

# **UNIVERSITÄTSKLINIKUM HAMBURG-EPPENDORF**

Institut für Tumorbiologie des Universitätsklinikums Hamburg-Eppendorf

Direktor: Prof. Dr. med. K. Pantel

## **BREAST CANCER HETEROGENEITY: GENETICS, ESTROGEN RECEPTOR, METASTASIS, AND TREATMENT**

### **Dissertation**

zur Erlangung des Doktorgrades PhD  
an der Medizinischen Fakultät der Universität Hamburg

vorgelegt von:

Anna Babayan  
aus Moskau

Hamburg 2016



## Table of Contents

|   |    |
|---|----|
| 1. SYNOPSIS .....   | 5  |
| 1.1. INTRODUCTION.....  | 5  |
| 1.1.1. Cancer and carcinogenesis.....   | 5  |
| 1.1.2. Breast cancer.....   | 5  |
| 1.1.3. Molecular subtypes of breast cancer.....   | 6  |
| 1.1.4. The role of estrogen receptor in breast cancer.....  | 8  |
| 1.1.5. Circulating tumor cells as source of distant metastases.....   | 8  |
| 1.1.6. Intra-patient heterogeneity in breast cancer.....  | 9  |
| 1.1.7. Cancer progression models .....  | 10 |
| 1.1.8. Radiotherapy resistance as function of cancer heterogeneity .....  | 11 |
| 1.1.9. Detection, isolation, and characterization of circulating tumor cells .....  | 12 |
| 1.1.10. Whole genome amplification of single (tumor) cells .....  | 13 |
| 1.1.11. Next generation sequencing of single (tumor) cells.....   | 14 |
| 1.2. THE PROJECTS (MATERIAL AND METHODS).....   | 16 |
| 1.2.1. Heterogeneity of estrogen receptor expression in circulating tumor cells<br>from metastatic breast cancer patients .....   | 16 |
| 1.2.2. Comparative study of whole genome amplification and next generation<br>sequencing performance of single cancer cells ..... | 17 |
| 1.2.3. Clonal evolution of metastatic breast cancer: two cases – two progression<br>models .....                                  | 18 |
| 1.2.4. RHAMM splice variants confer radiosensitivity in human breast cancer<br>cell lines .....                                   | 18 |
| 1.3. RESULTS.....   | 20 |
| 1.3.1. Heterogeneity of estrogen receptor expression in circulating tumor cells<br>from metastatic breast cancer patients .....   | 20 |
| 1.3.2. Comparative study of whole genome amplification and next generation<br>sequencing performance of single cancer cells ..... | 21 |
| 1.3.3. Clonal evolution of metastatic breast cancer: two cases – two progression<br>models .....                                  | 22 |
| 1.3.4. RHAMM splice variants confer radiosensitivity in human breast cancer<br>cell lines .....                                   | 26 |

|  |     |
|--|-----|
| 1.4. DISCUSSION .....  | 28  |
| 1.4.1. Analysis of circulating tumor cells as “liquid biopsy” .....                            | 28  |
| 1.4.2. Estrogen receptor heterogeneity in circulating tumor cells .....                        | 29  |
| 1.4.3. Intra-tumor heterogeneity as source of metastases .....                                 | 30  |
| 1.4.4. Whole genome amplification and next generation sequencing in single cell genomics ..... | 31  |
| 1.4.5. The role of circulating tumor cells in investigation of breast cancer clonality .....   | 33  |
| 1.4.6. Clonality-driven evolution of breast cancer.....  | 34  |
| 1.4.7. Parallel progression of breast cancer .....   | 35  |
| 1.4.8. Linear progression of breast cancer .....   | 36  |
| 1.4.9. Breast cancer heterogeneity and radiotherapy resistance .....                           | 37  |
| 1.5. CONCLUSION .....  | 40  |
| 2. LIST OF ABBREVIATIONS .....   | 41  |
| 3. REFERENCES .....  | 42  |
| 4. PUBLICATION 1 .....   | 56  |
| 5. PUBLICATION 2.....  | 68  |
| 6. PUBLICATION 3.....  | 103 |
| 7. PUBLICATION 4.....  | 132 |
| 8. SUMMARY (IN ENGLISH) / ZUSAMMENFASSUNG (AUF DEUTSCH) .....                                  | 145 |
| 9. ERKLÄRUNG DES EIGENANTEILS AN DEN PUBLIKATIONEN .....                                       | 148 |
| 10. DANKSAGUNG .....   | 149 |
| 11. LEBENSLAUF .....   | 150 |
| 12. EIDESSTATTLICHE ERKLÄRUNG .....  | 154 |

## 1. SYNOPSIS

### 1.1. INTRODUCTION

#### 1.1.1. Cancer and carcinogenesis

Cancer is a group of diseases caused by deregulation of cell cycle machinery, whereby normal cells undergo uncontrolled cell division. Normally cell cycle regulation is maintained by dynamic balance between proliferation and programmed cell death stimuli. These stimuli are produced by systems of proto-oncogenes (genes that stimulate cell proliferation) and tumor suppressor genes (genes that promote cell cycle arrest and programmed cell death), respectively. Activation of proto-oncogenes and inactivation of tumor suppressor genes lead to deregulation of the cell cycle and uncontrolled cell division [1].

Genetic and epigenetic aberrations cause activation of proto-oncogenes and inactivation of tumor suppressor genes. Genetic aberrations include mutations, copy number aberrations (CNAs), leading to gene dosage changes, and copy number neutral chromosomal aberrations, such as translocations. Epigenetic changes include aberrant methylation. Genetic and epigenetic changes ultimately reprogram a cell, promoting carcinogenesis. Uncontrolled cell division in combination with further evolution of cancer cells by natural selection in the body leads to cancer development. Cancer hallmarks include sustaining proliferative signaling, evading growth suppressors, resisting cell death, inducing angiogenesis, enabling replicative immortality, activating invasion and metastasis, avoiding immune destruction, tumor-promoting inflammation, genome instability and mutation, deregulation cellular energetics [2].

Cancer may derive from almost any cell type of the human body. However, each cancer is different according to its biology and pathophysiology. The most common cancers are lung, breast, colorectal, and prostate cancer, accounting each more than 1 million cases in 2012 worldwide [3].

#### 1.1.2. Breast cancer

Breast cancer is the most common malignancy in women with more than 1.6 million cases diagnosed in 2012 worldwide, accounting for approximately 25% of all cancer cases in women [3].

