplets are important to set an appropriate threshold. Quantification of the partition of positive droplets to negative droplets, are used in Poisson statistics to determine the number of target templates in each sample. This can be used to calculate the concentration of cDNA without the use of a standard curve [87]. Therefore, all the samples had to be diluted to 0,125ng/µL before the PCR analysis, as described in 4.6.1.2 Optimization process, to have some negative droplets for the positive partition quantification. To compare the positive and negative reads, no template control (NTC) was used as a negative control and HCC2218BL single cell cDNA was used as a positive control for the *CD45* gene.

The mastermix had a high viscosity and even with 20% more volume of all the reagents, it was not enough for all the wells. The well for the no template control (NTC) was not prioritized and water was used for the remaining volume needed for droplet generation. This resulted in a lower background signal from the probes and the mastermix in the NTC and is why the negative signal from the NTC is lower than the negative signal from the samples. The threshold for positive signal should be set just above by the NTC, but as it is lower than the other negative samples, this could not be done with the current results. For all the experiments, there was a clear separation of the positive signal in the positive control and negative signals in the samples and NTC, the threshold was set for all the samples above the negative partitions.

5.3.1 Gene expression results from the microinjection method (1.0)

The results from ddPCR generated by droplet generation and droplet read of possible amplified gene copies are shown in Figure 21 and Figure 22. The gene assays for *KRT18* and *CD45* was performed in multiplex, with both assays in the same experiment.

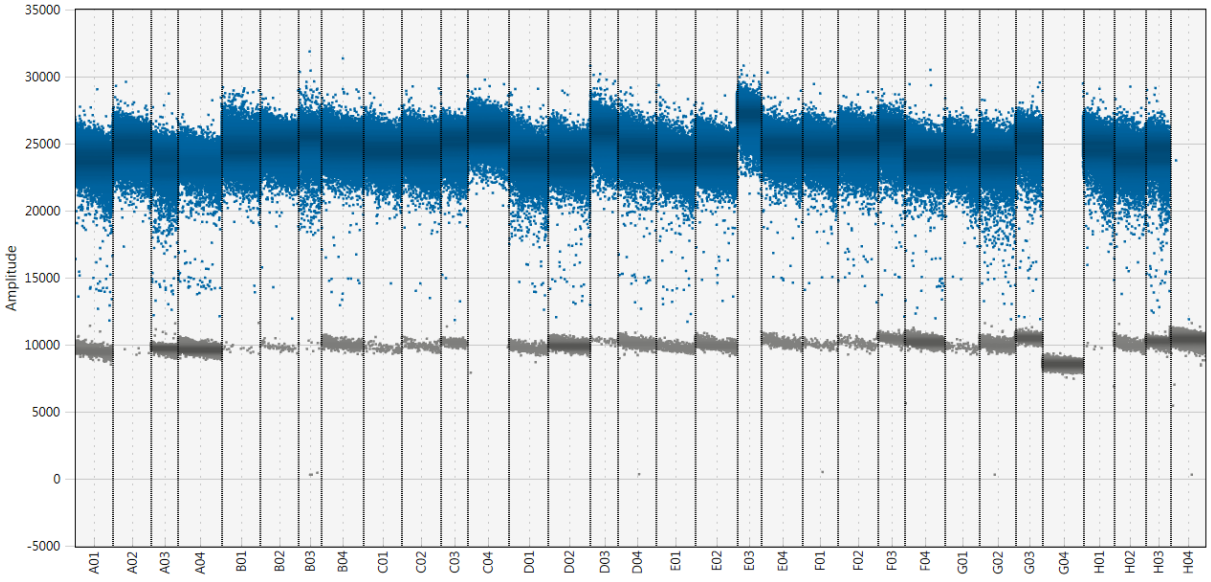

Figure 21. Results from droplet read of samples from method 1.0 in channel 1 with FAM for *KRT18*. The X-axis represent the samples A01-H04 corresponding to the plate set-up described in Table 15 in appendix 1. The Y-axis represent the amplitude of the fluorescent signal from the FAM probe for *KRT18*. The blue dots represent positive signal from amplified target sequences and the grey dots represent negative signal from no expression.

All the samples have a positive signal for *KRT18*, except G04 (no template control) and H04 (the HCC2218BL single cell cDNA). All the HCC38 single cells show high expression of the *KRT18* gene, with little variation.

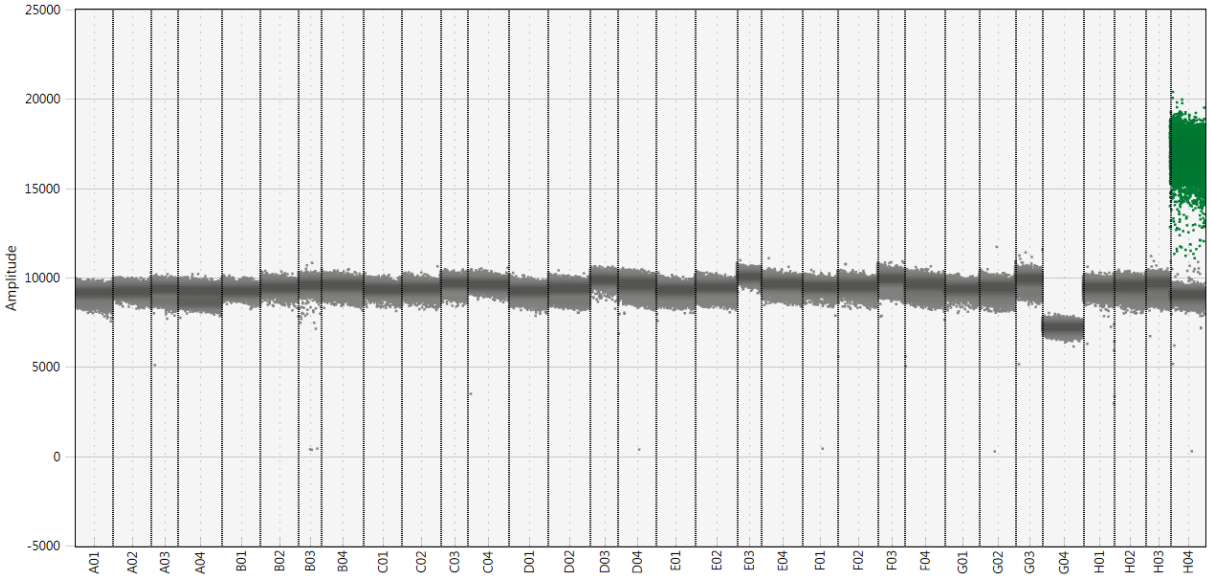

Figure 22. Results from droplet read of samples from method 1.0 in channel 2 with HEX for *CD45*. The X-axis represent the samples A01-H04 corresponding to the plate set-up described in Table 15 in appendix 1. The Y-axis represent the amplitude of the fluorescent signal from the HEX probe for *CD45*. The green dots represent positive gene expression of the *CD45* gene, and the grey dots represent negative signal from no expression.

The results of the *CD45* gene expression shows that only the HCC2218BL cDNA is positive with a green signal. This is expected as the cells from HCC38 cells should not express this gene and the HCC2218BL is used as a positive control.

The gene expression from all the cells tested in this plate set-up is shown in a 2D-plot in Figure 23.

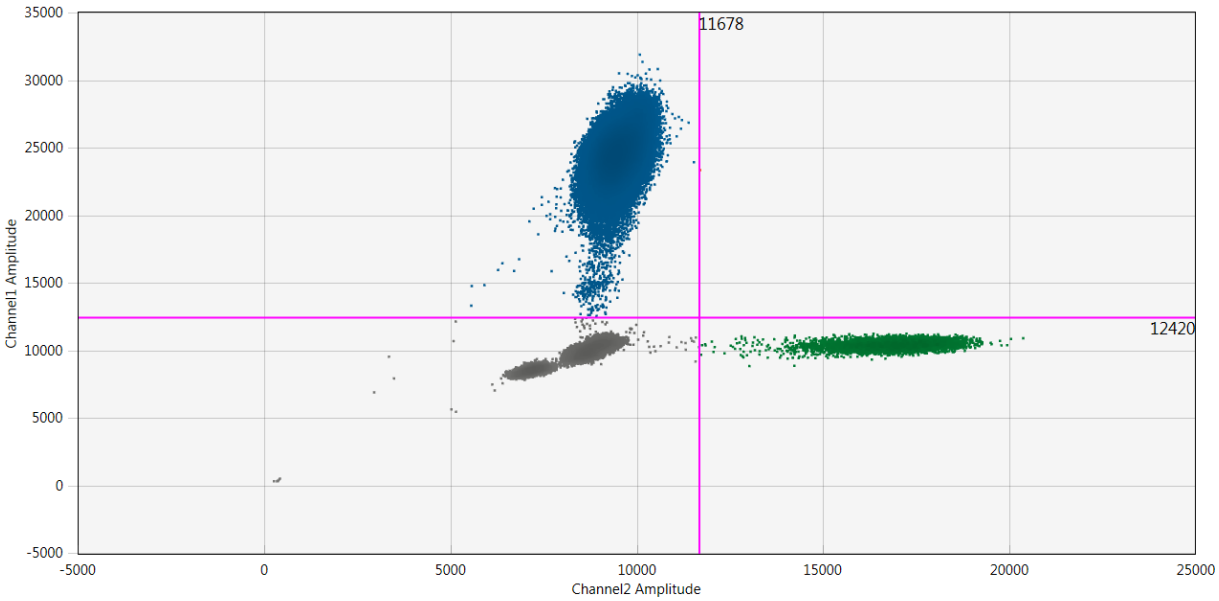

Figure 23. 2D plot of the gene expression from all the samples. The X-axis present the amplitude from channel 2 with HEX, and the Y-axis present the amplitude from channel 1 with FAM. The amplitude is higher for the Y-axis, which reaches 35000, and the X-axis reaches 25000 amplitude. The blue dots show positive expression of *KRT18* from the HCC38 single cells, the green dots show positive expression of the HCC2218BL single cell and the grey dots shows the negative droplets.

This shows that none of the HCC38 single cells expressed both genes, as this would have appeared as dots in the upper right quadrant. Such a 2D plot would be more applicable for patient samples to conclude their origin.

5.3.2 Gene expression results from the micromanipulator method (1.1)

Cells 1-10 and 19-30 from method 1.1 was analysed by the ddPCR as seen in Figure 24 and Figure 25. Not all the cells were tested because of low probe volume. The cDNA concentration from these samples were the lowest from the four methods (Table 11), so these samples were not prioritized. The number of droplets was lower for these samples, but the samples are positive if the number of positive droplets is three times higher than the NTC.

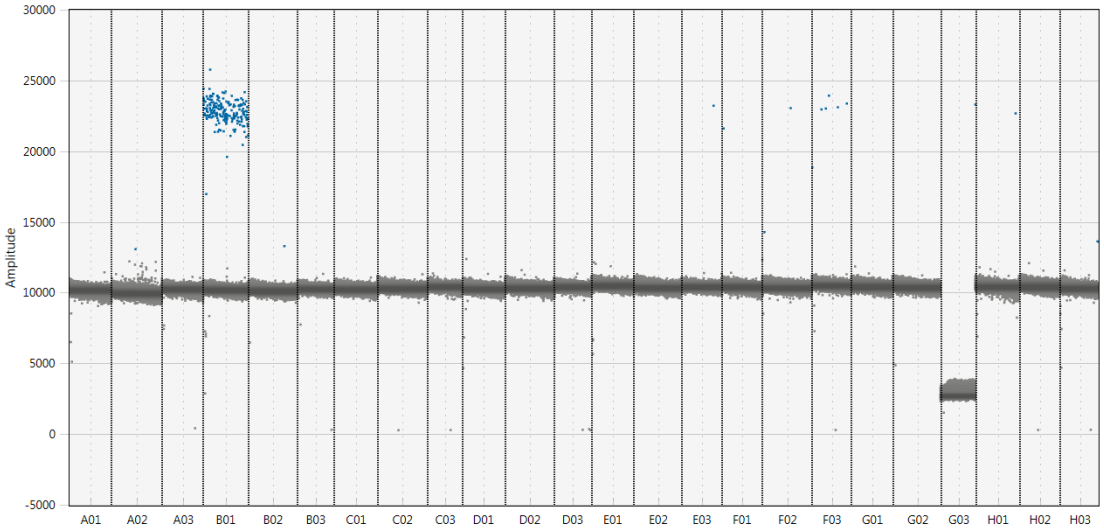

Figure 24. Results from droplet read of samples from method 1.1 in channel 1 with FAM for *KRT18*. The X-axis represent the samples A01-H04 corresponding to the plate set-up described in Table 16 in appendix 1. The Y-axis represent the amplitude of the fluorescent signal from the FAM probe for *KRT18*. The blue dots represent positive signal from amplified target sequences and the grey dots represent negative signal from no amplification.

The results show that some of the cells have some expression of the *KRT18* gene, but only in a few droplets per sample and a lower amplitude in the fluorescence compared to the reference method.

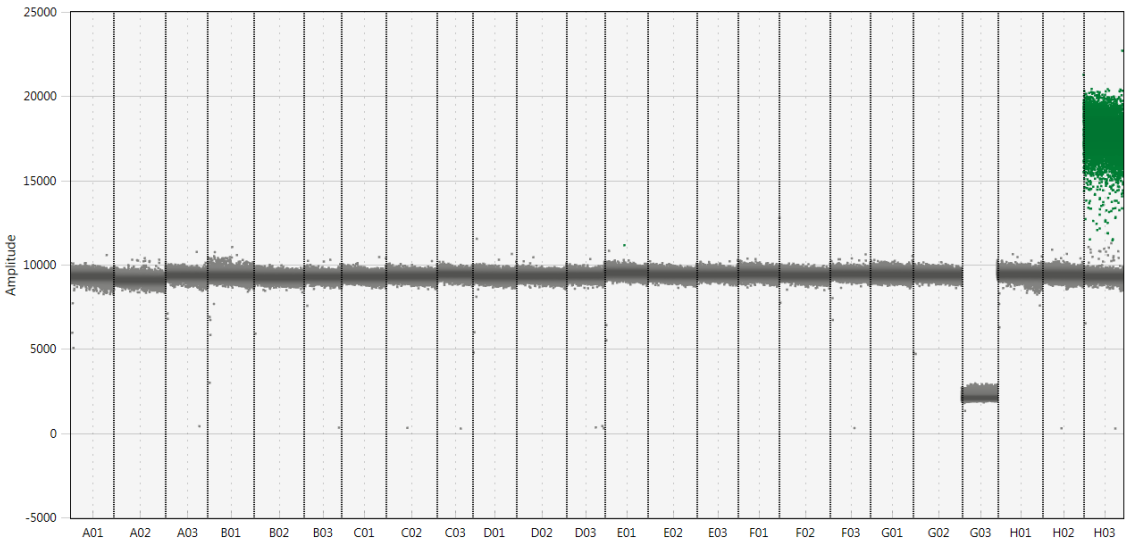

Figure 25. Results from droplet read of samples from method 1.1 in channel 2 with HEX for *CD45*. The X-axis represent the samples A01-H04 corresponding to the plate set-up described in Table 16 in appendix 1. The Y-axis represent the amplitude of the fluorescent signal from the HEX probe for *CD45*. The green dots represent positive gene expression of the *CD45* gene, and the grey dots represent negative signal from no expression.

The results of the *CD45* gene expression again shows that only the HCC2218BL cDNA is positive with a green signal. This is expected as the cells from HCC38 cells should not express this gene and the HCC2218BL is used as a positive control.

5.3.3 Gene expression results from the automatic FACS method (2.0)

The results from ddPCR of samples from the FACS method are shown in Figure 26 and Figure 27.

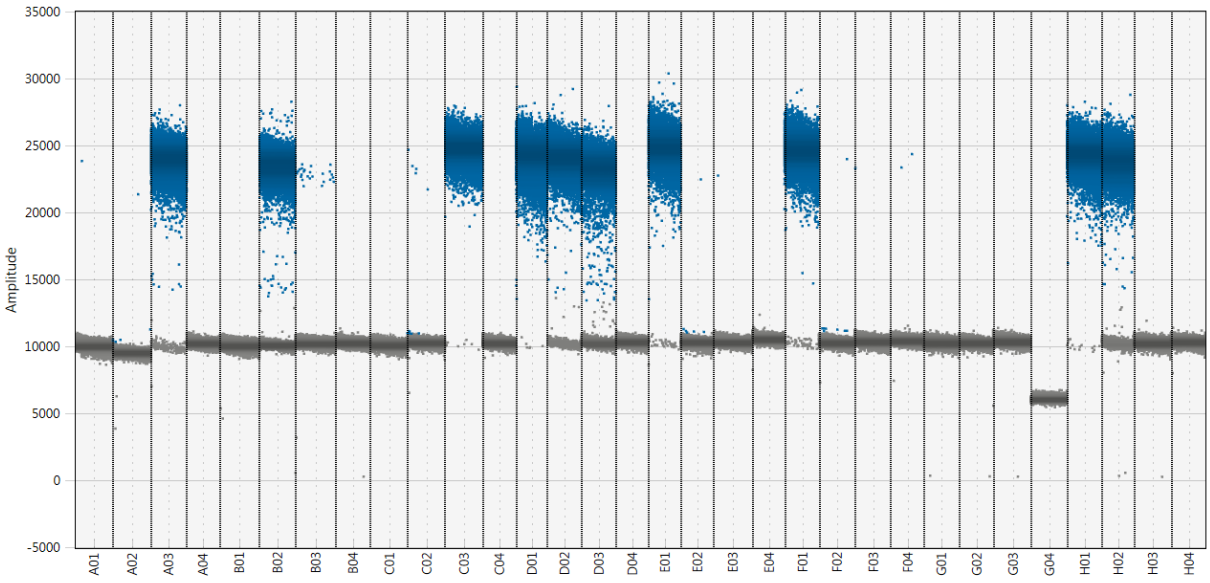

Figure 26. Results from droplet read of samples from method 2.0 in channel 1 with FAM for *KRT18*. The X-axis represent the samples A01-H04 corresponding to the plate set-up described in Table 17 in appendix 1. The Y-axis represent the amplitude of the fluorescent signal from the FAM probe for *KRT18*. The blue dots represent positive signal from amplified target sequences and the grey dots represent negative signal from no amplification.