About 5-10% of all breast cancer cases are hereditary, caused by germ-line mutations in e.g. the *BRCA1* and *BRCA2* genes, the remaining 90-95% are sporadic.

Risk factors for developing sporadic breast cancer are classified as variable and invariable. Variable risk factors for breast cancer development include obesity, giving first birth late, dietary factors, hormone replacement therapy, intake of oral contraceptives, alcohol consumption, radiation, and exposure to mutagens; invariable factors include early menarche, late menopause, ageing, and having a family history of breast cancer [4].

Breast cancer is a heterogeneous disease in regard to epidemiology, morphology, histology, molecular organization, clinical behavior, therapy response, and dissemination patterns to distant sites. The most common histopathological types of breast cancer are invasive ductal carcinoma (IDC), ductal carcinoma in situ (DCIS), and invasive lobular carcinoma (ILC), with a prevalence of 55, 13, and 5%, respectively [5].

Therapy opportunities for breast cancer include surgery, irradiation, and systemic therapy in neoadjuvant and/or adjuvant setting. The therapeutic approach for a particular patient depends on the stage of the disease, presence or absence of metastases, expression of certain markers, and present comorbidities [6-8]. More specifically, type of surgery and irradiation regiment mostly depend on stage of the disease, its spread and present comorbidities, whereas prescription of systemic therapy, including chemotherapy, endocrine therapy, and targeted therapy, is mostly dependent on the subtype of the tumor [6, 7].

### **1.1.3. Molecular subtypes of breast cancer**

Subtypes of breast cancer have originally been identified on molecular level based on different gene expression profiles of a large set of breast tumors [9, 10].

The most common molecular subtype of breast tumors is luminal A, presenting 50-60% of all sporadic breast cancer cases [11, 12]. This subtype is characterized by the expression of genes activated downstream of the estrogen receptor (ER) pathway as in normal luminal epithelium of the mammary ducts. Because of low expression of genes related to cell proliferation, luminal A tumors are characterized by a relative good prognosis. On protein level, luminal A tumors demonstrate expression of ER, progesterone receptor (PR), keratin (K) 8/18, low expression of Ki67, and lack of expression of human epidermal growth factor receptor 2 (ERBB2) [9, 10, 13, 14].

The luminal B subtype represents 10-20% of all breast tumors and is characterized by a mixed expression of ER, PR, and uncommonly ERBB2. In contrast

to luminal A, the luminal B subtype is characterized by higher proliferation rates measured by the expression of Ki67, MKI67, and cyclin B1. Consequently, patients with luminal B tumors have a worse prognosis than patients with luminal A tumors [14].

The ERBB2-enriched subtype, often called HER2-positive, entails up to 20% of all breast cancer cases. It is characterized by amplification-related overexpression of the *ERBB2* gene, as well as high expression of ERBB2-pathway associated genes, and lack of ER/PR expression. Histopathological ERBB2-positive breast cancer is characterized by a highly proliferative phenotype and worse prognosis [14].

Basal-like breast cancers represent another molecular subgroup and are usually characterized by absence of ER, PR, and ERBB2 expression. In most cases these tumors demonstrate positivity for EGFR or K5/6. Additionally, basal-like tumors often demonstrate mutations in *TP53* gene, explaining their high aggressiveness [10]. Patients with basal-like tumors have a worse prognosis than patients with luminal tumors [14].

Normal-like breast tumors account up to 10% of all breast cancer cases. They are poorly characterized and have a prognosis and clinical outcome between that of luminal and basal-like tumors. Normal-like tumors are negative for ER, PR, ERBB2, but in contrast to basal-like tumors, normal-like carcinomas are also EGFR, and K5/6 negative [14, 15].

The latest identified molecular subtype is the claudin-low subtype (12-14% of all cases). Despite this subtype shares some characteristics with basal-like tumors, such as low expression of ER, PR, and ERBB2, claudin-low tumors overexpress a set of genes related to immune response, mesenchymal phenotype, and epithelial-mesenchymal transition (EMT). These features condition a poor prognosis [16-18].

Sophisticated molecular characterization of breast tumors has been adapted for simplified pathological examination to be used in the routine clinical practice. Pathological identification of breast cancer subtype is based on evaluation of ER and PR expression by immunohistochemistry (IHC), as well as fluorescence in situ hybridization (FISH) analysis of ERBB2 overexpression. These markers are important for therapy indication. ERBB2 overexpressing tumors are mostly treated with anti-ERBB2 therapy, whereas ER-positivity of a tumor is considered being a surrogate marker for endocrine therapy indication.

#### **1.1.4. The role of estrogen receptor in breast cancer**

ER-signalling plays a key role in the development of both normal and neoplastic breast tissue. Physiological activation of ER through binding with its ligands, estrogens, promotes and controls the development of the female secondary sex characteristics, regulation of menstrual cycle, and genesis of breast tissue and its further development after puberty and during pregnancy [19]. Moreover, ER-mediated signalling is involved into growth of ER-positive breast tumors [20]. Therefore, pharmacological inhibition of ER action through selective ER modulators (SERM), selective ER down-regulators (SERD), or aromatase inhibitors (AI) leads to interruption of the ER signalling pathway in cancer cells [21].

Endocrine therapy is widely used as adjuvant therapy in women with ER-positive breast cancer [22, 23]. Nevertheless, failure of endocrine therapy is observed in 30-40% of these women [24, 25]. Resistance to endocrine therapy can be caused by different mechanisms, leading to either lack of functional ER protein expression or dysfunction of the ER pathway [26]. As a consequence, endocrine therapy failure in ER-positive breast cancer patients leads to metastatic progress, which is the cause of 90% of the cancer-related deaths [27].

#### **1.1.5. Circulating tumor cells as source of distant metastases**

A putative source of distant metastases are circulating tumor cells (CTCs) – cells that have detached from the primary tumor or metastases and have spread into the circulation [28]. The ability to invade surrounding tissue and intravasate appear to be associated with epithelial-mesenchymal transition (EMT). EMT is a reversible process leading to dedifferentiation and promoted motility of tumor cells. EMT is associated with loose of cell-cell contacts, apical-basal polarization, altered adhesion, rearrangement of molecular markers and cytoskeleton organization. By undergoing the EMT, tumor cells switch partially or fully their epithelial phenotype into a mesenchymal one (rev. in [29]). By undergoing mesenchymal–epithelial transition (MET), the reverse process to EMT, CTCs obtain the ability to settle down in distant organs and give rise to metastases.