The results show that 10 of the 30 single cells selected and isolated by the FACS method have a positive expression of the *KRT18* gene. This shows that 1/3 of the cells get cDNA of good quality and could possibly be used for further sequencing or other analysis of the transcriptome. However, with an output at only 1/3, the number of cells should be high in order to acquire enough cDNA from sufficient number of single cells.

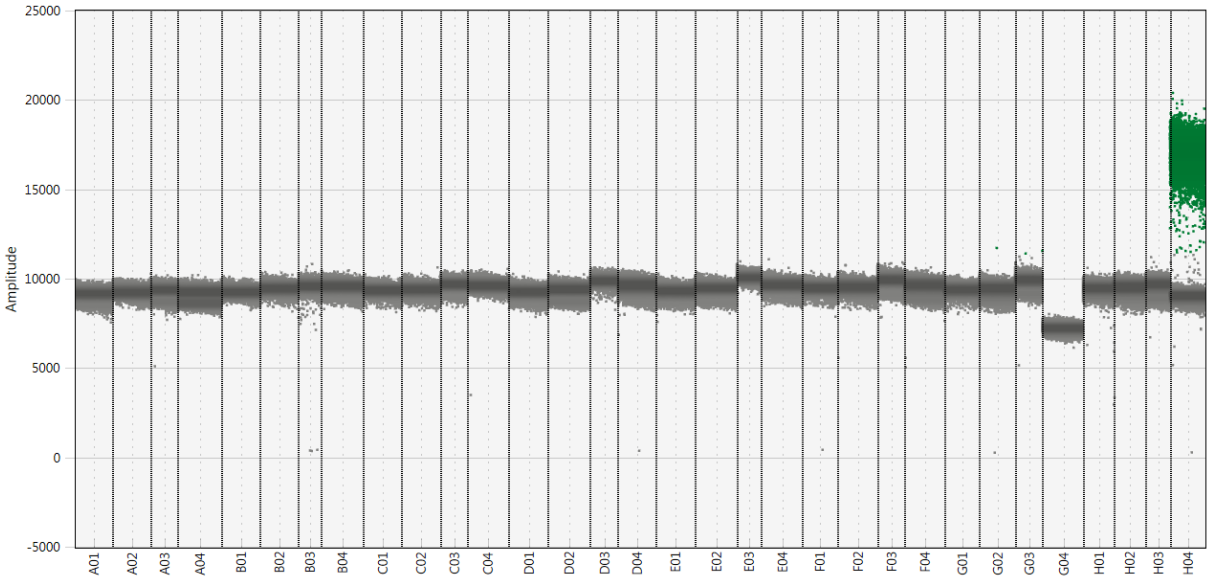

Figure 27. Results from droplet read of samples from method 2.0 in channel 2 with HEX for *CD45*. The X-axis represent the samples A01-H04 corresponding to the plate set-up described in Table 17 in appendix 1. The Y-axis represent the amplitude of the fluorescent signal from the HEX probe for *CD45*. The green dots represent positive gene expression of the *CD45* gene, and the grey dots represent negative signal from no expression.

The results of the *CD45* gene expression again shows that only the HCC2218BL cDNA is positive with a green signal. This is expected as the cells from HCC38 cells should not express this gene and the HCC2218BL is used as a positive control.

5.3.4 Gene expression results from the automatic DEPArray method (3.0)

The results from ddPCR gene expression of samples from the DEPArray method are shown in Figure 28 and Figure 29.

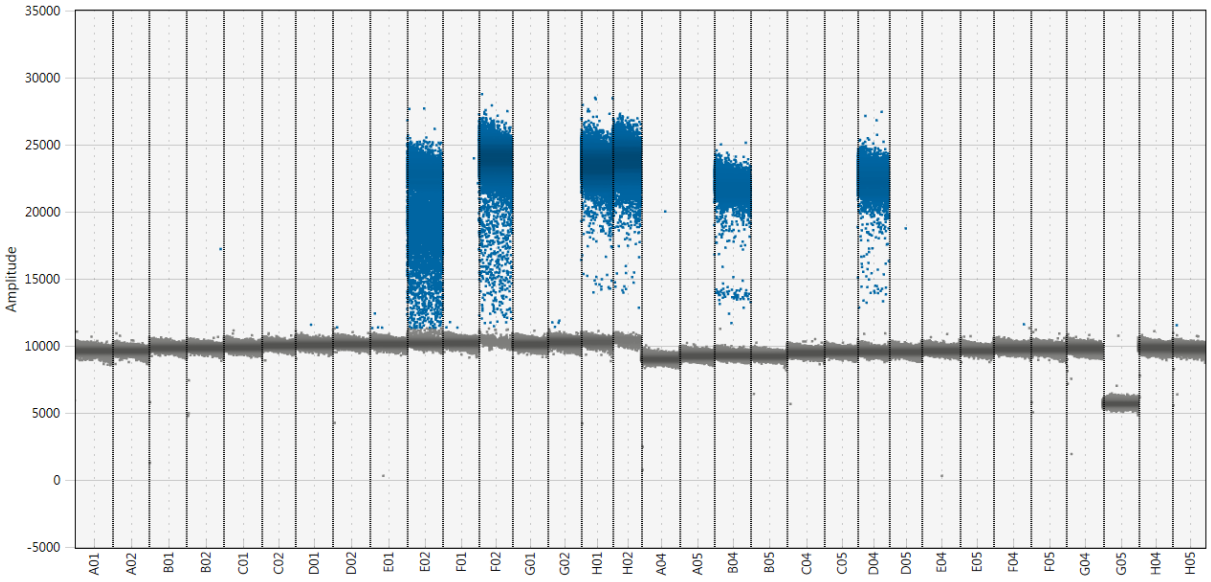

Figure 28. Results from droplet read of samples from method 3.0 in channel 1 with FAM for *KRT18*. The X-axis represent the samples A01-H04 corresponding to the plate set-up described in Table 18 in appendix 1. The Y-axis represent the amplitude of the fluorescent signal from the FAM probe for *KRT18*. The blue dots represent positive signal from amplified target sequences and the grey dots represent negative signal from no amplification.

The results show that six of the 30 samples identified and selected by the DEPArray method are positive. The number of positive droplets is lower for three of the samples, E02, B04 and D05 compared to H01, H02, B04 and D04. Two of the samples have some "rain" and investigated more closely in Figure 30.

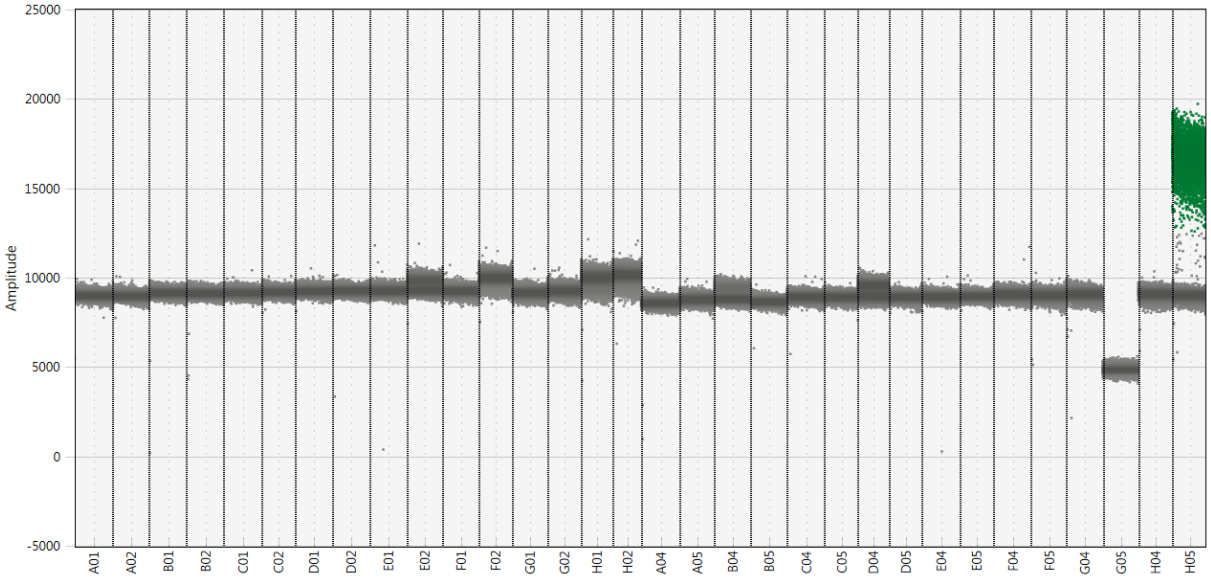

Figure 29. Results from droplet read of samples from method 3.0 in channel 1 with HEX for *CD45*. The X-axis represent the samples A01-H04 corresponding to the plate set-up described in Table 18 in appendix 1. Only the HCC2218BL single cell have a positive signal. The Y-axis represent the amplitude of the fluorescent signal from the HEX probe for *CD45*. The green dots represent positive gene expression of the *CD45* gene, and the grey dots represent negative signal from no expression.

The results of the *CD45* gene expression again shows that only the HCC2218BL cDNA is positive with a green signal. This is expected as the cells from HCC38 cells should not express this gene and the HCC2218BL is used as a positive control.

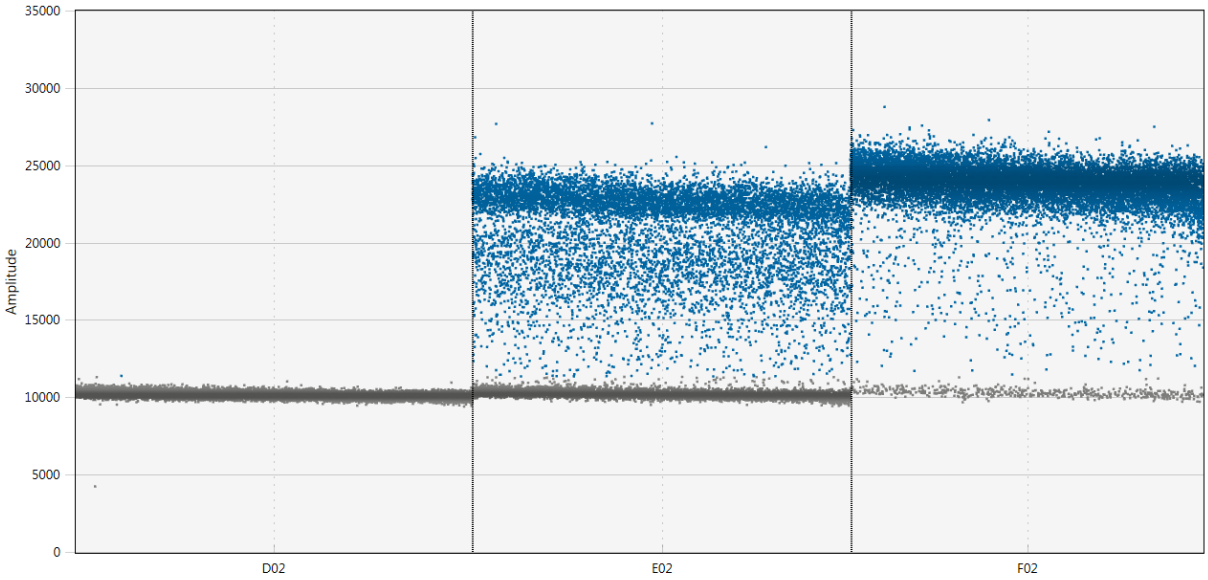

Figure 30. Results of D02-F02 from method 3.0 in channel 1 with FAM for *KRT18*. Closer view of D02, E02 and F02, with fluorescent amplitude on the Y-axis. The blue dots represent positive signal from amplified target sequences and the grey dots represent negative signal from no amplification.

Droplets with an intermediate amplitude of the fluorescent signal, ranging between the explicit positive (blue) and negative (grey) droplets, are called “rain” [88]. For these samples, the threshold had to be set manually, based on the clear negative droplets in the other samples. Usually, the NTC can be used to set this threshold, but in this project the NTC is lower than the negative droplets in the samples, as the NTC has less of the background signal from the mastermix. Sample E02 has a lot of "rain" compared to F02, while D02 is a negative sample. Note that sample F02 also has some “rain” but there is still a clear positive fraction of droplets around 25000 amplitude, which is at the same level of expression as the cells in the reference method. The reason for the “rain” is unclear, but it may be a consequence of a delayed PCR amplification in the droplets that represent the “rain”, with intermediate fluorescent signal, or a partial PCR inhibition in the individual droplets. Other reasons could be damaged positive droplets that get a lower amplitude in the fluorescent signal, or damaged negative droplets with background fluorescence, or a mix of both [88].

5.4 Patient sample concentration and gene expression analysis

The patient samples #1 and #2 was stained and selected by FACS into plates, but also into a small Eppendorf tube with culture medium, and from this some single cells were isolated by the microinjection pipette. For some of the single cells isolated by the microinjection pipette into an 8-well strip, cDNA was synthesized from RNA with the SmartSeq2 protocol. The concentration, fragment quality and origin of the cells were tested for an indication of how this method perform with old bone marrow samples from patients.

5.4.1 Concentrations of cDNA obtained from single cells from patient samples

The cDNA from the samples after the SmartSeq2 protocol, was measured by Qubit using HS DNA assay and is shown in Table 13.

Table 13. Concentration of cDNA from patient samples by Qubit (ng/µL).

Well	Patient	Qubit (ng/µL)
A1	LATE_2.0_1649_1	1,89
B1	LATE_2.0_1649_2	0,63
C1	LATE_2.0_1649_3	1,62
D1	LATE_2.0_1649_4	2,02
E1	LATE_2.0_2138_1	3,30
F1	LATE_2.0_2138_2	0,67
G1	LATE_2.0_2138_3	1,08
H1	RNA control	44,8
A2	NTC	0,66

The results show that the concentration from the single cells obtained from the patient samples are between 0,6-3,3 ng/µL. From the previous results in the methodology, samples with concentration above 1ng/µL, were shown to have a good integrity and show positive gene expression. The concentration of the RNA control and the NTC are included to show that the SmartSeq2 protocol have successfully converted and amplified the presentable RNA to cDNA.

5.4.2 Quality analysis of single cell cDNA from single cells obtained from patient samples

Fragment size and number of fragments of the cDNA from the patient samples was measured by Bioanalyzer and is shown in Figure 31.

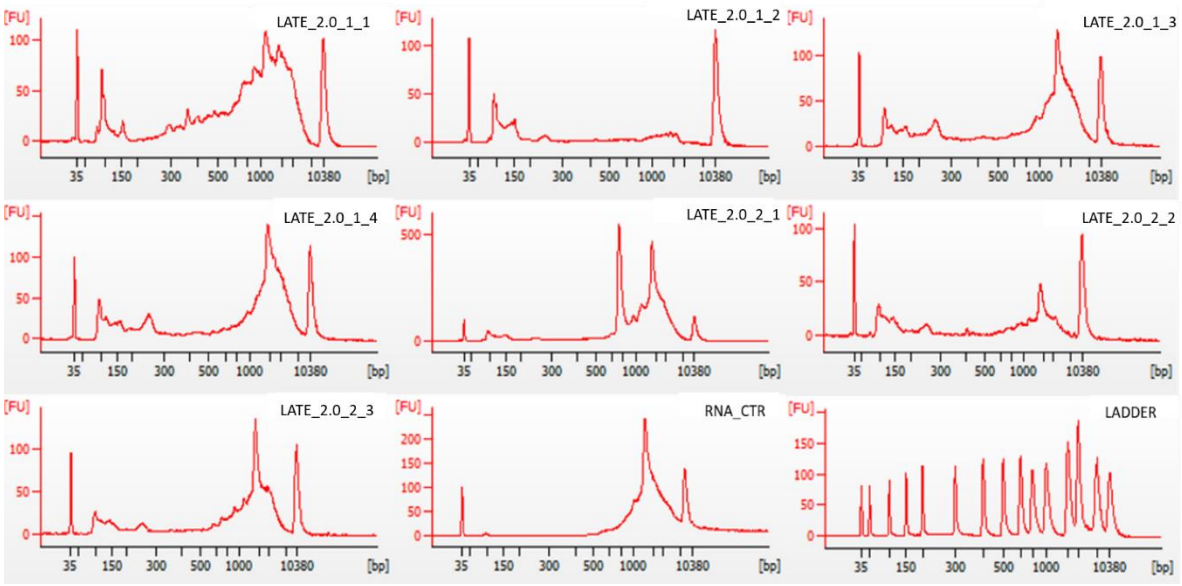

Figure 31. Quality analysis of single cell cDNA from single cells obtained from patient samples. The patient number and cell ID is written above each graph. The X-axis represents the length of the fragments with number of basepairs (bp) and the Y-axis represent the number of fragment units (FU).

Inspection of the results show variation between the samples. Three of the four samples from sample #1 have long fragments with wide peaks above 1000 basepairs, and up to 100 fragment units at the peak. This means the RNA fragments are of a good quality. The last sample from sample #1 has up to 50 fragment units that are shorter, which means the RNA is more fragmented and not in as good a quality. The first sample from #2 have two peaks of up to 500 fragment units, with fragment length around 1000 basepairs. This means it has RNA of high quality. The second sample from #2 have up to 50 fragment units that are above 1000 base pairs, and the third sample has more than 100 fragment units above 1000 base pairs. Overall, the RNA fragments output from the patient samples is of a relative high quality, with large fragments.

5.4.3 Gene expression analysis of single cells from patient samples

The gene expression analysis was performed to test the amplificability of the cDNA from the single cells obtained from the patient samples. The samples with the highest concentrations were tested, along with the RNA control, a positive control (HCC2218BL) and a no template control, as seen in Figure 32 and Figure 33.

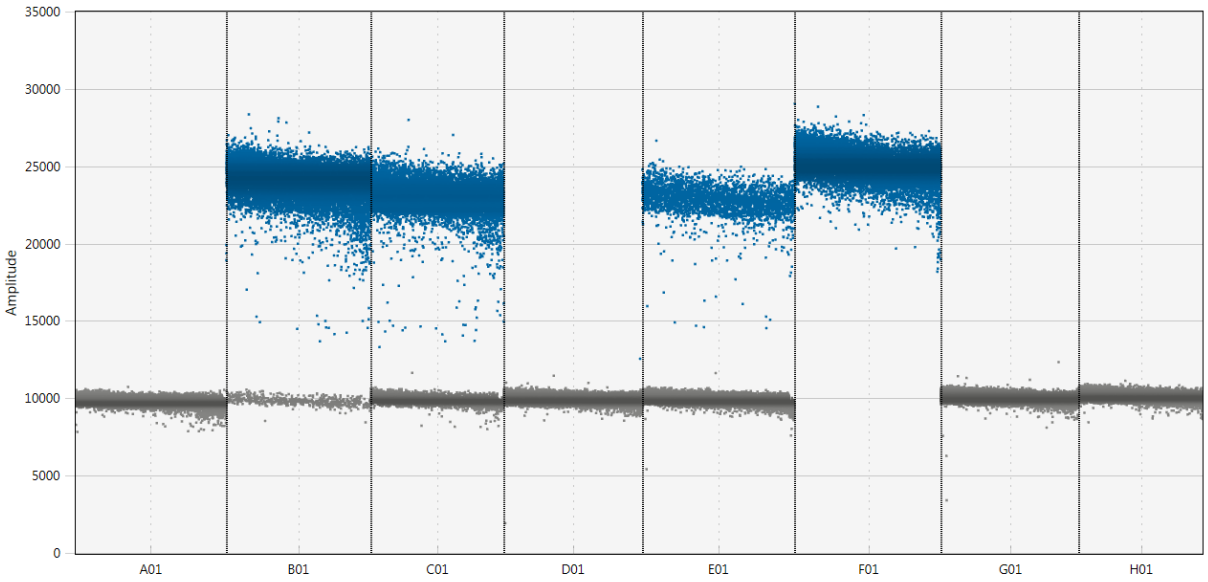

Figure 32. Gene expression by *KRT18* for patient samples. A01= LATE_2.0_1_1, B01= LATE_2.0_1_3, C01= LATE_2.0_1_4, D01= LATE_2.0_2_1, E01= LATE_2.0_2_3, F01= RNA control, G01= HCC2218BL positive control, H01= NTC. The samples were diluted down to approximately 0,125ng/µL. The blue dots represent positive signal from amplified target sequences and the grey dots represent negative signal from no amplification.

The results in Figure 32 show that two of the three cells from #1 and one of the two cells from #2 show positive expression of *KRT18*. The RNA control also shows positive expression, while the negative control and the HCC2218BL were negative.

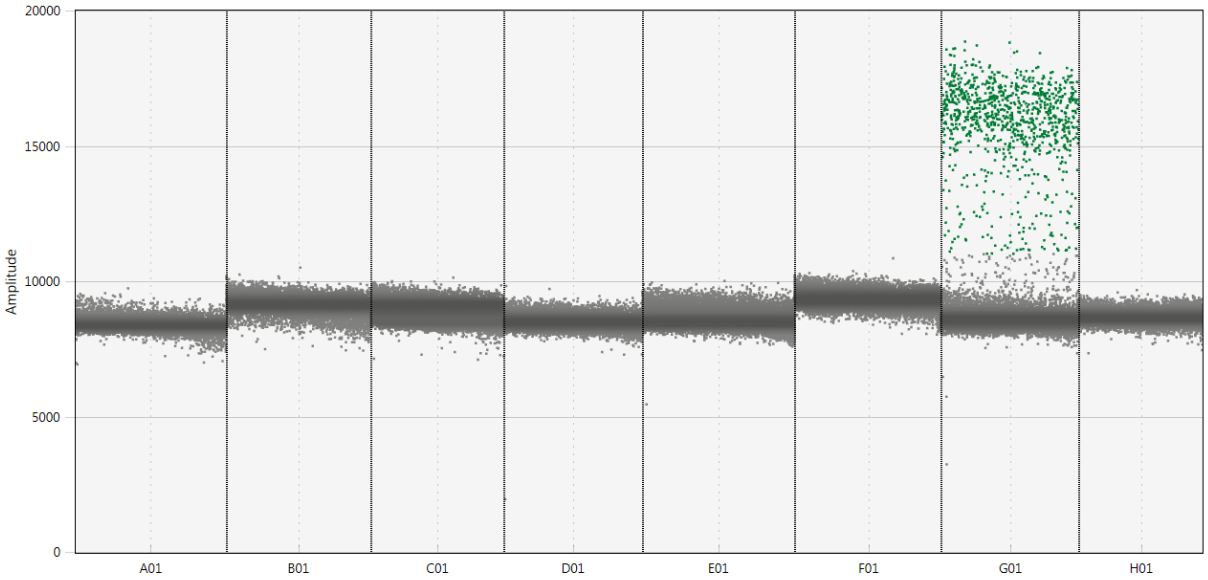

Figure 33. Gene expression of CD45 for patient samples. A01= LATE_2.0_1_1, B01= LATE_2.0_1_3, C01= LATE_2.0_1_4, D01= LATE_2.0_2_1, E01= LATE_2.0_2_3, F01= RNA control, G01= HCC2218BL positive control, H01= NTC. The samples were diluted down to approximately 0,125ng/µL.

The results of the CD45 gene expression shows that none of the samples express the CD45 gene, except the positive HCC2218BL positive control. This shows that all the single cells identified and selected from the patient samples do not have the CD45 gene. This is expected as the goal was that the single cells isolated were tumor cells.

5.5 Comparison of results

The comparison of number of cells with measurable cDNA concentration, fragment distribution, and positive expression of *KRT18* is presented in

Table ***14***, revised from Table 20, in Appendix 4. Master table.

Table 14. Comparison of method results for concentration of cDNA, fragment distribution and positive expression of *KRT18*.

	# of cells with measurable cDNA concentration (>1,0ng/μL)	**# of cells with visual larger fragments on Bioanalyzer**	**# of cells with positive expression of *KRT18***
Reference method (n=30)	30	29	30
Micromanipulator method (n=30)	0	0	(2) *
FACS method (n=30)	27	16	10 (12) *
DEPArray method (n=30)	2	7	6

*Only few droplets of positive gene expression, but more than three droplets, which is more than three times the negative control.

All the samples in the reference method had measurable cDNA concentration, all but one sample had many larger fragments (above 1000 bp), and all samples showed positive expression of *KRT18* by ddPCR. For the micromanipulator method none of the samples had measurable cDNA concentration above 1,0ng/μL, all had flat lines in the fragment analysis by Bioanalyzer and only two samples had a few positive droplets for positive *KRT18* expression. The two samples that had the few droplets with positive expression had 0,00ng/μL measured cDNA concentration. For the samples identified and selected by FACS, only three samples did not have measurable cDNA concentration, but not all the samples with measurable concentration had larger fragments by Bioanalyzer. The samples with concentration above 2ng/μL had many larger fragments and positive expression of *KRT18* by ddPCR. Twelve out of thirty samples had a positive expression by ddPCR. The samples from the DEPArray method had mostly cDNA concentration below 1ng/μL, and only one sample with 2,34ng/μL. Many of the DEPArray samples had peaks from shorter fragments (below 150bp), but some

had both a smaller peak and a high peak above 1000bp. Six of the samples had positive expression of *KRT18*, and the samples with the highest concentration had the strongest fluorescent signal. Two samples with concentration down to 0,4ng/μL also had positive expression.

6. Discussion

The aim of the work in this master thesis has been to test methods for single cell identification and selection, with respect to the quality and quantity of obtained RNA from the cells after the different enrichments and selection procedures. The microinjection pipette method has been used as a reference, with unstained and viable cells as close to physiological conditions as possible.

6.1 The impact of identification and selection procedures on single cell RNA

Several methodologies can be used for cell identification and selection. The methods tested in this methodology thesis, were selected based on availability and from other references in the field [89, 90]. All methods illustrate that identification and selection cannot be viewed as separate methods but are dependent on each other.

6.1.1 The identification process

Micromanipulation is a method that has been widely used for single cell isolation, as it in principle is a simple way of picking a selected cell and placing it in a PCR tube [89, 91]. In the micromanipulator method (4.3.2 Micromanipulation (method 1.1)), the identification process included fixation and staining with cytokeratin antibodies bound to epithelial cells and the blue/purple colour from alkaline phosphatase reacting with the bound antibodies. This staining enhances the tumor cells and for patient samples it would be possible to separate tumor cells from normal cells visually in a microscope. Cytokeratin staining has been widely used for detection of DTCs, often in combination with an antibody directed towards a leukocyte specific protein [50, 63]. The advantage with a cytoplasmic staining is that epithelial cells are easy to see, the disadvantage is that the cell membrane need permeabilization. By using FACS, the selection and sorting of tumor cells were based on variable fluorescent signals [71]. Before the automatic selection of the single cells with FACS, the cells were identified by staining with anti-EpCAM and anti-CD45. In contrast to cytokeratin, both of them are cell surface proteins. In addition, Hoechst 33258 was used as a live/dead cell marker. Before the methodology testing, the use of the Hoechst 33258 was tested, by single cells stained with and without Hoechst in addition to the other two antibody stains. The result from this preliminary testing is not

included, but the RNA concentration was measured, and the Hoechst staining did not affect the RNA amount or quality. Therefore, the staining process in this method, which does not include fixation of the cells, makes it possible to identify and select viable tumor cells among normal cells by fluorescence, as no visual identification is needed. This makes the process more automatized, un-biased and scalable. The disadvantage is that cells need to fall into a pre-defined gate, and no visual check of the selected cells can be performed.

The third method, DEPArray, is also more automatized as it sorts single tumor cells from an enriched cell suspension labelled with immunofluorescent antibodies [89]. Every cell is photographed, and in addition to the automatic detection of fluorescence levels, a visual inspection of each cell is performed. This secured both that doublets are avoided (i.e., two cells sorted as one) and that dubious staining are registered as such if cell morphology does not reflect a tumor cell. The challenge with DEPArray is the need for enrichment upfront, and the system is developed to be compatible with the cell identification system CellSearch. As the CellSearch system is dependent on having red blood cells present in the sample, the tumor cells were spiked into a fullblood sample, from a healthy donor, before identification and enrichment. This introduces a different starting point than the two other methods. The CellSave tubes for the fullblood sample also contain fixation elements, and the enrichment process by CellSearch fixated and stained the cells. In the CellSearch instrument, the cells are enriched in two steps; first by immunomagnetic beads coated with anti-EpCAM which extract out the cells with EpCAM surface proteins, followed by staining of anti-cytokeratin for identification and enumeration. It also uses DAPI, a blue fluorescent DNA staining, which is similar to Hoechst 33258 that was used for FACS staining. The processing time was longer in this method compared to the other two methods, with manual extraction of cells from cartridges between CellSearch and DEPArray, before sorting of single tumor cells by DEPArray.

6.1.2 The selection process

The three methods did not only differ with regards to identification method, but also type of selection and procurement of the cells. In particular time, number of procedure steps and temperature is of importance. The microinjection pipette has been used in the last step in the micromanipulator because the glass capillary used in the micromanipulator is too fragile to depose the single cells straight into a PCR strip. When the capillary uses the oil pump to suction the loosened single cells from the cytospin, it is not possible to be certain that the single cell

has been deposed from the capillary. With the microinjection pipette and the single cell on the separate petri dish, the accuracy and certainty of picking the single cell is much higher. The micromanipulator has been used on single cell isolation approaches by other researchers, and there it was concluded that the method involved laborious manual work with low output [92]. The use of the microinjection pipette alone on viable, unstained cells in suspension was an available option as the reference method, with the equipment and experienced help for demonstration at the beginning. With the equipment in the lab, the processing time from the cells in suspension kept on ice, could be low. The number of processing steps differed between the methods, where the reference method had the fewest steps, and DEPArray had the most. Many of the steps are manual, and this can infer a variation. Also, manual steps are time consuming and cannot easily be performed under stable temperature conditions. Newer technology includes pipetting robots, which have a higher precision, possibly faster procedure time, and can have chambers with temperature control, which all represents factors influencing degrading and quality of RNA.

6.1.3 Variation in RNA output from single cells

The success of retrieving RNA, and to be able to amplify it, showed a huge variation both between the methods, but also between the individual cells.

Most of the cells selected by the micromanipulation method showed non-detectable cDNA and thus a mean cDNA concentration of 0,02ng/µL, which is close to zero. This was reflected by almost no fragment units detected by the Bioanalyzer. Interestingly, the cDNA extracted from some of these cells were amplificable and showed a few droplets of positive expression of the *KRT18* gene, which means that the cells can be defined as epithelial cells. Still, the number of positive droplets and amplitude of fluorescence was much lower than seen from cells produced by the reference method. The method involves pre-treatment of the cells by cytospin preparation (including drying the cells and freezing at -80°C), fixation, as well as processing time. Several of these steps can be suspected to reduce the quality of the RNA in the cells compared to the reference method. For instance, air drying of the cells on the cytospin, and the freezing makes the cell membrane permeable, and leakage of RNA might occur. Further, a fixation in acetone followed by the intra cytoplasmic staining can possibly trap or interfere with the intracytoplasmic RNA molecules. The amount and quality of the RNA is so low from this method, that it would be challenging to have enough cells to actually get information from

some, and it would need more testing to know whether the RNA present represent the cell's global transcriptome or only a small, selected part of it.

Some of the single cells identified and selected via FACS did also not have a measurable cDNA level, but the 30 cells had a mean cDNA concentration of 1,48ng/µL. As we did not visualize the selected cells, some of the wells that had no measurable concentration, could be due to no single cell present after the selection. This reveals one of the limitations with FACS, some of the wells can be empty as cells can "fail" by missing the well, or land on the wall of the well. Each cell is sorted by FACS in a droplet with a volume of approximately 0,03µL, and this will quickly evaporate if the cell does not reach the lysis buffer at the bottom of the well. Therefore, time is important, as the plates must be spun down as quickly as possible after sorting, to get cells that might be stuck to the wall, down into the buffer. For the cells that have a cDNA concentration above 1ng/µL, there were many larger fragments obtained, as demonstrated by the quality analysis by Bioanalyzer seen in Figure 20. There was, however, still a difference in the quality of these fragments compared to the reference method, with lower number of fragment units. There was a difference between the cells from FACS on Bioanalyzer, where we observed that some had a "sharper" distribution curve than others. This curve could be due to the presence of fewer long fragments, and that some of those are preferentially amplified and thus visualized as a peak in the curve. Twelve (two had only few positive droplets) of the samples identified and selected by FACS, had positive expression of *KRT18*, as seen in Figure 26, which means that there could be potential RNA output. None of the samples had positive expression of *CD45*, which was expected as the cells were identified and selected from a suspension of only the tumor cell line cells.

The results from samples from the DEPArray isolation method was not as good as for FACS, and most had concentrations below 1,0ng/µL, with only one sample of 2,0ng/µL. The fragments were mostly shorter than the reference method and FACS, as seen in Figure 20 with the Bioanalyzer results, but one sample had a relatively good quality with a high peak up to 300 fragment units with a length above 1000 base pair. The concentration and quality of the RNA is higher from these samples compared to the micromanipulator, but lower than the FACS and reference method. Interestingly, six of the samples had positive expression of *KRT18*, as seen in Figure 28, but only two out them had concentrations above 1ng/µL, which means that there could be a potential RNA output from samples with lower concentrations. None of the cells had positive expression of *CD45*, which means that the selection of tumor cells only, from the mix of tumor and white blood cells works well. The reduced levels of RNA in these cells, could be

due to fixation, staining procedures during the enrichment by CellSearch as well as the many processing steps. It is also possible to speculate that the anti-cytokeratin can cause problems, as this is the same staining used in the micromanipulation method. Further testing and modification of this method can be feasible, as this has been the leading technology for single cell detection in diagnostics [92].

6.1.4 Rare single tumor cells from bone marrow samples

Bone marrow aspirates were available for two patient samples (#1 and #2) with known high numbers of DTCs (previously analysed in [44]). These were selected for identification and isolation of multiple DTCs by the FACS and micromanipulation method. After FACS, four of the single cells from sample #1 and three from sample #2 were further processed by converting RNA to cDNA by the SmartSeq2 protocol. All cells had measurable concentrations and various number of cDNA fragments, as seen in Table 13 and Figure 31. The cells providing the highest concentration were analysed for gene expression with ddPCR. Interestingly LATE_2.0_1_1 and LATE_2.0_2_1, both had measurable concentrations and a proper fragment distribution, but as seen in Figure 32 do not express either the *KRT18* or the *CD45* gene. The reason for this is unknown. The rest of the samples had positive expression of the *KRT18* gene and could be concluded as epithelial tumor cells. This is an interesting result, as the FACS then had identified and successfully selected the rare DTCs in the bone marrow samples.

Although the results from FACS were quite successful, some limitations must be considered when using it for identification and sorting of patient samples. The FACS instrument needs a high number of DTCs to set a gate identifying viable tumor cells and would not be applicable for patients with few DTCs. The patient samples tested here (3.2 Patient samples), had more 765 and 820 DTCs per two million normal cells. Further testing should be done to find a cut-off for the number of tumor cells that can be selected by FACS. The sample must also be free of clots, and this was difficult for one of the 20-year-old samples, and this thus need more optimizing and testing. If the instrument can sort the cells, there are still some important steps to do, including centrifugation as quickly as possible to prevent the cells that might be along the wall of the wells to dry out. This might be what happened to the three cells with no output from FACS in the methodology testing. The selection and identification method for CTC extraction from blood samples with DEPArray and enrichment by CellSearch, has been used by many researchers, and is an FDA approved method. It is not common to use it for DTC

identification and selection. To test the performance on bone marrow samples, further investigation must be done where the cell line will be spiked into a bone marrow sample as well as red blood cells.