Despite half-life time of CTCs in circulation is <2.4h [30], investigation of CTCs present in blood of a patient at any certain time moment provides a snapshot of the actual disease status.

Quantification and characterization of CTCs in blood of cancer patients was introduced as a concept of “liquid biopsy” despite short half-life of CTCs in circulation. Regular enumeration of CTCs as a validated clinical biomarker can be utilized for disease prognosis, diagnosis of minimal residual disease, and monitoring of therapy effectiveness for breast, prostate, and colon cancer [31-34].

It has been shown that the presence of CTCs after completion of adjuvant therapy is a predictor of metastatic relapse and poor survival [32, 35]. Moreover, information provided by CTCs might be extended over the CTCs’ enumeration. Namely, CTCs might be investigated on proteomic, transcriptomic, and genomic levels. Despite transcriptome analysis on single cells is challenging, investigations of protein expression and genome-wide studies on single cells are becoming the state of the art in cancer research [36]. Characterization of CTCs provides insights into heterogeneity of the cancer and metastases.

#### **1.1.6. Intra-patient heterogeneity in breast cancer**

Heterogeneity of cancer is not limited to disease differences between patients, but also occurs within one patient. This intra-tumor, or intra-patient, heterogeneity can be observed on all levels of molecular organization: genomic, epigenomic, transcriptomic, metabolomic, and proteomic [36].

The current view on tumor heterogeneity is based on principles of Darwinian evolution. Natural selection leads to elimination of subclones with unfavorable for tumor progression genomic and epigenomic aberrations, while tumor promoting aberrations are maintained among subclones and confer survival advantage on the cells. Sequential waves of clonal expansion and changes in tumor microenvironment further drive genetic divergence of the subclones (rev. in [37]).

Investigation of protein expression in CTCs can provide a valuable information about intra-patient heterogeneity on proteomic level. ER expression in single CTCs can be used as marker of endocrine therapy efficacy and is therefore of particular interest. ER positivity of breast tumors determined by IHC is based on a cut-off of 1% of tumor cell positivity for the ER nuclear reactivity [38]. Therefore, CTCs arising from primary ER-positive breast tumors are not necessarily expected to be ER-positive. Heterogeneous ER expression in CTCs might be one of the reasons for endocrine therapy failure and the development of metastases in patients with ER-positive tumors treated with hormone therapy. It has been shown that divergence of ER status between

primary tumor and CTCs is not a rare event. Initially, these studies were based on PCR measurement of mRNA expression levels in an enriched for CTCs cell fraction [39-41]. However, this approach does not allow for investigation of intra-patient heterogeneity between individual CTCs. Investigation of ER expression on single cell level might shed light on the cause of endocrine therapy resistance in individuals and could ultimately lead to treatment optimization.

Intra-tumor heterogeneity on functional level, such as transcriptome, metabolome, and proteome, might be caused by niche adaptation mechanisms and varies through cell cycle dynamics, and thus does not necessarily reflect clonality of the cancer. Genomic heterogeneity, reflecting clonal origin of a cell lineage, is supposedly more stable and thereby providing accessible information about clonal evolution of cancer.

Molecular characterization of CTCs provides a powerful tool for investigation of intra-patient heterogeneity, obtaining information about the clonal origin of CTCs and clonal selection under therapy. Identification of therapy sensitive and resistant clones may provide new insights and potential targets for cancer treatment. Herewith, investigation of single cell genomics may provide the next step towards individualized therapy.

#### **1.1.7. Cancer progression models**

Genetic intra-tumor heterogeneity caused by clonal evolution of cancer is a well-known phenomenon in human cancers. Nevertheless, it has been long discussed whether metastatic dissemination is an early or late event in cancer evolution, resulting in development of two progression models.

The first model, the linear progression model, postulates that metastasis-initiating cells originate from most progressed clone(s) of the primary tumor, which were developed during evolution of the primary tumor with selection for clones with high metastatic proclivity [42, 43]. On the other hand, data showing the metastatic potential of primary tumors at early stages, led to the coinage of the parallel progression model [44, 45]. This model proposes the presence of metastatic potential already in the early disease progression, leading to early dissemination of CTCs into circulatory system with subsequent parallel and independent evolution of the primary tumor and metastases [46, 47]. An alternative scenario of cancer metastasis, proposed in our institute, suggests continuous dissemination of tumor cells from a primary tumor

developing higher metastatic potential over the time during further evolution of the primary tumor [48].

Understanding tumor progression and the metastatic cascade in breast cancer is of tremendous value because distant metastases development is the most challenging issue in clinical management of cancer. Investigation of progression mechanisms and clonal evolution in cancer could identify molecular signatures, involved in progression and metastatic process. Parallel genetic evolution of the primary tumor and distant metastases might explain failure of systemic endocrine therapy, which prescription is based on ER-positivity of the primary tumor. ER-positive primary breast cancers, treated with endocrine therapy, often demonstrate presence of ER-negative metastases, insensitive to anti-estrogen therapy [49, 50].

#### **1.1.8. Radiotherapy resistance as function of cancer heterogeneity**

Clonality of breast cancer might not only play a role in endocrine therapy resistance, but also in sensitivity and resistance to radiotherapy. Radiotherapy is almost never given alone as its accompanying application is beneficial in women with early and metastatic breast cancer [51-54]. The combination of radiotherapy and endocrine therapy is widely used in treatment of ER-positive breast cancer to improve patient survival [55], it is mostly provided as sequential to endocrine therapy (rev. in [56]). However, radiotherapy increases risk of ischemic heart disease [57] and thus should not be given without a clear marker-based indication. Moreover, correlation of radiotherapy resistance with resistance to endocrine therapy has been shown [58, 59]. Therefore discovery of markers able to predict cross-resistance to endocrine and radiotherapy is of particular interest.

On molecular level, overexpression of CD44 – a receptor for hyaluronan (HA), is associated with acquired endocrine therapy resistance in breast cancer cells [60]. The mechanism of CD44-associated endocrine therapy resistance relies on the ability of CD44 to promote proliferative signaling through its interaction with ERBB2 and EGFR [60, 61]. The hyperactivation of the ERBB2 and EGFR signaling pathways is known to limit response to endocrine therapy in ER-positive breast cancer [62, 63].