6.2 Technological limitations and considerations

This work tested three different methodologies for single cell identification and selection, using cell lines and a few patient samples. Tumor cells from a cell line were chosen as the material for testing. These cells could be different from the patient samples that the methods would ultimately be used for. Cells from a cell line are, however, useful for testing the experimental set-ups as they have been produced to be able to grow and divide under *in vitro* conditions, as well as surviving thawing and freezing procedures. This makes it possible to produce enough cells, assumed to be similar, to perform many experiments allowing thorough comparisons of the different methodologies. On the other hand, the cell line cells might be more robust and possibly more sustainable to "survive" identification and selection procedures compared to single cells obtained directly from patients. It was therefore important to perform a final testing of a few patient samples, to get an impression of the applicability of some of the methods to actual patient samples.

Other single cell isolation methods that could have been used are for instance laser capture microdissection (LCM), microfluidics, Parsortix and Rare cyte. Of these, LCM would have been possible to test as several types were available in Oslo, but these techniques would need the same identification process (i.e., cytokeratin staining) as the micromanipulator method used here. We therefore postponed testing of LCM. Other technologies have not been accessible for this work.

6.2.1 Microinjection method as a reference

The manual microinjection method was performed without any staining of the cells and as close to physiological conditions as possible. This method had the shortest processing time, and the cells were kept on ice and then dry ice to prevent RNA degradation. Therefore, it was used as a reference for the isolation of cell line cells. It had high amount and quality of RNA, and as seen in Table 11, measurable cDNA concentrations for all single cells with this method. However, since there are no cell selection and identification processes, this method could not

have been used on a patient sample, as it would not be possible to identify the few tumor cells among the million normal cells in a sample.

Some cells had a concentration of approximately double the mean, which could be because two single cells were picked, as the method is manual and could, despite best efforts, be faulty. The quality of the RNA is also determined as satisfying, by inspecting the results from the Bioanalyzer. The peaks from the bioanalyzer are wide and on the higher end of the scale from 1000 basepairs in length, which means there are many large fragments. The analysis of the *KRT18* gene expression, a gene encoding cytokeratin in epithelial cells, show high expression of the epithelial gene, which confirms that the selected cells were tumor cells, as was expected with the use of the HCC38 tumor cell line [93]. None of the cells have expression of *CD45*, which is expected as these tumor cells do not express this gene. If the samples showed expression of both KRT18 and CD45, it would be a third population in the 2D plot show in Figure 23. This would be useful information when using patient samples, were the risk of isolating a normal cell instead of a tumor cell, or both a tumor cell and a normal cell, is higher.

6.2.2 Choice of cDNA conversion and amplification protocol

The conversion and amplification of the RNA with the SmartSeq2 protocol is common for all the methods. For an improved pipeline, this protocol could be done automated on a pipetting robot. With a more automatic pipeline for the reverse transcription and conversion of the RNA to cDNA, some uncertainties of faulty pipetting can be eliminated. The SmartSeq2-protocol was chosen for this master thesis based on previous comparison analysis of single cell RNA sequencing methods, by Ziegenhain (2017). They found that the SmartSeq2 protocol had the highest sensitivity and detected the most genes per cell and across cells [94]. The SmartSeq2 protocol has many manual pipetting steps, with small sample volumes and the need of high precision. In this work the RLT lysate buffer was used, and this buffer destroy enzymes used in the SmartSeq2, which is why the additional first step of sample clean-up was performed. For this step, the RNA Clean beads were used to magnetically keep the RNA from the samples while the RLT buffer was removed by pipette, and the remains of the buffer were washed away with 80% ethanol.

6.2.3 Variation in quality control and ddPCR analysis

The Qubit HS DNA assay was chosen for measurement of concentration of the cDNA in the samples, because it has low variation and high accuracy for low dsDNA concentrations. This was tested initially, with parallels before measuring the samples. For the samples with concentration below the measurable limit, the sample input was increased to 5µL to try to measure the lowest concentrations, down to the sample with 0,02ng/µL.

The analysis on Bioanalyzer gave some failed experiments, with no results and needed to be re-run, due to unknown reasons. It had some differences between samples from the same method from FACS as seen in the different sharpness of the curves in Figure 20, which leaves some uncertainty with these results. Small volumes and risk for faulty pipetting also gives some uncertainty and room for errors in the manual operation of this analysis. Tapestation by Agilent is a different instrument for analysis of fragment size and integrity which might have been more appropriate. This was used in the initial testing, but delayed deliveries from the producer, resulted in the use of the Bioanalyzer analysis instead, which was used by Picelli et. al. in their SmartSeq2 protocol.

The ddPCR analysis was used to test the amplificability of the RNA from the samples and at the same time served as a cell-type assay. For the ddPCR, slow pipetting during droplet generation was of importance, as the droplets could be damaged or left behind in the generation cartridge. This was taken into consideration and attempted endeavoured by cautious pipetting. There are automatic droplet generators that could eliminate the possible user differences [95]. The possibility of damaged or lost droplets due to manual droplet generation could have affected the results some, but this analysis was used as a positive/negative test, and the results show a clear distinction between positive and negative gene expression.

6.3 Pending results

For the additional analyses of 10X Genomics with bulk sequencing of single cells by FACS and from suspension, the results are pending from the Genomics Core facility, Norwegian Radium Hospital. The sequencing results of patient sample #2 that was sent to the Sanger Institute is also pending, as the analysis time is long and sequencing queues at Sanger are extensive due to the Covid-pandemic.

6.3.1 10X distribution of RNA loss in the transcriptome

By the 10X Genomics method, we will provide sequence data of unstained, viable single cells from suspension, and stained cells that have been sorted through the FACS instrument. The 10X sequencing of the transcriptome of two parallels from each of these methods can provide insight into how the distribution of the RNA loss could be across the transcriptome. The questions regarding which genes are lost, or affected by FACS sorting, can help decide if this method can be used to further understand how the dormant DTCs are activated. This could help decide if the RNA from single cells selected by FACS can be used for sequencing, which gives an easier pipeline for sorting of many single cells automatically.

6.3.2 Sequencing of patient samples by the Sanger institute

One of the patient samples' single cells and corresponding primary tumor, have been sent to the Sanger institute. The single cells were identified and selected by FACS and the micromanipulator methods but results from the sequencing are not finalized. At the Sanger institute, the Smartseq2 protocol is automated, and the results could give an indication of how much this affects the cDNA output. The hope is that the results will show how the transcriptome is in the DTCs in the bone marrow, compared to the tumor cells in the primary tumor. With single cell transcriptomes the heterogeneity and cell-to-cell variation can be clearer than from a population of tumor cells sequenced by bulk sequencing.

For a majority of breast cancer patients, the number of cancer cells in bone marrow are very low, and the FACS method might not be possible.

7. Conclusion

The five-year relative survival for breast cancer patients in Norway today is increasing, but there are still patients that experience relapse many years after diagnosis and successful surgery. The status of disseminating tumor cells present in the bone marrow can be used to predict the survival and treatment response in breast cancer patients. Study of single DTCs transcriptomes could help understand how they are activated and the potential to metastasis, which can result in a relapse of the disease.

In this work we found a huge variation in the RNA amount and quality obtained from single tumor cells. The difference was related to the identification and selection methods, but we also found cell-to-cell variation within each method. The three methods tested have several differences, both with regards to cell staining, time, and number of procedures.

The methodology testing showed that FACS was the most successful method for identification and selection og single cells. Here we found that the majority of the cells had measurable cDNA concentration, but only 1/3 had positive gene expression of cytokeratin. However, for this method to work successfully, the initial number of tumor cells needs to be high, in order to separate the rare tumor cells from the normal cells in the sample. DEPArray could also be a possible method, but this needs further testing. This would include improving the CellSearch enrichment and staining process, to keep RNA at a good quality. The micromanipulator method does not provide RNA output in a high enough amount to measure, but in this method the single cells picked was easier to manually distinguish as tumor cells. This work shows that RNA can be preserved during several steps of single tumor cell identification and selection, but further testing would be beneficial to increase the success rate.

7.1 Future aspects

Previous studies have used single cells from bone marrow samples isolated by the micromanipulator to study the single cell genomes [50]. Therefore, the DNA is presumed to have a good quality, even though the RNA did not. For the methods where the RNA was measurable, the possible DNA output is presumed to be of better quality as DNA is more stable than RNA.

A combined pipeline for studying both the genome and transcriptome from single cells could give many answers about unique circulating tumor cells. This have been tested, and the RNA

and DNA are first separated and then amplified and sequenced in parallel [96]. The G&T pipeline require enough and high-quality RNA, and so far, the sorting of DTCs by FACS will be the preferred method. This means that clinical samples with many DTCs in the bone marrow will be analysed, and those with fewer cells will be spared until the protocols are optimal. Still, it will be of major interest to now be able to study both the genome and the transcriptome of these cells residing in breast cancer patients bone marrow years before they have a detectable metastasis.

References

1. Crick, F., *Central Dogma of Molecular Biology.* Nature, 1970. **227**(5258): p. 561-563.
2. James Watson, T.B., Stephen Bell, Alexander Gann, Michael Levine, Richard Losick, Stephen Harrison, , *Molecular Biology of the Gene*. Seventh ed. 2014, Cold Spring Harbor, New York.
3. Bruce Alberts, A.J., Julian Lewis, David Morgan, Martin Raff, Keith Roberts, Peter Walter, , *Molecular Biology of The Cell*. 6th ed. 2015.
4. National Cancer Institute. *What is cancer?* 2015 01.12.2020]; Available from: https://www.cancer.gov/about-cancer/understanding/what-is-cancer.
5. Weinhold, B., *Epigenetics: the science of change.* Environ Health Perspect, 2006. **114**(3): p. A160-7.
6. Hanahan, D. and Robert A. Weinberg, *Hallmarks of Cancer: The Next Generation.* Cell, 2011. **144**(5): p. 646-674.
7. Klein, C.A., *Parallel progression of primary tumours and metastases.* Nat Rev Cancer, 2009. **9**(4): p. 302-12.
8. Lucia, M.S. *CTC/DTC for Prognostic Value for Prostate Cancer*. 2015 31.03.2021]; 09.01.2015:[Available from: https://grandroundsinurology.com/prostate-cancer-m-scott-lucia-ctc-dtc-prognostic-value/.
9. Fares, J., et al., *Molecular principles of metastasis: a hallmark of cancer revisited.* Signal Transduction and Targeted Therapy, 2020. **5**(1): p. 28.
10. Fidler, I.J., *The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited.* Nat Rev Cancer, 2003. **3**(6): p. 453-8.
11. Russnes, H.G., et al., *Insight into the heterogeneity of breast cancer through next-generation sequencing.* J Clin Invest, 2011. **121**(10): p. 3810-8.
12. Navin, N., et al., *Tumour evolution inferred by single-cell sequencing.* Nature, 2011. **472**(7341): p. 90-94.
13. Yuan, S., R.J. Norgard, and B.Z. Stanger, *Cellular Plasticity in Cancer.* Cancer Discov, 2019. **9**(7): p. 837-851.
14. da Silva-Diz, V., et al., *Cancer cell plasticity: Impact on tumor progression and therapy response.* Seminars in Cancer Biology, 2018. **53**: p. 48-58.
15. Yang, X., et al., *Cellular Phenotype Plasticity in Cancer Dormancy and Metastasis.* Front Oncol, 2018. **8**: p. 505.
16. Johns Hopkins University. *Overview of the breast*. 2021; Available from: https://pathology.jhu.edu/breast/overview.
17. The American Society of Breast Surgeons Foundation. *Breast Anatomy*. 2021; Available from: https://breast360.org/topic/2017/01/01/breast-anatomy/.
18. Cleveland clinic. *Breast anatomy*. 2021 14.10.2020; Available from: https://my.clevelandclinic.org/health/articles/8330-breast-anatomy.
19. Kreftregisteret. *Cancer Statistics*. 2020 13.11.2020; Available from: https://www.kreftregisteret.no/en/The-Registries/Cancer-Statistics/.
20. Cancer registry of Norway, *Cancer in Norway 2019*. 2020: Oslo.
21. Kreft registeret. *Brystkreft*. 2021 21.01.2021; Available from: https://www.kreftregisteret.no/Temasider/kreftformer/Brystkreft/.
22. Kreftforeningen. *Brystkreft*. 2020 07.01.2021; Available from: https://kreftforeningen.no/om-kreft/kreftformer/brystkreft/#utbredelse-og-overlevelse.
23. Breastcancer.org. *Breast Cancer Risk Factors*. 2020 25.06.2020; Available from: https://www.breastcancer.org/symptoms/understand_bc/risk/factors#:~:text=%20Risk%20factors%20you%20can%20control%20%201,risk.%20The%20American%20Cancer%20Society%20recommends...%20More.

24. Lesk, A.M., *Introduction To Genomics*. Third ed. 2017, Great Clarendon Street, Oxford: Oxford University Press.
25. Institute, N.C. *BRCA Gene Mutations: Cancer Risk and Genetic Testing*. 2020 19.11.2020; Available from: https://www.cancer.gov/about-cancer/causes-prevention/genetics/brca-fact-sheet.
26. Polyak, K., *Breast cancer: origins and evolution.* J Clin Invest, 2007. **117**(11): p. 3155-63.
27. Fantozzi, A. and G. Christofori, *Mouse models of breast cancer metastasis.* Breast Cancer Research, 2006. **8**(4): p. 212.
28. Feng, Y., et al., *Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis.* Genes Dis, 2018. **5**(2): p. 77-106.
29. Ottesen, G.L., *Carcinoma in situ of the female breast. A clinico-pathological, immunohistological, and DNA ploidy study.* APMIS Suppl, 2003(108): p. 1-67.
30. The American Cancer Society. *Types of breast cancer*. 2019 20.09.2019; Available from: https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/types-of-breast-cancer.html.
31. Johns Hopkins University. *Staging & Grade*. 2021 13.04.2021]; Available from: https://pathology.jhu.edu/breast/staging-grade/.
32. Iqbal, B. and A. Buch, *Hormone receptor (ER, PR, HER2/neu) status and proliferation index marker (Ki-67) in breast cancers: Their onco-pathological correlation, shortcomings and future trends.* 2016. **9**(6): p. 674-679.
33. Helsedirektoratet. *Nasjonalt handlingsprogram med retningslinjer for diagnostikk, behandling og oppfølging av pasienter med brystkreft*. 2020 14.08.2020 20.04.2021]; Available from: https://www.helsebiblioteket.no/retningslinjer/brystkreft/forord.
34. Sørlie, T., et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(19): p. 10869-10874.
35. Breastcancer.org. *Molecular subtypes of breast cancer*. 2020 25.06.2020; Available from: https://www.breastcancer.org/symptoms/types/molecular-subtypes.
36. Fougner, C., et al., *Re-definition of claudin-low as a breast cancer phenotype.* Nature Communications, 2020. **11**(1): p. 1787.
37. Wallden, B., et al., *Development and verification of the PAM50-based Prosigna breast cancer gene signature assay.* BMC Med Genomics, 2015. **8**: p. 54.
38. Parker, J.S., et al., *Supervised risk predictor of breast cancer based on intrinsic subtypes.* J Clin Oncol, 2009. **27**(8): p. 1160-7.
39. Bordvik, M., *Sykehusene får ta i bruk ny brystkreft-test*, in *Dagens Medisin*. 2019.
40. Rakha, E.A., et al., *Breast cancer prognostic classification in the molecular era: the role of histological grade.* Breast Cancer Res, 2010. **12**(4): p. 207.
41. Roche. *What is liquid biopsy?* 2021 04.02.2021; Available from: https://www.roche.com/research_and_development/what_we_are_working_on/oncology/liquid-biopsy.htm.
42. Nguyen, Q.H., et al., *Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity.* Nat Commun, 2018. **9**(1): p. 2028.
43. Pantel, K. and R.H. Brakenhoff, *Dissecting the metastatic cascade.* Nature Reviews Cancer, 2004. **4**(6): p. 448-456.
44. Mathiesen, R.R., et al., *High-resolution analyses of copy number changes in disseminated tumor cells of patients with breast cancer.* Int J Cancer, 2012. **131**(4): p. E405-15.
45. Müller, V. and K. Pantel, *Bone marrow micrometastases and circulating tumor cells: current aspects and future perspectives.* Breast Cancer Res, 2004. **6**(6): p. 258-61.
46. Møller, E.K., et al., *Next-generation sequencing of disseminated tumor cells.* Front Oncol, 2013. **3**: p. 320.
47. Wiedswang, G., et al., *Comparison of the clinical significance of occult tumor cells in blood and bone marrow in breast cancer.* Int J Cancer, 2006. **118**(8): p. 2013-9.