CD44 in conjunction with CD24 is a well-known marker for cancer stem cells (CSC) [64]. Moreover, it has been shown that CSC-like phenotype CD44<sup>+</sup>/CD24<sup>-/low</sup> is associated with radiotherapy resistance in cancer cells and may be induced by radiation even in differentiated breast cancer cells [65, 66]. The CD44<sup>+</sup>/CD24<sup>-/low</sup> tumor

cells are known to be more often present in basal-like breast tumors [67]. However, total expression of CD44 as measured by qRT-PCR was significantly higher in the luminal A subgroup compared to basal-like, luminal B, and ERBB2-enriched tumors [68]. This discrepancy might be explained by strong heterogeneity of luminal-type breast cancer with demonstrated presence of basal-like cells in luminal tumors [69-71]. Taken together, these data demonstrate the need of reliable markers responsible for sensitivity and resistance to radiotherapy. Recent studies suggest that aberrant apoptosis, driven by the p53 protein, may contribute to radiotherapy resistance [72, 73]. Receptor for hyaluronan-mediated motility (RHAMM), characterized as potential target protein of p53, is involved in radiation-induced apoptosis [74] and is highly expressed in luminal breast cancer cell lines [75]. It has been observed that luminal breast cancers are radiotherapy sensitive (rev. in [76]), however, very little is known about role of the 4 RHAMM isoforms in breast cancer development, progress, and therapy response.

One of the presented studies addresses the functional role of RHAMM-proteins in breast cancer as well as the relevance of its interaction with p53 with regard to therapeutic interventions supporting radiotherapy-based treatment decisions. In particular, the hypothesis was tested if RHAMM and its binding partner HA are eligible as therapeutic targets to sensitize breast cancer cells to ionizing radiation.

#### **1.1.9. Detection, isolation, and characterization of circulating tumor cells**

Minimal-invasiveness, easy accessibility, and the possibility of sequential blood collection make CTC analysis to a promising new blood-based biomarker [31, 77]. However, the need for dedicated technologies and expertise hamper CTC analysis.

Investigation of protein expression patterns as well as genomic aberrations in individual CTCs requires the detection and isolation of these cells. Low concentration of CTCs in the circulation makes the isolation challenging. Several existing enrichment techniques are based on the physical or immunological properties of CTCs (reviewed in [78, 79]). Whereas physical properties of the CTCs, such as size, might be not necessarily CTC-specific, immunological characteristics of the tumor cells are more likely to be CTC-specific. CTCs that originate from epithelial tumors (carcinomas) normally express epithelial markers such as EpCAM and keratins and lack expression of molecules typical for leukocytes, such as CD45 molecules. Therefore, immunocytochemistry (ICC) with the use of differently labeled antibodies against these

specific markers allow the discrimination between CTCs and leukocytes with simultaneous investigation of target protein expression, such as ER.

Identification, analysis, and isolation of individual CTCs can be expanded by genome-wide characterization. Characterization of genomic aberrations on single cell level is a powerful tool, allowing for the investigation of intra-tumor clonal heterogeneity and the metastatic cascade. Genome-wide characterization of single cells became first possible with recent advances in isolation of single cells, establishment of whole genome amplification (WGA), and development of next generation sequencing (NGS).

#### **1.1.10. Whole genome amplification of single (tumor) cells**

WGA prior to downstream genetic analysis of individual CTCs is required since a single cell does not contain enough DNA for biomolecular investigation. WGA was established in 1992 and used primarily for sperm typing [80, 81]. Very soon, WGA became applied in the preimplantation genetic diagnostic of human embryos [82, 83] and investigation of single tumor cells [84, 85].

The current existing WGA techniques can be grouped into three classes. The first class includes polymerase chain reaction (PCR) based methods. Production of short fragments is often seen being a disadvantage as DNA fragments less than 1 kb cannot be used in many downstream applications [86] especially in preimplantation genetic diagnostic [87].

The second class of WGA techniques is the multiple-displacement amplification (MDA), which is a non-PCR-based amplification method. MDA utilize the highly processive Phi29 DNA polymerase and random hexamer exonuclease-resistant primers. The following strand-displacement synthesis is an isothermal process. Products generated by MDA can be more than 10 kb in length [88, 89].

The third WGA class includes techniques that combine a brief MDA pre-amplification and a PCR amplification phase. Unlike the first two WGA methods, combined MDA-PCR provides quasi-linear amplification [88, 89].

Different concordance rates between non-amplified genomic DNA and DNA amplified with different WGA strategies have been reported in single nucleotide polymorphism (SNP) genotyping studies and CNA analysis [88, 90-96]. Moreover, these studies demonstrate that WGA might cause imbalanced amplification of alleles, leading to inaccurate results of CNA analysis. It has been shown, that unequal amplification of different sites is random and is not reproducible in different experiments

with the same DNA [97]. Therefore the amplification approach has to be chosen carefully depending on its specific characteristics, advantages, disadvantages, and the subsequent analysis [95, 98].

An important factor influencing WGA is material preservation. CTCs in blood may be preserved in special CellSave tubes in order to overcome clotting and for longer periods of storage. However, fixatives may inhibit DNA amplification and thereby hamper downstream analysis [99, 100]. Most tissue samples are conserved by formalin-fixation, and paraffin-embedding (FFPE), which is difficult to handle in biomolecular analysis due to formalin-induced cross-links [101]. Therefore, it is essential to have WGA methods compatible with these types of materials.

#### **1.1.11. Next generation sequencing of single (tumor) cells**

Downstream analysis of (amplified) DNA can be performed by massive parallel sequencing using NGS in order to identify SNPs, indels (insertions-deletions), loss of heterozygosity, structural variations, and copy number aberrations (CNA).

Although genomic aberrations can be investigated by array-comparative genomic hybridization (aCGH), the analysis on single cell level is challenging. The combination of pre-selected targets on the array on one hand and the random and incomplete genome amplification during WGA [96, 97, 102] on the other hand, can result in a high signal-to-noise ratio [103]. Furthermore, the resolution for whole genome analysis by aCGH is limited [104], in contrast, NGS provides the possibility to examine each nucleotide of the entire amplified product with single base resolution.

Existing NGS platforms differ by library preparation and signal detection approaches. Illumina's NGS technology is based on sequencing-by-synthesis approach. Currently, Illumina's HiSeq machines offer the highest throughput per run, nevertheless, a sequencing run can last several days [105, 106]. ThermoFisher's IonProton sequencers utilize semiconductor sequencing technology, based on detection of dNTPs incorporation by pH change. Despite this approach allows to complete a sequencing run within 4 hours, homopolymer stretches might be called incorrectly [105].

Taken together, methods for single cell analysis of CTCs, allowing for simultaneous characterization of the cells on both protein expression and genomic levels are of particular interest as they can provide valuable information about cancer biology as well as for identifying potential new targets and biomarkers for cancer

treatment. Genomic characterization of CTCs provides insights into genetic heterogeneity of the cancer and metastases and might aid clinical management of cancer patients due to identification of therapy sensitive and resistant clones. Herewith, investigation of single cell genomics may provide the next step towards individualized medicine.

In the studies presented here we 1) established and validated a highly sensitive approach to detect CTCs and simultaneously investigate their ER expression in blood samples of metastatic breast cancer patients; 2) investigated methodological basis for single cell genome-wide analysis; 3) investigated clonal evolution of human breast cancer on primary tissue and CTCs from two metastatic breast cancer patients; 4) investigated the functional role of RHAMM-proteins in BC as well as the relevance of its interaction with p53 with regard to therapeutic interventions supporting radiotherapy-based treatment decisions.

## 1.2. THE PROJECTS (MATERIAL AND METHODS)

The total study consisted of four major projects, each performed on the basis of previous project(s). Every project has been individually published or submitted for publication as follows:

1. Heterogeneity of estrogen receptor expression in circulating tumor cells from metastatic breast cancer [107].
2. Comparative study of whole genome amplification and next generation sequencing performance of single cancer cells [submitted]. Published 19.07.2016 in *Oncotarget*
3. Clonal evolution of metastatic breast cancer: two cases – two progression models [manuscript in preparation].
4. RHAMM splice variants confer radiosensitivity in human breast cancer cell lines [108].

### 1.2.1. Heterogeneity of estrogen receptor expression in circulating tumor cells from metastatic breast cancer patients

In the first project, we established a triple staining protocol for the detection and characterization of CTCs in blood of breast cancer patients. In order to simulate CTCs in blood, blood from healthy donors was spiked with human breast cancer cell line cells: ER-positive BT-474 and MCF-7 cells lines, and ER-negative BT-20, and MDA-MB-231. Prepared cytopspins of mononuclear cell fraction were used for the protocol establishment.

The established triple staining protocol allowed for the visualization of ER, CD45, and keratin (K) with the use of the dyes AlexaFluor488 (fluorescent green), NBT/BCIP (chromogenic dark blue), and Cy3 or AlexaFluor555 (fluorescent red), respectively. Additionally, nuclei were visualized by DAPI staining.

Subsequently, the protocol was applied to blood samples obtained from metastatic breast cancer patients. Keratin and DAPI positive, but CD45 negative cells were considered as CTCs. As proof of principal, 8 CTCs from 4 patients were individually picked by micromanipulation [109]. The quality of the WGA products was assessed by a multiplex PCR of the 100, 200, 300, and 400bp non-overlapping fragments of *GAPDH* gene as described elsewhere [110]. Subsequently mutation analysis of exon 4, 6, and 8 of the *ESR1* gene was performed.

Statistical analysis included comparison of CTC-positive and negative groups depending on clinical disease status (Fisher's exact test), survival analysis in

dependence on CTC status (Kaplan-Meier test), the groups of patients who received endocrine therapy vs. chemotherapy at the time of blood collection was calculated by Mann-Whitney-U-test.

### **1.2.2. Comparative study of whole genome amplification and next generation sequencing performance of single cancer cells**

In this project, we comprehensively investigated the performance and effectiveness of commercially available WGA techniques for whole exome sequencing by NGS on single and pooled tumor cells, and the independence of blood preservative. The performance of 3 WGA kits, representing 3 WGA methods, was analyzed in 4 groups of source material, different by origin and preservation method: A) individual SK-BR-3 cells obtained from EDTA-preserved blood; B) individual SK-BR-3 cells obtained from CellSave-preserved blood; C) single SK-BR-3 cells picked from a FFPE SK-BR-3 cell pellet; and D) individual CTCs obtained from EDTA-preserved blood from a breast cancer patient. Single tumor cells were obtained by spiking of healthy donors' blood, obtained in either EDTA or CellSave tubes, with SK-BR-3 cell line cells. The same cell line was previously formalin-fixed, paraffin-embedded, and stored for >3 years. Additionally, CTCs from metastatic breast cancer patients' blood, collected in EDTA tubes, were available. Single tumor cells from the fraction of mononuclear cells were enriched, stained, and picked according to the previously established protocol.

After DNA yield and quality per WGA kit were estimated, DNA of a single cell from each WGA group was used for whole exome NGS on 2 different platforms. Briefly, three SK-BR-3 cells, obtained from EDTA-preserved blood and amplified with Ampli1, PicoPlex, and REPLI-g kit, were analyzed with both HiSeq2000 and IonProton platforms.

Based on the obtained results, 1 NGS platform and 1 WGA kit were excluded for further experiments as they yielded results of insufficient quality. The next round of experiments included WGA of single and pooled cells in duplicate and NGS of obtained DNA in order to investigate the performance and the limit of detection with increasing amounts of material. Duplicates of 1, 3, 5, and 10 pooled SK-BR-3 cells obtained from CellSave-preserved blood and amplified with Ampli1 and PicoPlex kits were sequenced on HiSeq2000. Subsequently, a proof of principle experiment was performed on 2 individual CTCs obtained from EDTA-collected blood of a breast

cancer patient. The cells were individually amplified with PicoPlex WGA kit and sequenced on HiSeq2000.

### **1.2.3. Clonal evolution of metastatic breast cancer: two cases – two progression models**

In the presented project, we used methods and data analysis workflows, established in the described above projects for the investigation of clonal evolution of human breast cancer on primary tissue and CTCs from two metastatic breast cancer patients.

CTCs from blood of two metastatic breast cancer patients were enriched, stained for ER and K and picked according to the established protocol. From the two enrolled subjects, formalin-fixed, paraffin-embedded archival material of the primary tumors was cut in 5 µm thick sections, processed as described before [111], stained for ER, and used for laser microdissection of ER-positive and ER-negative tumor fragments.

CTCs and tissue samples underwent WGA with the PicoPlex WGA kit and whole genome sequencing with Illumina's HiSeq2000 NGS platform. Raw NGS data was processed with the previously established pipeline for CNA analysis with the use of Control-FREEC tool with a window size of 500kb for segment calling [112, 113]. The analyses of samples were done for each patient separately and included unsupervised phylogenetic cluster analysis and support vector machine (SVM) analysis. Based on the obtained results, we were able to reconstruct clonal organization of the two investigated tumors and evolutionary pathways of the patients' diseases.