48. Pantel, K., C. Alix-Panabières, and S. Riethdorf, *Cancer micrometastases.* Nature Reviews Clinical Oncology, 2009. **6**(6): p. 339-351.
49. Hayes, D.F., et al., *Circulating tumor cells at each follow-up time point during therapy of metastatic breast cancer patients predict progression-free and overall survival.* Clin Cancer Res, 2006. **12**(14 Pt 1): p. 4218-24.
50. Demeulemeester, J., et al., *Tracing the origin of disseminated tumor cells in breast cancer using single-cell sequencing.* Genome Biol, 2016. **17**(1): p. 250.
51. Pantel, K., R.H. Brakenhoff, and B. Brandt, *Detection, clinical relevance and specific biological properties of disseminating tumour cells.* Nature Reviews Cancer, 2008. **8**(5): p. 329-340.
52. Tang, X., et al., *The single-cell sequencing: new developments and medical applications.* Cell & Bioscience, 2019. **9**(1): p. 53.
53. Rossi, E., A. Facchinetti, and R. Zamarchi, *Notes for developing a molecular test for the full characterization of circulating tumor cells.* Chin J Cancer Res, 2015. **27**(5): p. 471-8.
54. Nawy, T., *Single-cell sequencing.* Nature Methods, 2014. **11**(1): p. 18-18.
55. Simone Picelli, O.R.F., Åsa K Björklund, Gösta Winberg, Sven Sagasser, Rickard Sandberg, *Full-length RNA-seq from single cells using Smart-seq2.* Nature, 2014. **9**(1).
56. Picelli, S., et al., *Smart-seq2 for sensitive full-length transcriptome profiling in single cells.* Nature Methods, 2013. **10**(11): p. 1096-1098.
57. Camp, J.G., R. Platt, and B. Treutlein, *Mapping human cell phenotypes to genotypes with single-cell genomics.* Science, 2019. **365**(6460): p. 1401.
58. Powell, A.A., et al., *Single cell profiling of circulating tumor cells: transcriptional heterogeneity and diversity from breast cancer cell lines.* PLoS One, 2012. **7**(5): p. e33788.
59. Gallego Romero, I., et al., *RNA-seq: impact of RNA degradation on transcript quantification.* BMC Biology, 2014. **12**(1): p. 42.
60. ATCC, *HCC38 (ATCC® CRL-2314™)*, ATCC, Editor. 2020: USA.
61. ATCC, *HCC2218 BL (ATCC® CRL2363™)*. 2018: USA.
62. Altin, J.G. and E.K. Sloan, *The role of CD45 and CD45-associated molecules in T cell activation.* Immunol Cell Biol, 1997. **75**(5): p. 430-45.
63. Borgen, E., et al., *NR2F1 stratifies dormant disseminated tumor cells in breast cancer patients.* Breast Cancer Res, 2018. **20**(1): p. 120.
64. Thermo Fisher Scientific. *Introduction to Cell Culture*. 28.10.2020]; Available from: https://www.thermofisher.com/no/en/home/references/gibco-cell-culture-basics/introduction-to-cell-culture.html.
65. Thermo Fisher Scientific. *Aseptic Technique*. Available from: https://www.thermofisher.com/no/en/home/references/gibco-cell-culture-basics/aseptic-technique.html.
66. Green Bioresearch LLC. *Cell Number Density At 100% Confluency In Cell Culture Dish, Plate, Flask*. 2016 16.10.2016; Available from: https://greenbioresearch.com/cell-number-density-percentage-confluency-cell-culture-dish-plate-flask/.
67. Thermo Fisher Scientific. *RPMI 1640 Media*. 28.10.2020]; Available from: https://www.thermofisher.com/no/en/home/life-science/cell-culture/mammalian-cell-culture/classical-media/rpmi.html.
68. Invitrogen, *Countess™ II Automated Cell Counter*. 2015.
69. Scientific, T.F. *1-Step™ NBT/BCIP*. 2011; Available from: https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2FMAN0011289_1Step_NBT_BCIP_UG.pdf&title=VXNlciBHdWlkZTogIDEtU3RlcCBOQlQvQkNJUA==.
70. Flow Cytometry Core Facility. *FACS Aria II cell sorter (Montebello)*. 2021; Available from: https://www.ous-research.no/home/flow/Services/18714.

71. Sino Biological. *Fluorescence-activated Cell Sorting (FACS)*. 2020 30.11.2020]; Available from: https://www.sinobiological.com/category/fcm-facs-facs.
72. Menarini Silicon Biosystems. *What is the CELLSEARCH® system?* 2020 24.08.2020 29.10.2020]; Available from: https://www.cellsearchctc.com/product-systems-overview/cellsearch-system-overview.
73. Biosystems., M.S. *DEPArray™ Technology*. 2020 29.10.2020]; Available from: http://www.siliconbiosystems.com/deparray-technology.
74. Menarini Silicon Biosystems, *DEPArray™ NxT How to Prepare a CellSearch® Enriched Sample*. 2020.
75. Schirmer, C.B., *Ekstraksjon av celler fra CS cartridge*, OUS, Editor. 2019.
76. Schirmer, C.B., *DepArray CTC RUO*, D.o.P. OUS, Editor. 2019.
77. Menarini Silicon Biosystems. *VRNxT™ - Volume Reduction Instrument*. 2020 16.11.2020]; Available from: http://www.siliconbiosystems.com/vrnxt.
78. Menarini Silicon Biosystems, *VRNxT™ Volume Reduction Instrument* 2019.
79. illumina. *A highly sensitive method for measuring gene expression from single cells*. 2021; Available from: https://www.illumina.com/techniques/sequencing/rna-sequencing/ultra-low-input-single-cell-rna-seq.html.
80. Thermo Fisher Scientific. *Qubit 4 Fluorometer*. 2021; Available from: https://www.thermofisher.com/no/en/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/fluorometers/qubit/qubit-fluorometer.html.
81. invitrogen. *Qubit™ 1X dsDNA HS Assay Kits*. 2019 16.11.2020]; Available from: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0017455_Qubit_1X_dsDNA_HS_Assay_Kit_UG.pdf.
82. Technologies, A., *2100 Expert SoftwareUser's Guide*. 2020.
83. Agilent Technologies, *Agilent High Sensitivity DNA Kit Guide*, A.T. Inc, Editor. 2020: Germany.
84. Bio-Rad Laboratories. *Droplet Digital™ PCR (ddPCR™) Technology*. 2021; Available from: https://www.bio-rad.com/en-no/applications-technologies/droplet-digital-pcr-ddpcr-technology?ID=MDV31M4VY.
85. Laboratories, B.-R., *Digital Droplet PCR Applications Guide*, Bio-Rad, Editor. 2021.
86. Liane D. Fairfull. *Assay Optimization for ddPCR*. 2020; Available from: https://www.genetics.pitt.edu/sites/default/files/pdfs/TechTalk/Liane%20TT%20DDPCR%20OPT%20SLIDES.pdf.
87. Jones, M., et al., *Low copy target detection by Droplet Digital PCR through application of a novel open access bioinformatic pipeline, 'definetherain'.* J Virol Methods, 2014. **202**(100): p. 46-53.
88. Gerdes, L., et al., *Optimization of digital droplet polymerase chain reaction for quantification of genetically modified organisms.* Biomol Detect Quantif, 2016. **7**: p. 9-20.
89. Swennenhuis, J.F. and L. Terstappen, *Sample Preparation Methods Following CellSearch Approach Compatible of Single-Cell Whole-Genome Amplification: An Overview.* Methods Mol Biol, 2015. **1347**: p. 57-67.
90. Zhang, K., et al., *Single-cell isolation by a modular single-cell pipette for RNA-sequencing.* Lab Chip, 2016. **16**(24): p. 4742-4748.
91. Sveli, M.A.T., *Immunocytochemistry straining protocol with bioninylated AE1AE3 single cell micromanipulation*. 2020.
92. Ding, S., X. Chen, and K. Shen, *Single-cell RNA sequencing in breast cancer: Understanding tumor heterogeneity and paving roads to individualized therapy.* Cancer Commun (Lond), 2020. **40**(8): p. 329-344.
93. Walker, L.C., et al., *Cytokeratin KRT8/18 expression differentiates distinct subtypes of grade 3 invasive ductal carcinoma of the breast.* Cancer Genet Cytogenet, 2007. **178**(2): p. 94-103.

94. Ziegenhain, C., et al., *Comparative Analysis of Single-Cell RNA Sequencing Methods.* Molecular Cell, 2017. **65**(4): p. 631-643.e4.
95. Košir, A.B., et al., *Droplet volume variability as a critical factor for accuracy of absolute quantification using droplet digital PCR.* Analytical and Bioanalytical Chemistry, 2017. **409**(28): p. 6689-6697.
96. Macaulay, I.C., et al., *G&T-seq: parallel sequencing of single-cell genomes and transcriptomes.* Nature Methods, 2015. **12**(6): p. 519-522.

Appendix 1. Plate set-up for digital droplet PCR.

Table 15. Plate set-up with multiplex of KRT18 and CD45, for samples from microinjection pipette method (1.0) method with 0,125ng/µL.

	1 - KRT18 + CD45	2 - KRT18 + CD45	3 - KRT18 + CD45	4 - KRT18 + CD45
A	COMP_1.0_HCC38_1	COMP_1.0_HCC38_9	COMP_1.0_HCC38_17	COMP_1.0_HCC38_25
B	COMP_1.0_HCC38_2	COMP_1.0_HCC38_10	COMP_1.0_HCC38_18	COMP_1.0_HCC38_26
C	COMP_1.0_HCC38_3	COMP_1.0_HCC38_11	COMP_1.0_HCC38_19	COMP_1.0_HCC38_27
D	COMP_1.0_HCC38_4	COMP_1.0_HCC38_12	COMP_1.0_HCC38_20	COMP_1.0_HCC38_28
E	COMP_1.0_HCC38_5	COMP_1.0_HCC38_13	COMP_1.0_HCC38_21	COMP_1.0_HCC38_29
F	COMP_1.0_HCC38_6	COMP_1.0_HCC38_14	COMP_1.0_HCC38_22	COMP_1.0_HCC38_30
G	COMP_1.0_HCC38_7	COMP_1.0_HCC38_15	COMP_1.0_HCC38_23	NTC
H	COMP_1.0_HCC38_8	COMP_1.0_HCC38_16	COMP_1.0_HCC38_24	HCC2218BL single cell

Table 16. Plate set-up with multiplex of KRT18 and CD45, for samples from micromanipulator (1.1) method with 0,125ng/ µL.

	1 - KRT18 + CD45	2 - KRT18 + CD45	3 - KRT18 + CD45
A	COMP_1.1_HCC38_1	COMP_1.1_HCC38_9	COMP_1.1_HCC38_25
B	COMP_1.1_HCC38_2	COMP_1.1_HCC38_10	COMP_1.1_HCC38_26
C	COMP_1.1_HCC38_3	COMP_1.1_HCC38_19	COMP_1.1_HCC38_27
D	COMP_1.1_HCC38_4	COMP_1.1_HCC38_20	COMP_1.1_HCC38_28
E	COMP_1.1_HCC38_5	COMP_1.1_HCC38_21	COMP_1.1_HCC38_29
F	COMP_1.1_HCC38_6	COMP_1.1_HCC38_22	COMP_1.1_HCC38_30
G	COMP_1.1_HCC38_7	COMP_1.1_HCC38_23	NTC
H	COMP_1.1_HCC38_8	COMP_1.1_HCC38_24	HCC2218BL single cell

Table 17. Plate set-up with multiplex of KRT18 and CD45, for samples from FACS (2.0) method with 0,125 ng/µL.

	1 - KRT18 + CD45	2 - KRT18 + CD45	3 - KRT18 + CD45	4 - KRT18 + CD45
A	COMP_2.0_HCC38_1	COMP_2.0_HCC38_9	COMP_2.0_HCC38_17	COMP_2.0_HCC38_25
B	COMP_2.0_HCC38_2	COMP_2.0_HCC38_10	COMP_2.0_HCC38_18	COMP_2.0_HCC38_26
C	COMP_2.0_HCC38_3	COMP_2.0_HCC38_11	COMP_2.0_HCC38_19	COMP_2.0_HCC38_27
D	COMP_2.0_HCC38_4	COMP_2.0_HCC38_12	COMP_2.0_HCC38_20	COMP_2.0_HCC38_28
E	COMP_2.0_HCC38_5	COMP_2.0_HCC38_13	COMP_2.0_HCC38_21	COMP_2.0_HCC38_29
F	COMP_2.0_HCC38_6	COMP_2.0_HCC38_14	COMP_2.0_HCC38_22	COMP_2.0_HCC38_30
G	COMP_2.0_HCC38_7	COMP_2.0_HCC38_15	COMP_2.0_HCC38_23	NTC
H	COMP_2.0_HCC38_8	COMP_2.0_HCC38_16	COMP_2.0_HCC38_24	HCC2218BL single cell

Table 18. Plate set-up with multiplex of KRT18 and CD45, for samples from DEPArray (3.0) method with 0,125ng/µL.

	1 - KRT18 + CD45	**2 - KRT18 + CD45**	**4 - KRT18 + CD45**	**5 - KRT18 + CD45**
A	COMP_3.0_HCC38_1	COMP_3.0_HCC38_9	COMP_3.0_HCC38_17	COMP_3.0_HCC38_25
B	COMP_3.0_HCC38_2	COMP_3.0_HCC38_10	COMP_3.0_HCC38_18	COMP_3.0_HCC38_26
C	COMP_3.0_HCC38_3	COMP_3.0_HCC38_11	COMP_3.0_HCC38_19	COMP_3.0_HCC38_27
D	COMP_3.0_HCC38_4	COMP_3.0_HCC38_12	COMP_3.0_HCC38_20	COMP_3.0_HCC38_28
E	COMP_3.0_HCC38_5	COMP_3.0_HCC38_13	COMP_3.0_HCC38_21	COMP_3.0_HCC38_29
F	COMP_3.0_HCC38_6	COMP_3.0_HCC38_14	COMP_3.0_HCC38_22	COMP_3.0_HCC38_30
G	COMP_3.0_HCC38_7	COMP_3.0_HCC38_15	COMP_3.0_HCC38_23	NTC
H	COMP_3.0_HCC38_8	COMP_3.0_HCC38_16	COMP_3.0_HCC38_24	HCC2218BL single cell

Table 19. Plate set-up with multiplex of KRT18 and CD45, for patient samples from FACS (2.0) with 0,125ng/µL.

	1 - KRT18 + CD45
A	LATE_2.0_1_1
B	LATE_2.0_1_3
C	LATE_2.0_1_4
D	LATE_2.0_2_1
E	LATE_2.0_2_3
F	RNA Control
G	HCC2218BL single cell
H	NTC

Appendix 2. Results from Bioanalyzer

All the samples were analysed on Bioanalyzer. The first samples that were analysed were not diluted, and because these results were so high compared to the reference peaks, the remaining samples were diluted down to approximately 2 ng/µL before analysis.

Method 1.0 – microinjection pipette isolation of single cells

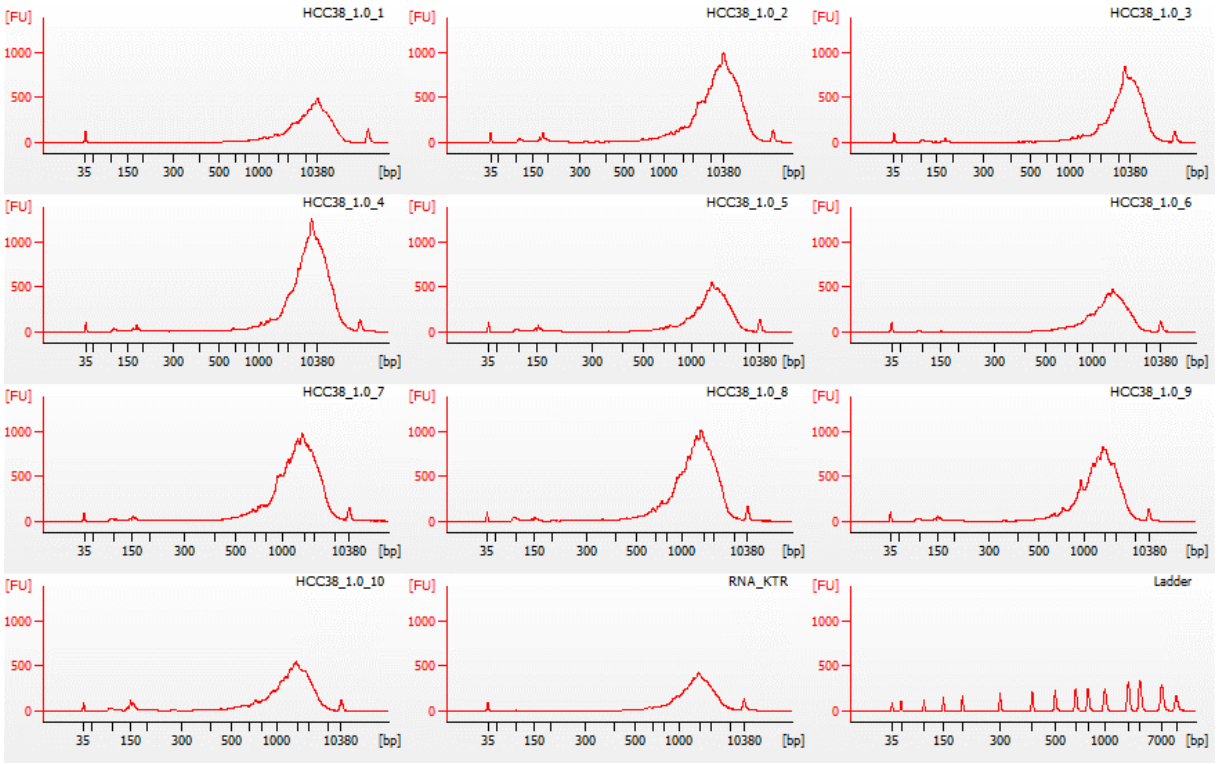

Figure 34. Bioanalyzer result from cells 1-10 from method 1.0. The X-axis represents the length of the fragments with number of basepairs (bp) and the Y-axis represent the number of fragment units (FU). These samples were not diluted before testing and have high peaks. Analysis was not repeated, to not misuse reagents and time, as well as it still shows the high peak around 1000 base pairs.