### **1.2.4. RHAMM splice variants confer radiosensitivity in human breast cancer cell lines**

To characterize the relevance of *RHAMM* expression in BC progression, mRNA expression data (Affymetrix) from 196 breast cancer tissue samples was analyzed in respect to clinicopathological factors. Two different BC cell line cells (MCF-7 and MDA-MB-231) were used to test whether *RHAMM* influences cell proliferation, apoptosis, or migration. To investigate the role of *RHAMM* in tumor progression in response to radiation, proliferation rate and cell death rate were characterized after 2Gy ionizing radiation. To investigate the role of *RHAMM* in response to radiation, both cell lines

were irradiated with 2Gy and *RHAMM* expression was evaluated by western blotting, ICC, and quantitative real-time PCR (qRT-PCR).

To establish the radiosensitizing ability of RHAMM - observed in terms of apoptosis - and to investigate the involvement of the different *RHAMM* splice variants, cells were treated with siRNA against the respective mRNAs and subsequently irradiated. Sub-G1 cell cycle (apoptosis) analysis was performed by fluorescence-activated cell sorting (FACS). Determination of live/dead cell was done with the use of TrypanBlue, proliferation rate of the cells was determined as shown previously [114].

Functional analyses of transfected and irradiated cells were performed with the use of migration assay, investigated via time-lapse microscopy starting 24h after ionizing radiation with 2Gy, and intracellular signaling array with the use of PathScan® Stress and Apoptosis Signaling Antibody Array Kit (Cell Signaling Technology, Danvers).

Knockdown of p53 was performed to clarify p53 involvement into RHAMM regulation. The effect of pharmacological inhibition of RHAMM with the HA-synthase inhibitor 4-methylumbelliferone (4-MU) was investigated by culturing with 4-MU.

## 1.3. RESULTS

### 1.3.1. Heterogeneity of estrogen receptor expression in circulating tumor cells from metastatic breast cancer patients

We established a triple staining protocol for the simultaneous investigation of ER, CD45, and K expression, suitable for further single cell downstream analysis.

The protocol was developed on a CTC model system with the usage of blood of healthy volunteers spiked with 500, 100, and 40 human breast cancer cell line cells. We demonstrated a recovery rate of 79%±4% using the density gradient Ficoll centrifugation as the method for mononuclear cell enrichment.

The triple staining protocol was used for detection and investigation of CTCs in blood of 35 metastatic breast cancer patients, initially diagnosed with ER-positive primary tumor. Metastatic disease was diagnosed in these patients on average 7.2 years (0.5-17.0 years) after primary tumor removal.

CTCs were detected in 16 out of 35 samples (45.7%). Survival analysis starting from the time point of blood analysis until the end of the study (median follow up: 13.1 months, range 1-30 month), demonstrated significant correlation of CTC presence in the blood with shorter disease-free survival ( $p=0.038$ ). Moreover, detection of CTCs was significantly associated with clinical progression of the disease ( $p<0.001$ , two-sided Fisher's exact test).

Among all 16 CTC positive cases, 8 samples (50.0%) demonstrated homogeneous ER status: 3 samples (18.7%) had ER-negative CTCs only and 5 cases (31.3%) had ER-positive CTCs only. Eight out of 16 samples (50.0%) displayed both ER-negative and ER-positive CTCs. The average fraction of ER-negative and ER-positive CTCs in samples with mixed population was 36.8% and 63.2%, respectively.

Among all 16 CTC positive cases, 14 women received endocrine therapy (87.5%), whereas two (12.5%) did not. In the blood samples of women with ER-positive primary tumors that received endocrine therapy, ER-negative CTCs were found in 3/14 cases (21.47%), ER-positive CTCs in 4/14 cases (28.6%), and both ER-positive and ER-negative CTCs were detected in 7/14 patients (50.0%). Thus, the presence of ER-positive CTCs in patients whom received endocrine therapy was detected in 11/14 cases in total (78.6%) and ER-negative CTCs could be found in 10/14 cases (71.4%). Among the three patients in which only ER-negative CTCs were detected, two had progression of disease and therefore received chemotherapy by the time of blood sampling. One patient who developed distant metastases during endocrine therapy

was switched to chemotherapy after which remission of the disease was documented. Sequence analysis of exons 4, 6, and 8 of the *ESR1* gene of 8 individual CTCs did not reveal any mutations.

### **1.3.2. Comparative study of whole genome amplification and next generation sequencing performance of single cancer cells**

Three different WGA kits were used to amplify single cell samples of: individual SK-BR-3 cells picked from EDTA- and CellSave-preserved blood spiked with SK-BR-3 cells, single SK-BR-3 cells picked from FFPE material, and individual CTCs picked from EDTA-collected blood of breast cancer patients. In total, 192 cells underwent WGA.

Among all tested WGA kits REPLI-g demonstrated the highest DNA yield along all sample groups (on average 34.0  $\mu$ g), however with the lowest success rate (50% on average). Ampli1 and PicoPlex kits demonstrated comparable success rates (on average 93 and 95%, respectively), however DNA yield was higher in Ampli1-amplified samples (on average 6.1 and 3.7 $\mu$ g in Ampli1- and PicoPlex-amplified samples, respectively).

Comparison of sequencing platforms revealed the HiSeq2000 performing better than the IonProton platform in respect to produced reads, depth and breadth of target base coverage, and mapping rate.

The number of total, known, and novel SNPs identified with HiSeq2000 platform in single cells was higher than for the same cells sequenced with IonProton NGS regardless of the WGA method. Sensitivity of the SNP analysis was also higher in samples sequenced with HiSeq2000 with maximum 41.3 and 27.1% for Ampli1 and PicoPlex WGA experiments, respectively.

Correlation between CNA profiles of single cells and genomic DNA, compared by Spearman correlation, did not depend on WGA kit, but on NGS platform. Cells amplified with Ampli1, PicoPlex, and REPLI-g kits demonstrated a good ( $r < 0.7$ ), strong ( $r > 0.8$ ), and weak ( $r < 0.3$ ) correlation with genomic DNA, respectively.