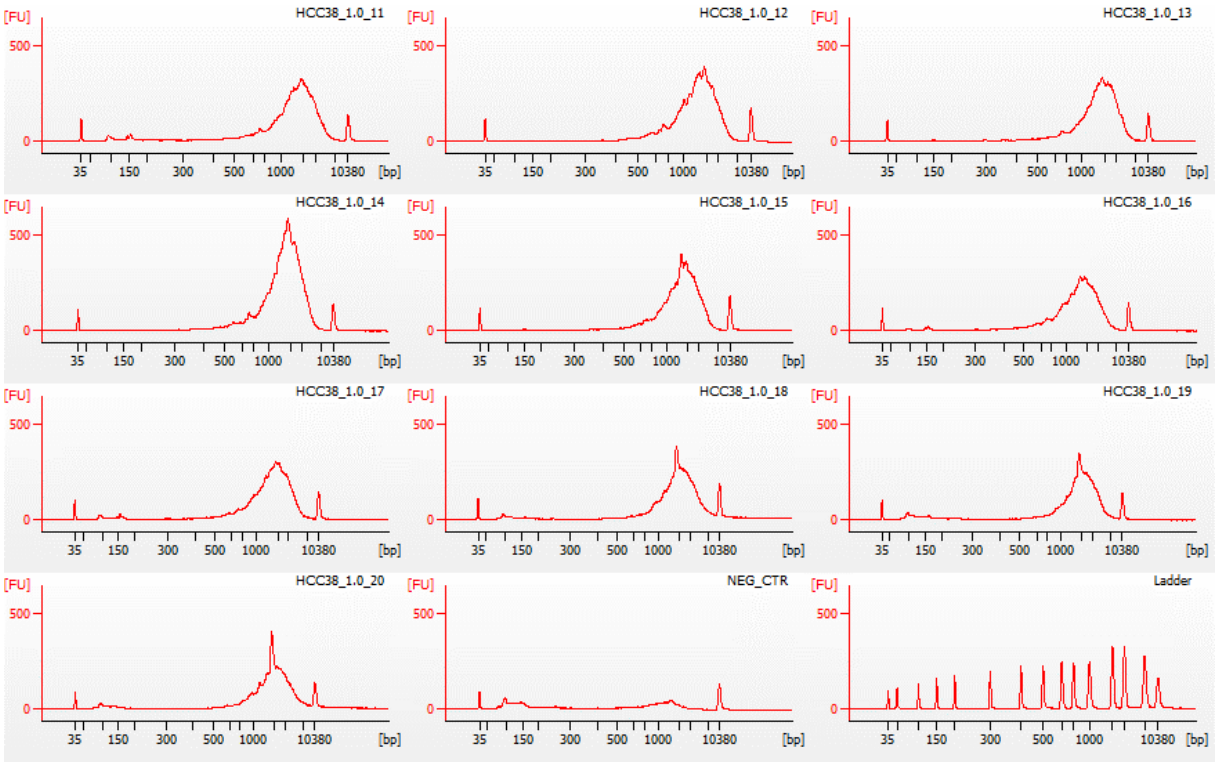

Figure 35. Bioanalyzer result from cells 11-20 from method 1.0. The X-axis represents the length of the fragments with number of basepairs (bp) and the Y-axis represent the number of fragment units (FU). Samples were diluted down to around 2ng.

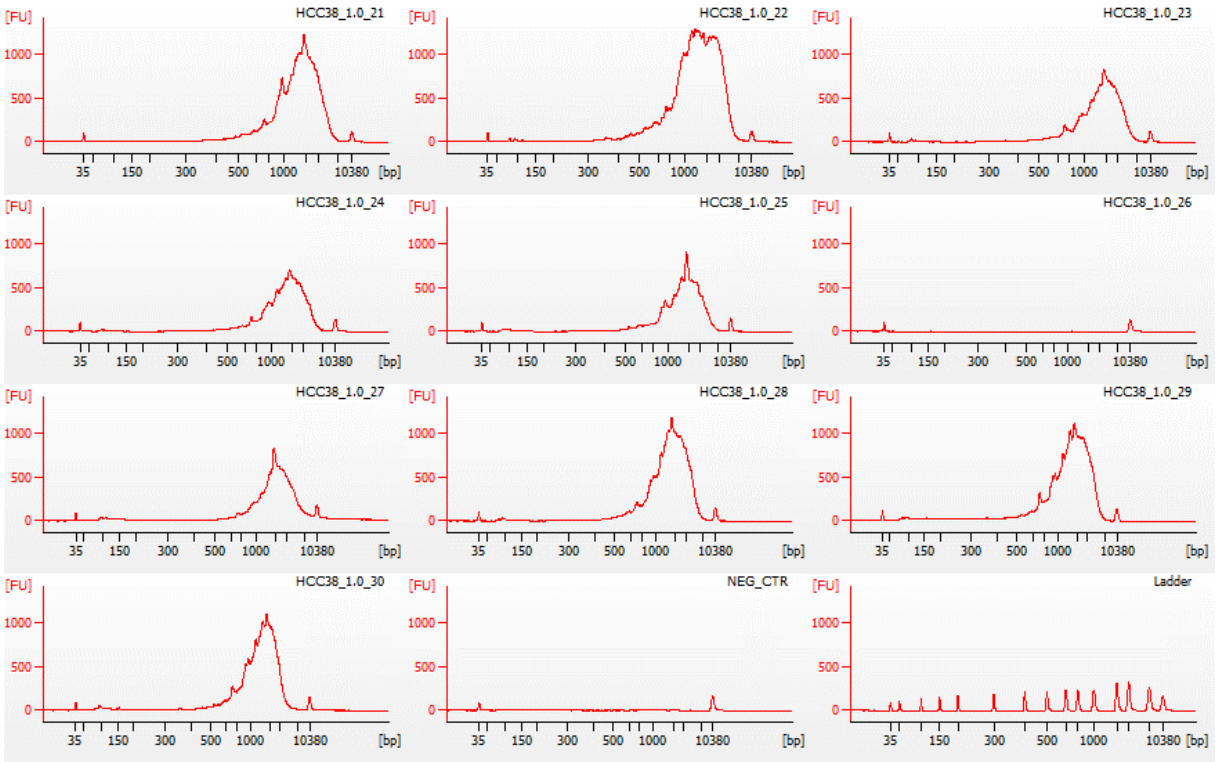

Figure 36. Bioanalyzer result from cells 21-30 from method 1.0. The X-axis represents the length of the fragments with number of basepairs (bp) and the Y-axis represent the number of fragment units (FU). Samples were not diluted before testing.

Method 1.1 – Cytospin and micromanipulator for isolation of single cells.

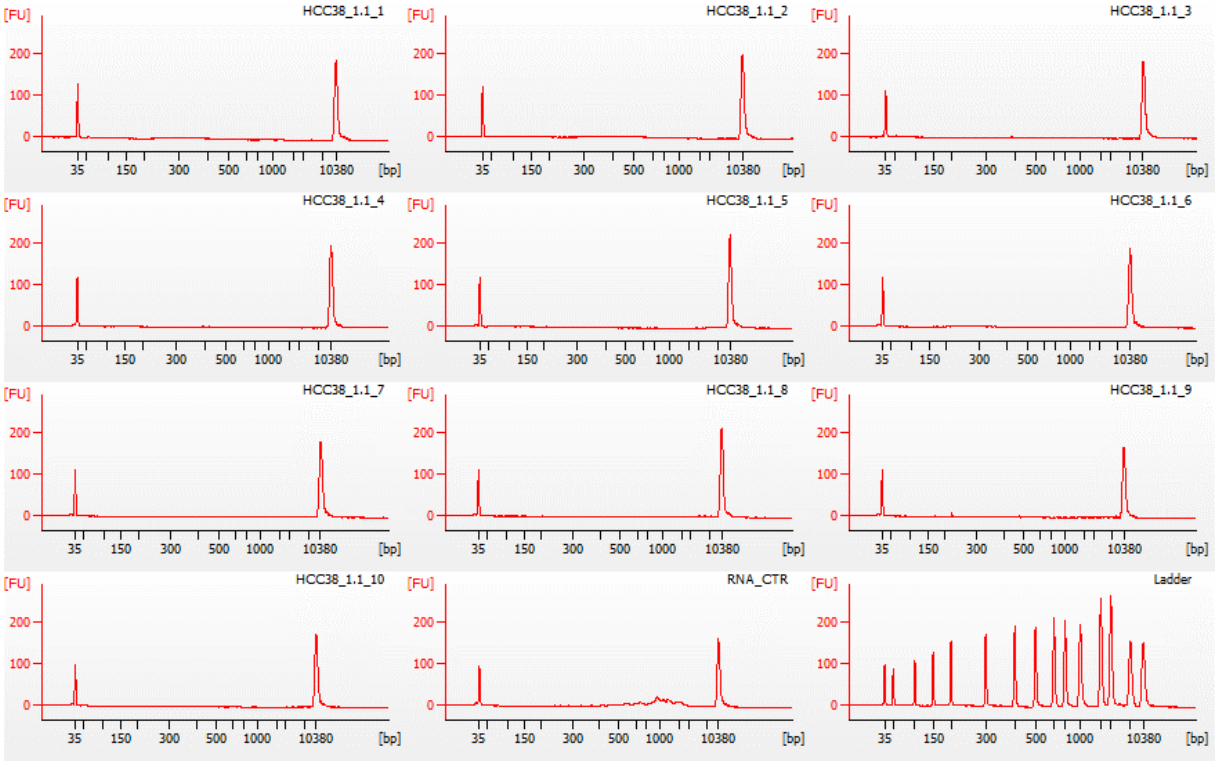

Figure 37. Bioanalyzer result from cells 1-10 from method 1.1. The X-axis represents the length of the fragments with number of basepairs (bp) and the Y-axis represent the number of fragment units (FU).

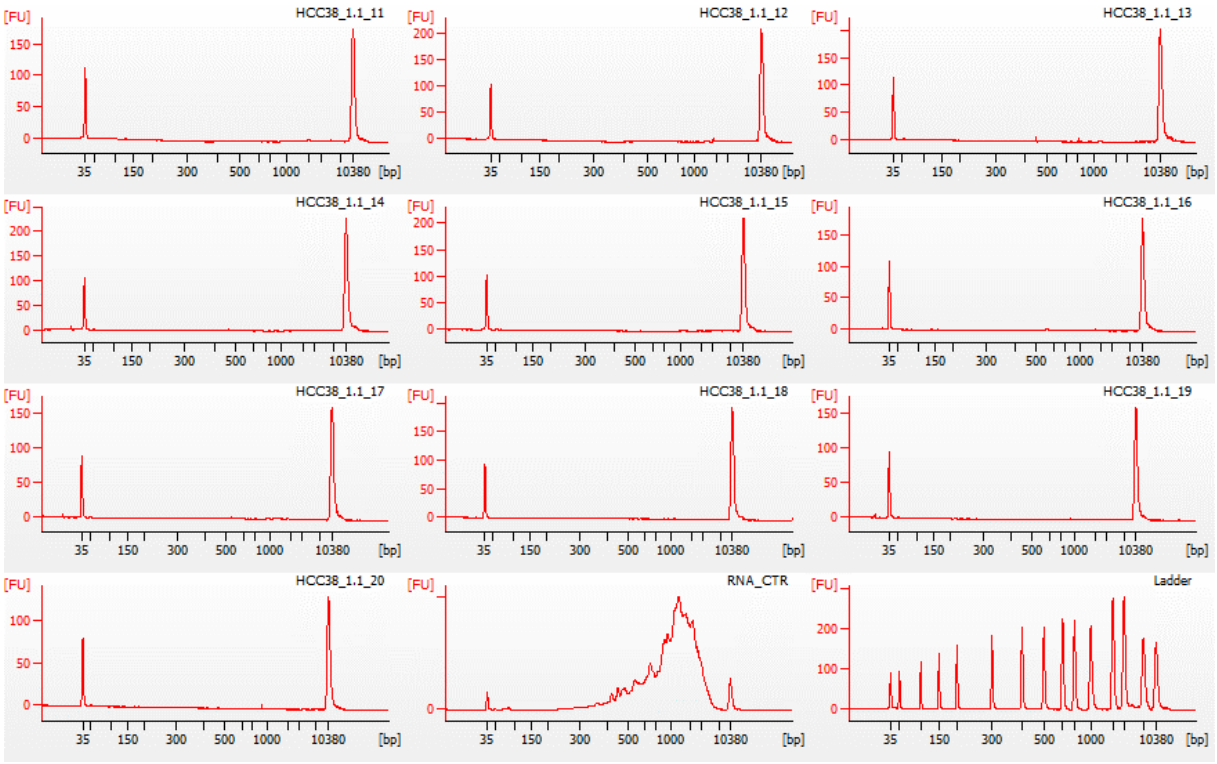

Figure 38. Bioanalyzer result from cells 11-20 from method 1.1. The X-axis represents the length of the fragments with number of basepairs (bp) and the Y-axis represent the number of fragment units (FU).

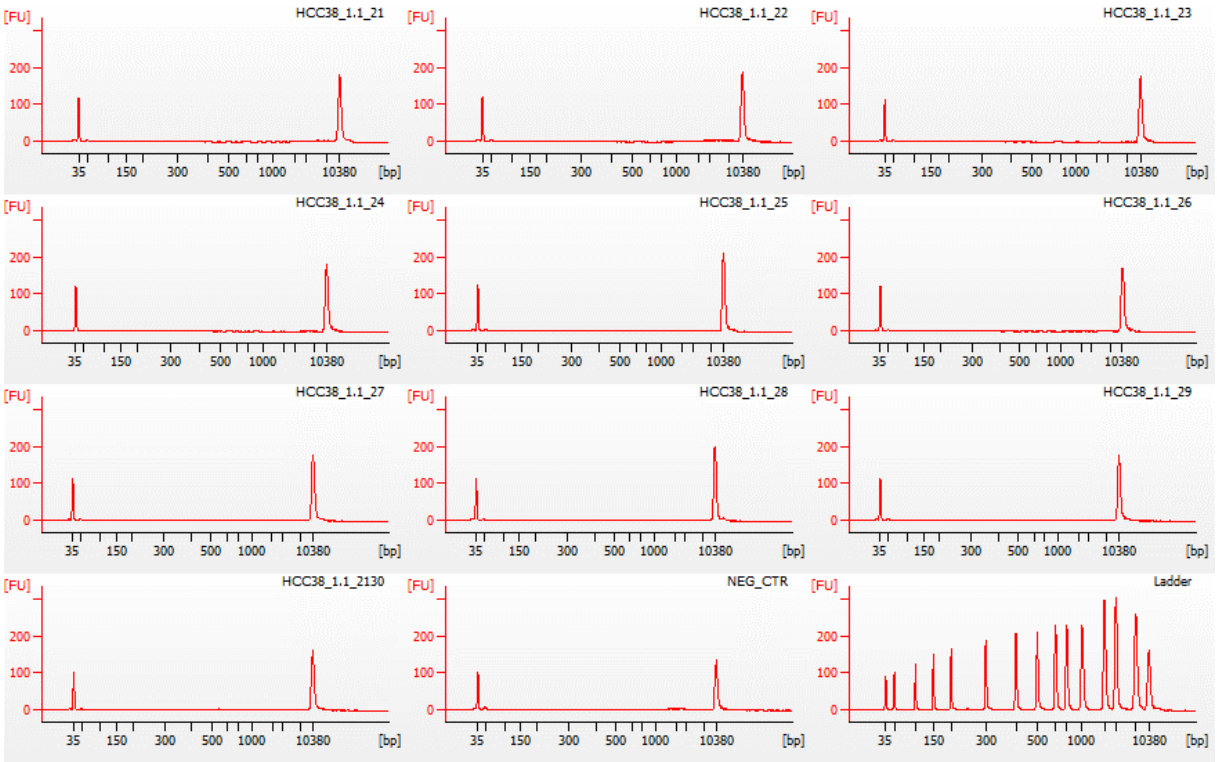

Figure 39. Bioanalyzer result from cells 21-30 from method 1.1. The X-axis represents the length of the fragments with number of basepairs (bp) and the Y-axis represent the number of fragment units (FU).

Method 2.0 – FACS automatic isolation of single cells.

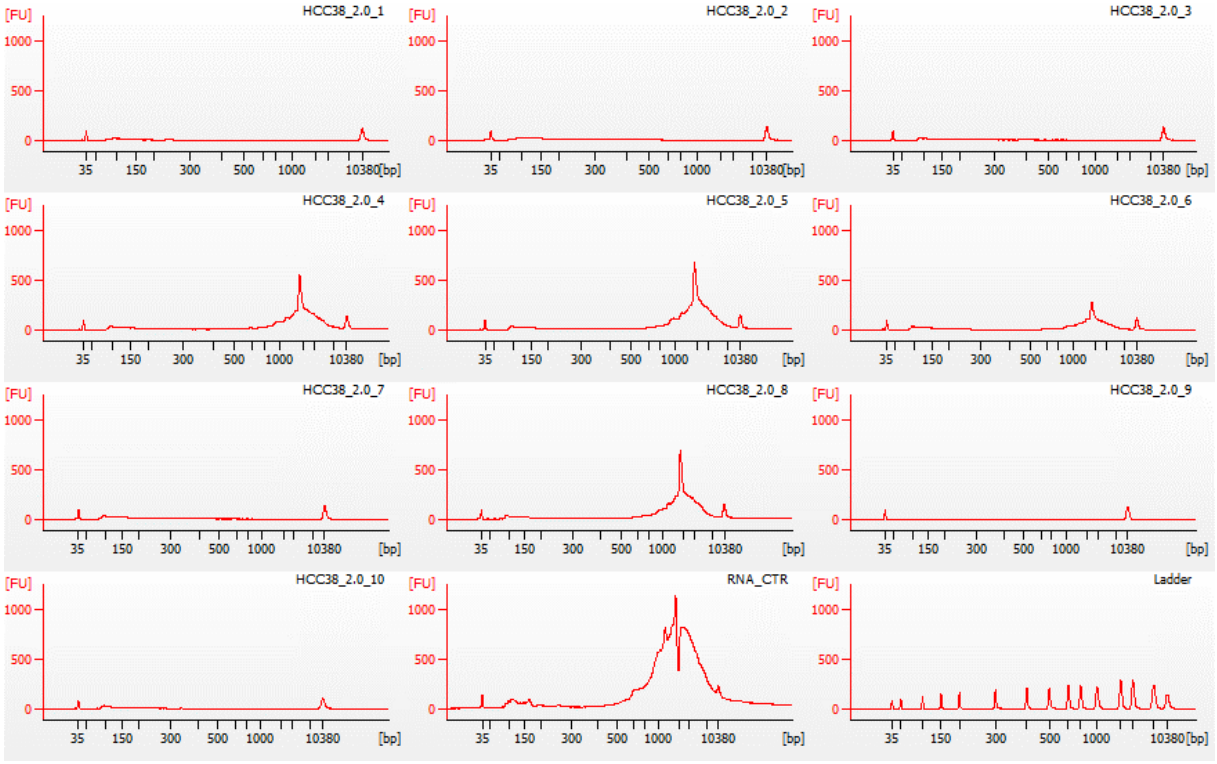

Figure 40. Bioanalyzer result from cells 1-10 from method 2.0. The X-axis represents the length of the fragments with number of basepairs (bp) and the Y-axis represent the number of fragment units (FU). Samples were not diluted before testing, and the undiluted RNA control makes this experiment difficult to inspect but sample 4,5,6 and 7 have high peaks from up to 500 fragment units of longer fragments.