To investigate the detection limit with increasing amounts of starting material for WGA, as well as the influence of CellSave preservative on WGA and NGS performance, we analyzed duplicates of 1, 3, 5, and 10 pooled SK-BR-3 cells amplified with Ampli1 and PicoPlex WGA kits and on Illumina's HiSeq2000. Fewer total and known SNPs and indels and more novel SNPs and indels were identified in cells from

CellSave-preserved blood than in cells from EDTA collected blood, resulting in lower sensitivity of SNP and indel calling for CellSave-preserved cells. Moreover, known SNPs identified in single cells from EDTA- and CellSave-preserved blood were overlapping only partly. Comparison by WGA kit demonstrates superiority of Ampli1 WGA over PicoPlex for SNP and indel analysis in single cells. Analyses of pooled cells demonstrated that PicoPlex performance significantly improved with the number of pooled cells (increasing amount of input DNA), whereas PicoPlex performance of CNA analysis was not affected by input amount. Ampli1 performance did not significantly improve with increase of input material in any case.

As proof of principle, 2 CTCs from a metastatic breast cancer patient were analyzed. CNA analysis demonstrated two different profiles, suggesting cancer genetic heterogeneity of this patient's disease. Both CTCs carry chromosome 1q gain, suggested being a universal genomic change in breast cancer [115]. Additionally, CTC-1 demonstrated chromosome 16p gain and 16q loss (typical aberrations for luminal breast cancer) in contrast to CTC-2, which was strongly characterized by chromosome 9p loss. SNP calling analysis revealed 1135 SNPs and 15 indels common in both cells. Mutation analysis revealed 5 missense mutations annotated in COSMIC database [116]. Mutations in genes *CHEK2*, *PRAME*, and *KIT* were present in both CTCs, mutation in gene *FGFR2* was detected in CTC-1 only and a mutation in gene *TP53* was found in CTC-2 only.

### **1.3.3. Clonal evolution of metastatic breast cancer: two cases – two progression models**

CTCs and FFPE primary tissue samples from two enrolled metastatic breast cancer patients were used for the investigation of the clonal organization of the breast cancer.

#### *Patient UKE243*

Patient UKE243 (1945-2012) was diagnosed with primary breast cancer of the right breast in 1992 and with collateral ER-positive and ERBB2-negative breast cancer of the left breast in 1999, and received endocrine treatment (aromatase inhibitor) in 2000-2005. The first metastasis (ER-positive, ERBB2-negative) was detected in 2009, at which the endocrine treatment with aromatase inhibitor (aromasin) was started. Due to further metastatic progress (2010, ER-positive) the treatment was switched to endocrine therapy with selective ER-modulator (fulvestrant), and in 2011 switched to

chemotherapy (docetaxel) due to further metastatic progress. Blood for CTC analysis was collected during the course of chemotherapy in November 2011.

The blood sample analysis revealed the presence of 270 CTCs in 1 ml of blood with heterogeneous ER expression (64% ER-positive and 36% ER-negative CTCs). In total, 42 CTCs were picked by micromanipulation for downstream analysis. The FFPE material of the second primary tumor, diagnosed in 1999, was used for obtaining 50 tissue sections containing each 10-20 cells using laser microdissection: 40% ER-positive, 40% ER-negative, and 20% with unknown ER status.

Unsupervised phylogenetic cluster analysis was performed on the CNA data from the primary tumor tissue; as a result, 5 clearly distinguishable clusters were formed. Next, CNA data of the CTCs were added to the CNA data of tissue samples and clustering was repeated. The resulting phylogenetic tree contained mixed CTC-tissue clusters. Subsequently support vector machine (SVM) analysis was performed to finally allocate the CTCs to the identified tissue clusters. Most of the CTCs were tackled by SVM analysis to the same tissue cluster as by phylogenetic cluster analysis on combined data. Among 42 analyzed CTCs, 12 CTCs resided to the first cluster, 11 to the second, 18 to the third, and 1 to the fourth, no CTCs were allocated to the fifth cluster. ER expression was heterogeneous among tissue samples and CTCs within each cluster. Chromosome 1q and 16p gain and chromosome 9p loss were present in all identified tissue clusters and respective CTCs. Based on distances between the clusters of the phylogenetic tree, we combined the clusters into 2 groups: the first group included clusters 1-3 and the second group contained clusters 4-5. Fisher's exact test of the 2 groups revealed significantly different CNAs: chromosome 4q and 8p losses were significantly more frequent in clusters 1-3, whilst chromosome 8p gains were more frequent in clusters 4-5. Because all CTCs except one resided to the tissue clusters 1-3, we compared the aberration frequencies between the two groups: tissue clusters 1-3 vs. CTCs. Significant differences were chromosomes 8q gain (tissue) and 1q and 7 gains and 16q and 22 losses (CTCs). Losses of chromosome 22 and 16q were found exclusively in CTCs.

Taken together, the evolutionary pathway of the disease could be schematically present as follows: initial or very early chromosomal aberrations included chromosome 1q and 16p gains, and 9p loss. These events probably caused chromosomal instability, required for further clonal evolution and progress of cancer. Chromosomal instability could lead to the development of at least 2 cell lineages. One lineage evolved towards

luminal subtype and gave rise to clones 1-3, depicted by clusters 1-3. These clones experienced further evolutionary progress encompassing gain of 8q after a number of cells had spread into the systemic circulation. These cells might have given rise to metastases after a certain dormancy period. CTCs, released from these metastatic lesions, reflect inherent CNAs from primary tumor clones, as well as CNAs of further evolution within metastatic lesion, like losses of chromosomes 16q and 22. Another lineage experienced further chromosomal aberrations, resulted in development of clones identified as clusters 4 and 5, characterized by high chromosomal instability.

Our results indicate that the metastases of the patient UKE243 arise from cells, disseminated from almost all subclones of the primary tumor, from the most earliest to very progressed ones. These findings are in line with parallel progression model of carcinogenesis and metastasis, suggesting that tumor cells acquire metastatic potential in the early stages of tumor progression.

#### *Patient UKE008*

Patient UKE008 (born 1978) was diagnosed with primary metastatic breast cancer in 2013 with multiple metastases in the spine and pelvis. Palliative therapy included irradiation of the primary tumor and systemic chemotherapy (paclitaxel, April – August 2013) in combination with anti-ERBB2 therapy (Trastuzumab and Pertuzumab, April 2013 – December 2015). The blood samples were collected before any systemic treatment was applied (1<sup>st</sup> sample) and 3 months after completion of the chemotherapy (2<sup>nd</sup> sample).

We detected 2 ER-negative CTCs in 7.5 ml blood of the 1<sup>st</sup> blood sample, collected before therapy (0.27 CTCs/ml) and 20 ER-negative CTCs per 1 ml in the 2<sup>nd</sup> blood sample. In total, 1 CTC from the 1<sup>st</sup> and 11 CTCs from the 2<sup>nd</sup> blood sample were collected for downstream analysis.