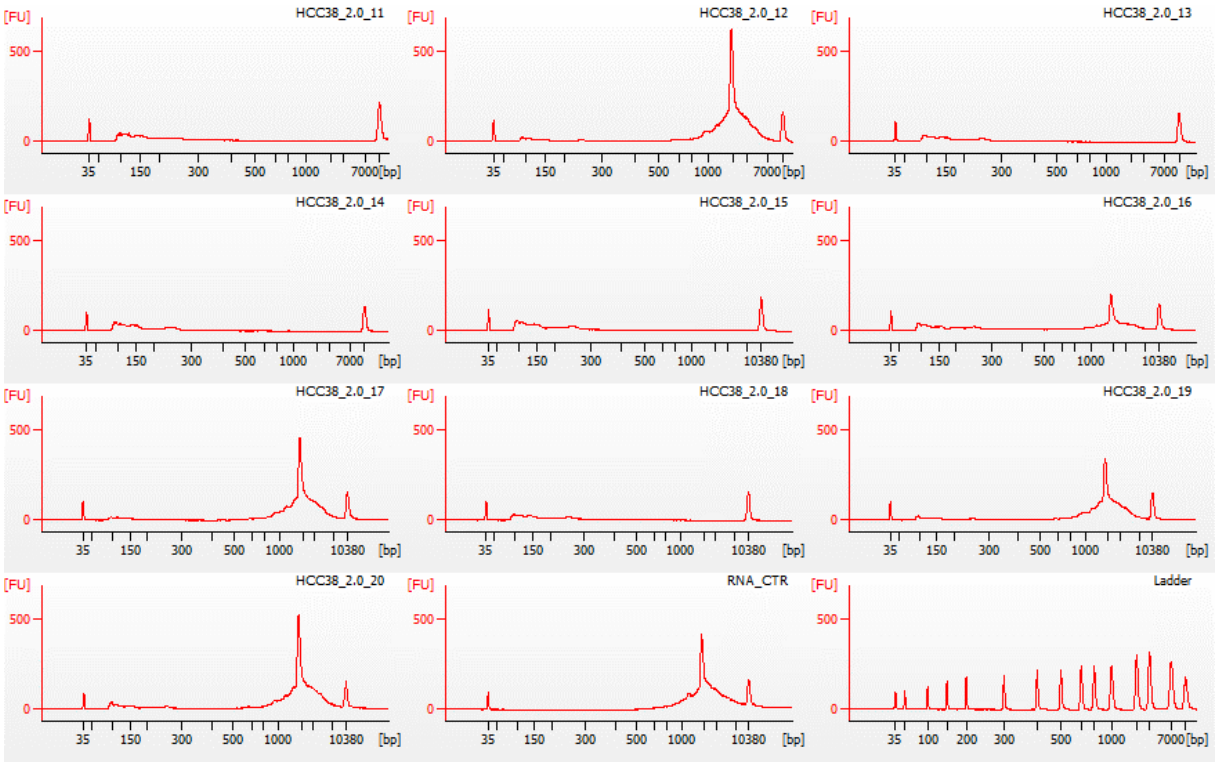

Figure 41. Bioanalyzer result from cells 11-20 from method 2.0. The X-axis represents the length of the fragments with number of basepairs (bp) and the Y-axis represent the number of fragment units (FU). Samples were not diluted before testing.

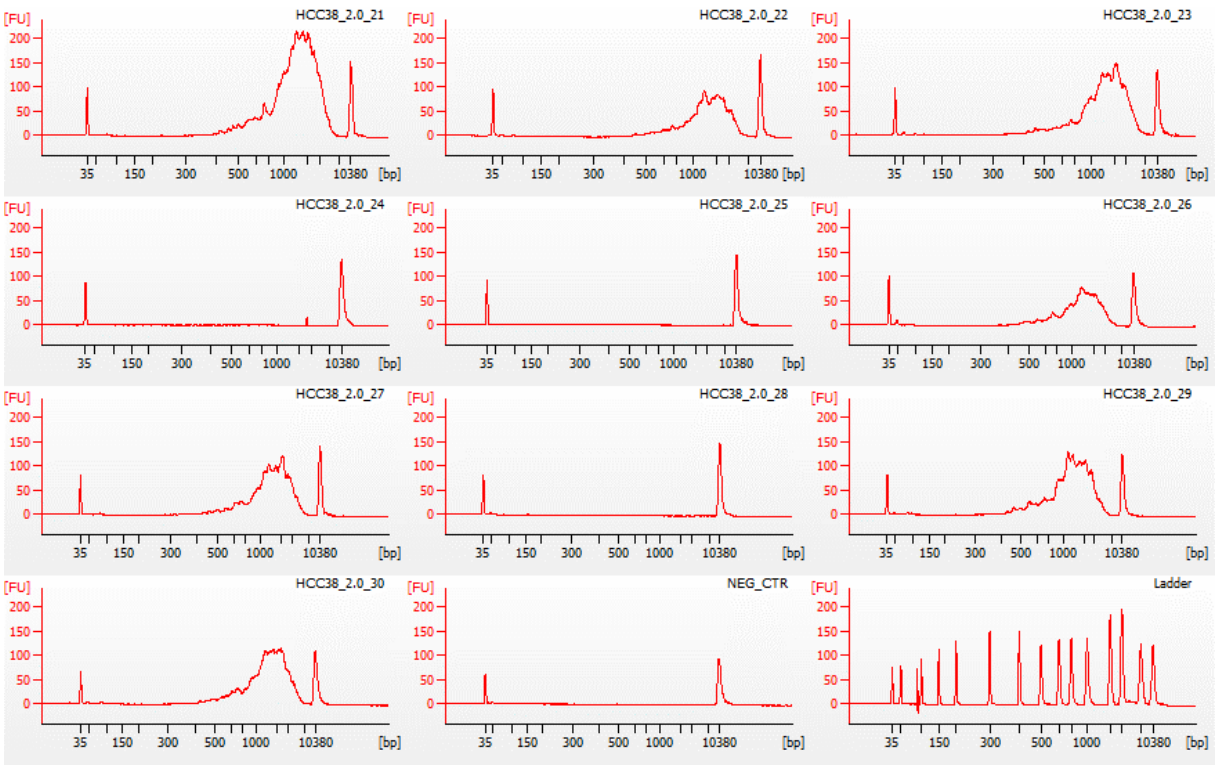

Figure 42. Bioanalyzer result from cells 2-30 from method 2.0. The X-axis represents the length of the fragments with number of basepairs (bp) and the Y-axis represent the number of fragment units (FU).

Method 3.0 – Cellsearch and DEPArray automatic isolation of single cells.

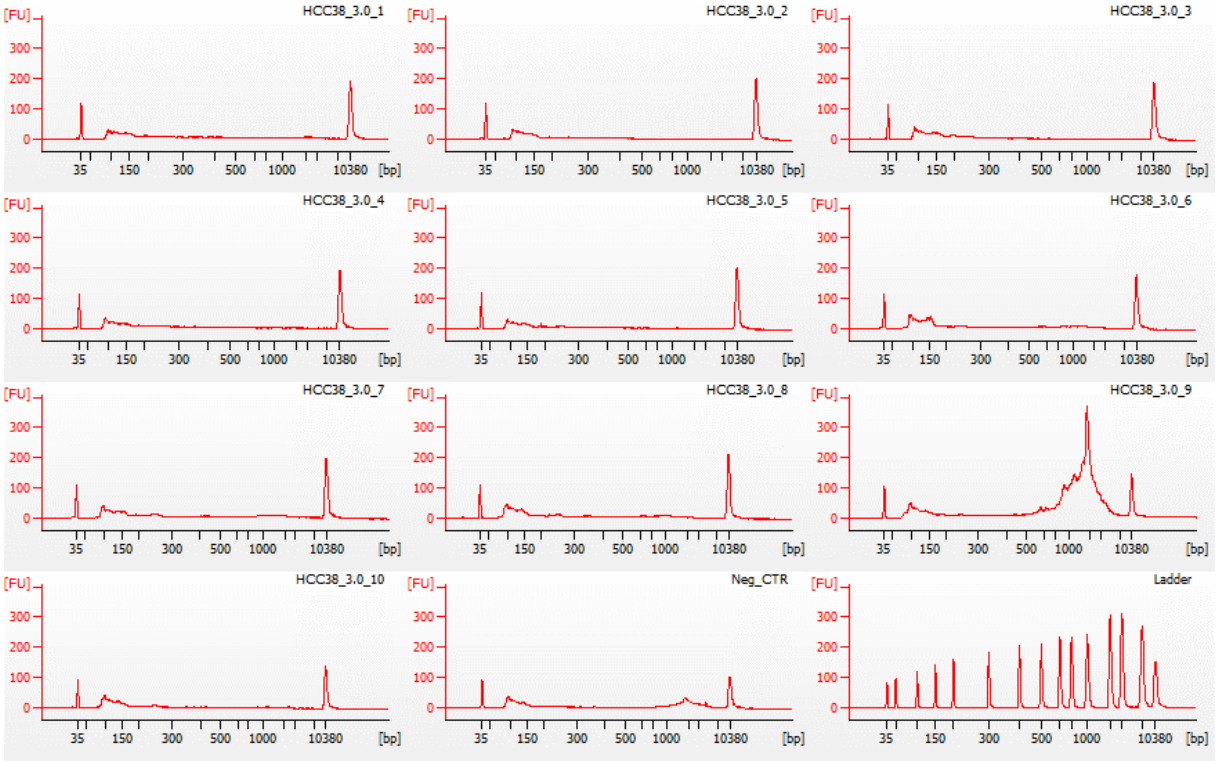

Figure 43. Bioanalyzer result from cells 1-10 from method 3.0. The X-axis represents the length of the fragments with number of basepairs (bp) and the Y-axis represent the number of fragment units (FU). Samples 8 and 9 have been pipetted in opposite wells by mistake, result in 9 is from sample 8.

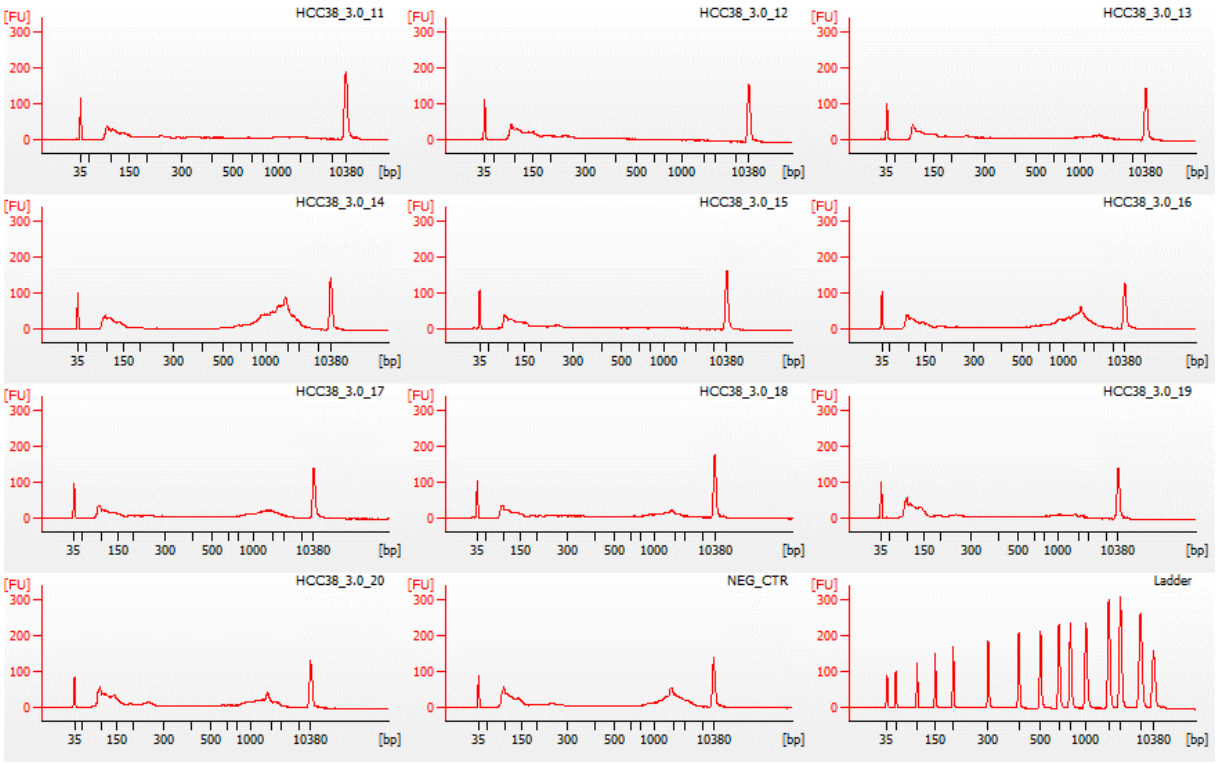

Figure 44. Bioanalyzer result from cells 11-20 from method 3.0. The X-axis represents the length of the fragments with number of basepairs (bp) and the Y-axis represent the number of fragment units (FU).

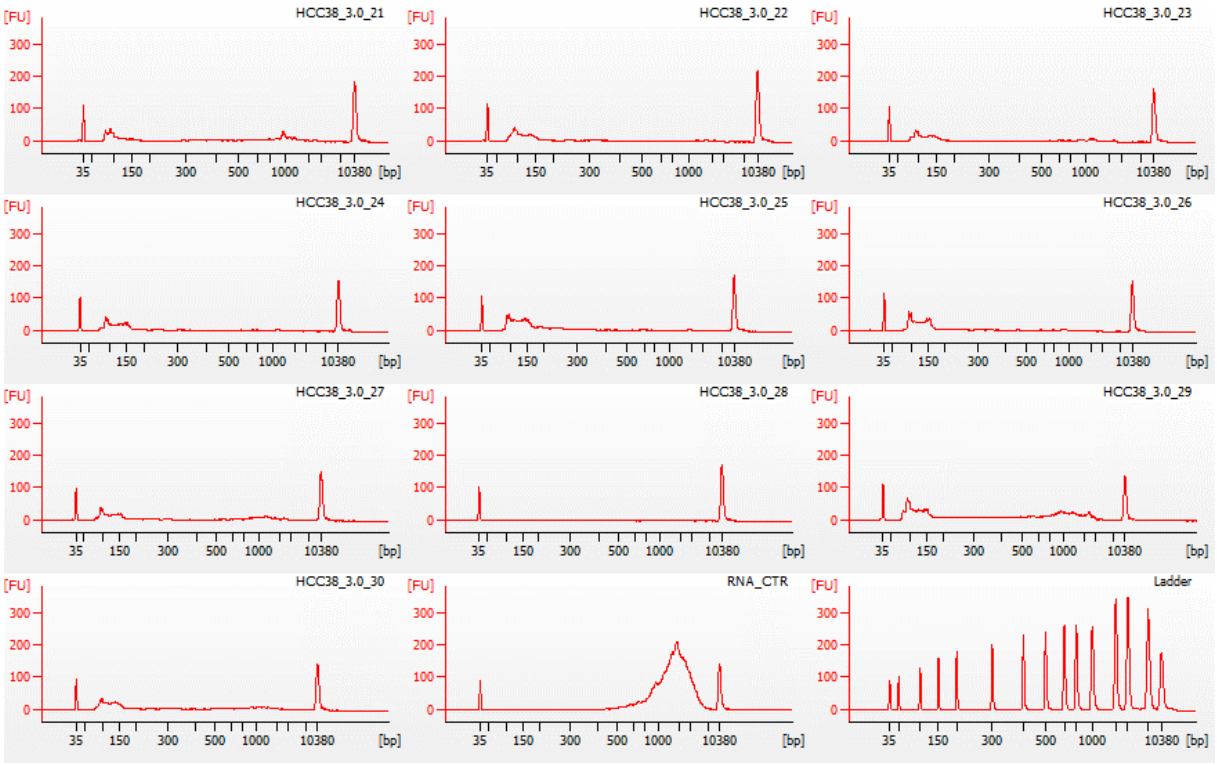

Figure 45. Bioanalyzer result from cells 21-30 from method 3.0. The X-axis represents the length of the fragments with number of basepairs (bp) and the Y-axis represent the number of fragment units (FU).

Appendix 3. Full list of reagents and equipment

Component	Cat. No. (LOT)	Manufacturer
Single use equipment		
Filtered pipette tips 0.1-20µL	70.1114.210	Sarstedt
Filtered pipette tips 2-100µL	70.760.212	Sarstedt
Filtered pipette tips 1000µL	70.3050.255	Sarstedt
Eppendorf Safe-lock tubes (1,5mL)	213100-328	VWR
Tube 15mL	42121	Sarstedt
Eppendorf Tubes® 5.0mL	30119460	Eppendorf
Protein LoBind Tube 1.5mL	22431081	Eppendorf
MicroAMP™ 96-well plate	N8010560	Thermo Fisher Scientific
Cell culture		
Dulbecco's Phosphate Buffered Saline (DPBS)	14190-094	Gibco
Fetal bovine serum	21568719H	Gibco
Pipetboy®	612-2964	VWR
Roswell Park Memorial Institute Medium 1640	72400-021	Gibco
Corning® flask, 25cm²	430639	Corning
TC-Flask T25, Susp.	6021011	Sarstedt
Nunc™ EasYFlask™ 75cm²	156499	Thermo scientific
Trypsin, 0,25% with EDTA	25200-072	Invitrogen
Dulbecco's Modified Eagle Medium	2198725	Gibco
Trypan blue stain 0.4%	10702404	Invitrogen
Countess cell counting slide	C10228	Invitrogen
DNase 1	11284932001	Sigma
Cell lines:		
HCC38 - human, mammary gland	CRL-2314™	ATCC®
HCC2218BL - human, peripheral blood	CRL-2363™	ATCC®
Manual single-cell isolation		
Polyvinylpyrrolidone (PVP)	P5288-100G	Sigma-Aldrich
The STRIPPER micropipette	MXL3-STR-CGR	Origio Benelux
STRIPPER TIPS	MXL3-75	Origio Benelux
Nunclon™ Delta Surface	159950	Thermo scientific
Buffer RLT Lysis buffer	1015750	Qiagen
PCR Tube Strips 0.2mL	951010022	Eppendorf
Mineral oil	MKBD9874V	Sigma

Cytospin & Immunocytochemistry staining		
Superfrost Plus object slide	10149870	Thermo scientific
Cytospinholders with chamber and filter	59910052	Thermo Fisher scientific
AE-1/AE-3 cytokeratin biotin	138538-AF-Bioti	Nordic BioSite
Streptavidin Alkaline Streptavidine Phosphatase	S-5100	Vector Laboratories
Levamisole Endogenous Alkaline Phosphatase Inhibitor	X3021	Dako
BCIP/NBT Substrate system, ready-to-use	K0598	Dako
Antibodies & Stains		
CD45-APC, human	130-113-114	MACS
Ep-CAM (O.N.276) PE	B2317	Santa Cruz Biotechnology
Hoechst 33258	H3569	Invitrogen
Smartseq		
RNAClean XP beads	A66514	Beckman Coulter
AMPure XP beads	A63881	Beckman Coulter
Oligo-dT 5' Biotin 10µM	227438216	Integrated DNA Technologies
dNTP Mix 10 mM	751795	Thermo Scientific
5x First Strand Buffer	2225915	Invitrogen
Super script® II Reverse Transcriptase	18064014	Thermo Fisher Scientific
DTT 100mM	2236778	Invitrogen
SMARTSEQ TSO 100µM	YCO0191258	Qiagen
Recombinant Rnase Inhibitor	SD0367	TaKaRa
RT-PCR Grade Water	744045	Invitrogen
KAPA Hifi HotStart	114575	Roche
IS PCR primers	223114093	Integrated DNA Technologies
Qubit		
Qubit™ 1x dsDNA HS Assay kit	Q32854	Life Technologies
Qubit™ RNA BR Assay kit 100 assay	2136835, 2172034	Life Technologies
Qubit™ Assay tubes	12920	Life Technologies
Eppendorf Safe-lock tubes (1,5mL)	213100-328	VWR
Bioanalyzer		
High Sensitivity DNA Reagents	5067-4626	Agilent Technologies
High Sensitivity DNA Chips	ZB08BK50	Agilent Technologies

Droplet Digital™ PCR		
2x ddPCR Supermix for probes (no dUTP)	1863024	Bio-Rad
Droplet Generation Oil for probes	1863005	Bio-Rad
Nuclease free water	9935G	Bio-Rad
ddPCR Plates 96-well, semi-skirted	12001925	Bio-Rad
DG8 Cartridges for QX200 Droplet Generator	1864008	Bio-Rad
DG8 Cartridge Holder	1863051	Bio-Rad
Droplet Generator DG8 Gasket	1863009	Bio-Rad
Pierceable Foil Heat Seal	1814040	Bio-Rad
DNA LoBind tube 1.5mL	022431021	Eppendorf
CELLSEARCH		
1-200 µL Gel loading tips	4853	Corning
Protein LoBind tube 1.5mL	H181243P	Eppendorf
Cellsave	916624	Menarini silicon biosystems
DEPArray		
Bovine Serum Albumin	SLCB1974	Sigma
DEPArray™ NxT Cartridge	CNXT1	Menarini silicon biosystems
DEPArray™ Buffer for Fixed cells	KI0066	Menarini silicon biosystems
Protein LoBind tube 1.5mL	H181243P	Eppendorf
MicroAMP® Reaction Tube with Cap (0.2mL)	N8010540	Life Technologies
VRNxT™ Cap v1.0.0	23.07.2019	Menarini silicon biosystems

Instrument	Model	Manufacturer
Bioanalyzer	2100	Agilent
Automated Cell counter	Countess II FL	Invitrogen™
FACS	Aria™ Ilu	Beckton Dickinson (BD)
Biowizard LAF bench	369-1455	Bryn Byggklima AS
CELLSEARCH®	Celltracks autoprep system	Menarini
Centrifuge	Rotina 420	Hettich Zentrifugen
Cytocentrifuge	Hettich Universal 30F/320	Hettich
DEPArray™	DEPArray™ NxT	Menarini
Vortex mixer	MS3	IKA®
CO_2 Incubator	NU-5700	Nuaire
Light microscope	TL3000 Ergo	Wetzlar, Germany
Micromanipulator	Transferman NK2	Eppendorf
PCR cycle machine	Life Touch	BIOER
PX1™ Plate Sealer	77BR5499	Bio-Rad
QIAcube	QIAcube	Qiagen
Qubit® Fluorometer	4.0	Life Technolgies
QX200 Droplet Generator	772BR7056	Bio-Rad
QX200 Droplet Reader	771BR5589	Bio-Rad
Stereo microscope	Leica TL3000 Ergo	Leica Microsystems
Swinging bucket centrifuge	Eppendorf 5810R	Eppendorf
Thermal cycler (PCR cycle)	VeritiDx	Applied Biosystems
Volume reduction instrument	VRNxT™	Menarini
Water bath	462-0557	VWR

Appendix 4. Master table

A table of all the single cells selected and isolated by all the methods, including patient samples selected and isolated by FACS. Each single cell is displayed with its cDNA concentration (ng/µL) by Qubit, fragment distribution by Bioanalyzer and expression of *KRT18* by ddPCR. The fragment distribution is ranged from "flat line" when there is no visual curve, to "Low" when there were smaller peaks from shorter fragments (below 150bp), to "Low-High" when there were both a smaller peak and a peak from larger fragments (above 1000bp) and to "High" when there was only a peak from larger fragments. The positive expression is defined as "Negative", "Positive" and "Positive, few droplets". The description "Positive, few droplets" were used on the samples with only a few positive samples, but at least three positive droplets, which is three times the number from the negative control.

Table 20. Master table.

CELL ID #	QUBIT	BIOANALYZER	EXPRESSION OF KRT18
COMP_1.0_HCC38_1	25,6	High	Positive
COMP_1.0_HCC38_2	8,5	High	Positive
COMP_1.0_HCC38_3	5,76	High	Positive
COMP_1.0_HCC38_4	9,22	High	Positive
COMP_1.0_HCC38_5	3,58	High	Positive
COMP_1.0_HCC38_6	19,6	High	Positive
COMP_1.0_HCC38_7	8,28	High	Positive
COMP_1.0_HCC38_8	8	High	Positive
COMP_1.0_HCC38_9	6,34	High	Positive
COMP_1.0_HCC38_10	4,42	High	Positive
COMP_1.0_HCC38_11	2,34	High	Positive
COMP_1.0_HCC38_12	4,88	High	Positive
COMP_1.0_HCC38_13	4,92	High	Positive
COMP_1.0_HCC38_14	9,78	High	Positive
COMP_1.0_HCC38_15	4,18	High	Positive
COMP_1.0_HCC38_16	5,8	High	Positive
COMP_1.0_HCC38_17	3,2	High	Positive
COMP_1.0_HCC38_18	5,66	High	Positive

COMP_1.0_HCC38_19	3,36	High	Positive
COMP_1.0_HCC38_20	3,72	High	Positive
COMP_1.0_HCC38_21	11,3	High	Positive
COMP_1.0_HCC38_22	22	High	Positive
COMP_1.0_HCC38_23	6,6	High	Positive
COMP_1.0_HCC38_24	7,44	High	Positive
COMP_1.0_HCC38_25	5,12	High	Positive
COMP_1.0_HCC38_26	11,7	Flate line	Positive
COMP_1.0_HCC38_27	6,34	High	Positive
COMP_1.0_HCC38_28	9,4	High	Positive
COMP_1.0_HCC38_29	8,4	High	Positive
COMP_1.0_HCC38_30	10,3	High	Positive
COMP_1.1_HCC38_1	0,00	Flate line	Negative
COMP_1.1_HCC38_2	0,00	Flate line	Positive, few droplets
COMP_1.1_HCC38_3	0,00	Flate line	Negative
COMP_1.1_HCC38_4	0,00	Flate line	Negative
COMP_1.1_HCC38_5	0,00	Flate line	Negative
COMP_1.1_HCC38_6	0,00	Flate line	Negative
COMP_1.1_HCC38_7	0,00	Flate line	Negative
COMP_1.1_HCC38_8	0,00	Flate line	Negative
COMP_1.1_HCC38_9	0,00	Flate line	Negative
COMP_1.1_HCC38_10	0,10	Flate line	Negative
COMP_1.1_HCC38_11	0,03	Flate line	NA
COMP_1.1_HCC38_12	0,00	Flate line	NA
COMP_1.1_HCC38_13	0,00	Flate line	NA
COMP_1.1_HCC38_14	0,00	Flate line	NA
COMP_1.1_HCC38_15	0,02	Flate line	NA
COMP_1.1_HCC38_16	0,02	Flate line	NA
COMP_1.1_HCC38_17	0,03	Flate line	NA
COMP_1.1_HCC38_18	0,00	Flate line	NA
COMP_1.1_HCC38_19	0,00	Flate line	Negative
COMP_1.1_HCC38_20	0,00	Flate line	Negative
COMP_1.1_HCC38_21	0,18	Flate line	Negative

COMP_1.1_HCC38_22	0,14	Flate line	Negative
COMP_1.1_HCC38_23	0,00	Flate line	Negative
COMP_1.1_HCC38_24	0,00	Flate line	Negative
COMP_1.1_HCC38_25	0,00	Flate line	Negative
COMP_1.1_HCC38_26	0,00	Flate line	Negative
COMP_1.1_HCC38_27	0,00	Flate line	Negative
COMP_1.1_HCC38_28	0,00	Flate line	Negative
COMP_1.1_HCC38_29	0,00	Flate line	Negative
COMP_1.1_HCC38_30	0,00	Flate line	Positive, few droplets
COMP_2.0_HCC38_1	0,52	Flat line	Negative
COMP_2.0_HCC38_2	0,52	Flat line	Negative
COMP_2.0_HCC38_3	0,70	Flat line	Negative
COMP_2.0_HCC38_4	2,04	High	Positive
COMP_2.0_HCC38_5	4,64	High	Positive
COMP_2.0_HCC38_6	1,95	High	Positive
COMP_2.0_HCC38_7	0,84	Flat line	Negative
COMP_2.0_HCC38_8	3,48	High	Positive
COMP_2.0_HCC38_9	0,60	Flat line	Negative
COMP_2.0_HCC38_10	0,96	Flat line	Positive
COMP_2.0_HCC38_11	1,03	Low	Positive, few droplets
COMP_2.0_HCC38_12	4,4	High	Positive
COMP_2.0_HCC38_13	0,61	Low	Negative
COMP_2.0_HCC38_14	0,81	Low	Negative
COMP_2.0_HCC38_15	0,87	Low	Negative
COMP_2.0_HCC38_16	1,44	High	Positive
COMP_2.0_HCC38_17	3,78	High	Positive
COMP_2.0_HCC38_18	0,71	Low	Positive, few droplets
COMP_2.0_HCC38_19	3,3	High	Positive
COMP_2.0_HCC38_20	2,3	High	Positive
COMP_2.0_HCC38_21	2,26	High	Negative
COMP_2.0_HCC38_22	0,85	High	Negative
COMP_2.0_HCC38_23	1,32	High	Negative
COMP_2.0_HCC38_24	0,00	Flat line	Negative

COMP_2.0_HCC38_25	0,00	Flat line	Negative
COMP_2.0_HCC38_26	0,79	High	Negative
COMP_2.0_HCC38_27	1,06	High	Negative
COMP_2.0_HCC38_28	0,00	Flat line	Negative
COMP_2.0_HCC38_29	1,31	High	Negative
COMP_2.0_HCC38_30	1,41	High	Negative
COMP_3.0_HCC38_1	0,76	Low	Negative
COMP_3.0_HCC38_2	0,55	Low	Negative
COMP_3.0_HCC38_3	0,86	Low	Negative
COMP_3.0_HCC38_4	0,76	Low	Negative
COMP_3.0_HCC38_5	0,58	Low	Negative
COMP_3.0_HCC38_6	1,09	Low	Negative
COMP_3.0_HCC38_7	0,94	Low	Negative
COMP_3.0_HCC38_8	2,34	Low	Positive
COMP_3.0_HCC38_9	0,53	Low	Negative
COMP_3.0_HCC38_10	0,57	Low	Negative
COMP_3.0_HCC38_11	0,77	Low	Negative
COMP_3.0_HCC38_12	0,78	Low	Negative
COMP_3.0_HCC38_13	0,73	Low	Positive
COMP_3.0_HCC38_14	0,96	Low-High	Positive
COMP_3.0_HCC38_15	0,5	Low	Negative
COMP_3.0_HCC38_16	0,8	Low-High	Positive
COMP_3.0_HCC38_17	0,64	Low-High	Negative
COMP_3.0_HCC38_18	0,64	Low-High	Negative
COMP_3.0_HCC38_19	0,67	Low	Negative
COMP_3.0_HCC38_20	0,74	Low-High	Negative
COMP_3.0_HCC38_21	0,38	Low-High	Negative
COMP_3.0_HCC38_22	0,32	Low	Negative
COMP_3.0_HCC38_23	0,31	Low	Negative
COMP_3.0_HCC38_24	0,43	Low	Negative
COMP_3.0_HCC38_25	0,48	Low	Negative
COMP_3.0_HCC38_26	0,51	Low	Positive
COMP_3.0_HCC38_27	0,41	Low	Negative

COMP_3.0_HCC38_28	0,43	Low	Positive
COMP_3.0_HCC38_29	0,59	Low-High	Negative
COMP_3.0_HCC38_30	0,39	Low	Negative
LATE_2.0_1649_1	1,89	High	Negative
LATE_2.0_1649_2	0,63	Low-High	NA
LATE_2.0_1649_3	1,62	High	Positive
LATE_2.0_1649_4	2,02	High	Positive
LATE_2.0_2138_1	3,3	High	Negative
LATE_2.0_2138_2	0,67	Low-High	NA
LATE_2.0_2138_3	1,08	High	Positive

Appendix 5. R-script for boxplot and dot plot

The scripts for the box plot and dot plots were run in R-studio and were made with help from Ole Christian Lingjærde and Arne Valebjørg Pladsen.

```
dat = read.table(paste0(file.dir, "DatatilR.txt"), header=F, sep="\t", stringsAsFactors = F, dec=",")
colnames(dat) = c("ID","concentration","Run")
dat$method=c(rep(0,30),rep(1,30),rep(2,30),rep(3,30))
dat$method.label = c(rep("1.0",30),rep("1.1",30),rep("2.0",30),rep("3.0",30))


file.name = paste0(file.dir, "qubit_data4.pdf")
pdf(file.name, width=12, height=8)
par(mfrow=c(3,1), mar=c(3,4,1,2))

# Boxplot of four populations
boxplot(concentration ~ method.label, data=dat, col=c("green","red","magenta","blue"), xlab="")

# Scatterplot gives a bit more info
#plot(dat$concentration, col=c(rep("green",30),rep("red",30),rep("magenta",30),rep("blue",30)),
 #   xaxt="n",xlab="",ylab="Concentration", pch=16)
plot(1,1, type="n", xaxt="n",xlab="",ylab="Concentration", xlim=c(0,120), ylim=c(0, max(dat$concentration)))
abline(v=c(30.5, 60.5, 90.5), col="grey30", lwd=3)
abline(v=seq(1,400, by=10)-0.5, col="grey80")
abline(h=0)
points(1:length(dat$concentration), dat$concentration, col=c(rep("green",30),rep("red",30),rep("magenta",30),rep("blue",30)), pch=16)
axis(1,15+c(0,30,60,90),c("1.0","1.1","2.0","3.0"),tick=F)



# Sorted scatterplot gives a slightly different view
plot(dat$concentration, col="white", xaxt="n", xlab="", ylab="Concentration")
abline(v=c(30.5, 60.5, 90.5), col="grey")
abline(h=0)
points(1:30, rev(sort(dat$concentration[1:30])), col="green", pch=16)
points(31:60, rev(sort(dat$concentration[31:60])), col="red", pch=16)
points(61:90, rev(sort(dat$concentration[61:90])), col="magenta", pch=16)
points(91:120, rev(sort(dat$concentration[91:120])), col="blue", pch=16)
axis(1,15+c(0,30,60,90),c("1.0","1.1","2.0","3.0"),tick=F)
```

Figure 46. R-script for box plot and dot plots of sample concentrations.

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

Postboks 5003
NO-1432 Ås
Norway