The primary tumor as well as one of the metastases in the L4 spine segment were biopsied and formalin-fixed and paraffin-embedded. The tumor as well as metastasis were ER- and ERBB2-positive. Microdissected fragments of the primary tumor (n=6) and spine metastasis (n=5) were ER-positive in 50% and 40% of cases, respectively.

Unsupervised phylogenetic cluster analysis of the tissue data only was performed. Because the patient was diagnosed with primary metastatic breast cancer, cluster analysis was performed on the combined data obtained from the primary tumor and metastasis. The obtained phylogenetic tree demonstrated the presence of 3

clearly distinguishable clusters. Phylogenetic cluster analysis of the combined CTC and tissue data demonstrated 1 distinct CTC cluster in addition to the 3 previously identified tissue clusters. Subsequent SVM analysis tackled all CTCs (n=12) to the third tissue cluster. This discrepancy is explainable by the difference between the phylogenetic cluster analysis and the SVM. The cluster analysis identifies as many clusters as necessary according to the differences between the samples, whereas SVM analysis is not able to define new clusters. Taking this explanation into consideration, phylogenetic tree built on combined CTC and tissue data was considered as reflecting clonal organization the best: 3 distinct tissue clusters and 1 CTC cluster were identified. CTCs demonstrated highest similarity with the most progressed clone identified in the primary tumor and metastasis, but low probability of arising directly from that clone.

The tumor subclone represented in the first cluster contained data obtained from 2 fragments of the primary tumor. The second cluster (represented by data of the metastasis only) might be considered an intermediate evolutionary step towards cluster 3. The third cluster, representing the most progressed evolutionary step, was made up of data from both primary tumor's and metastasis' tissue fragments. These results indicate that metastatic outgrowth could be initiated by collective dissemination of tumor cells from the 2 cooperating clones within a CTC cluster. However, it cannot be excluded that cells from primary tumor clones disseminated not in a CTC-cluster, but as individual CTCs, arrived at the same distant location and cooperated there. Investigation of further metastatic lesions is needed to clarify mechanisms of metastasis-initiating dissemination in the patient.

Evolutionary history of the UKE008 patient's cancer might have been as follows: chromosome 17p loss and chromosome 17q and 19q gain might be initial or very early events in the carcinogenesis because these CNAs we identified in frequency plots of all the identified clusters. Later during carcinogenesis this early cancerous cell population branched in its evolution. One subclone experienced chromosome 4q loss and chromosome 6 gain and developed the clone, depicted by cluster 1. Possibly lineage, represented by clusters 2-3, originated from another branch. Further evolution of the lineage led via chromosome 1q, 8q and 11p gain and chromosome 11q loss towards the second clone (cluster 2), and additional gain of chromosome 7q resulted in cell clone, depicted by cluster 3. Cells from these cooperating clones disseminated

either as CTC-cluster or as individual cells and built up distant metastasis we investigated, which is therefore reflecting the clonal structure of the primary tumor.

This scenario does not answer the question where the CTCs came from: the primary tumor or the metastasis. However, based on the phylogenetic tree, CTCs did not reside to any of tissue clusters, but formed separate clusters. The fact that the patient demonstrated multiple metastases suggests that CTCs of the patient UKE008 arise from the metastasis we did not investigate. In this case, the uninvestigated metastases embody further steps in evolutionary progression of the cancer in line with the linear progression model.

#### **1.3.4. RHAMM splice variants confer radiosensitivity in human breast cancer cell lines**

The relevance of RHAMM expression in breast cancer progression was investigated using mRNA expression data (Affymetrix) from 196 breast cancer tissues. Increased RHAMM expression was significantly correlated with a decrease in overall and recurrence-free survival, and high tumor grade.

Two different breast cancer cell line cells (MCF-7 and MDA-MB-231) were used to investigate RHAMM influence on cell proliferation, apoptosis, and migration. No effect on cellular proliferation was observed 48h after transient inhibition of all *RHAMM* splice variants. However, sub-G1 analysis revealed that *siRHAMM* treatment significantly increased the rate of cell death in MCF-7, but not in MDA-MB-231 cells.

The proliferative capacity was not altered by 2Gy of ionizing radiation of both cell lines. MCF-7 cells demonstrated a significant increase in the apoptotic rate as measured by sub-G1 analysis, in contrast to MDA-MB-231 cells, which were found to be radio-resistant. RHAMM mRNA levels in response to 2Gy radiation, measured by qRT-PCR, were significantly reduced in MCF-7 cells. Downregulation of RHAMM was confirmed by ICC staining in MCF-7 cells and can be explained by significant increase of p53 and p38 in MCF-7 cells 48h after initial radiation. In MDA-MB-231 cells no change in expression of p53 and p38 was detected. MDA-MB-231 cells are characterized by mutation in *TP53* gene and endogenously increased level of p53 in the nucleus.

Expression of all 4 *RHAMM* splice variants before irradiation was significantly lower in MDA-MB-231 cells than in MCF-7 cells. Expression of *RHAMM* splice variants v1 and v2 decreased in MCF-7 cells as a consequence of radiation treatment in

contrast to MDA-MB-231 cells demonstrated no further decrease of *RHAMM* v1/v2 expression after irradiation.

In MCF-7 cells siRHAMMpan as well as siRHAMM v1/v2 increased the rate of apoptosis, whilst knockdown of RHAMM v3 and v4 did not induce apoptosis. In MDA-MB-231 cells, treatment with neither siRHAMMpan nor siRHAMMv1/v2 induced a significant increase of cell death. However, knockdown of p53 and subsequent upregulation of RHAMM v1/v2 increased the rate of cellular death in MDA-MB-231 cancer cells.

Treatment of cells with pharmacological inhibitor of HA, the main binding partner of RHAMM, increased the radiosensitivity of the MCF-7 cells with respect to apoptosis fourfold. Whereas MDA-MB-231 cells did not respond to the treatment alone, the susceptibility of the cells was increased after additional radiation.

In conclusion, our data suggests that RHAMM is involved in the radio-resistant phenotype of breast cancer cells. Detection of *RHAMM* isoform expression in correlation with the *TP53* mutation status might allow for prediction of the responsiveness to radiation. Importantly, pharmacological inhibition of HA, the main binding partner of RHAMM, could help to increase the radiosensitivity of both *TP53* wild type and mutated cancer type.

## 1.4. DISCUSSION

### 1.4.1. Analysis of circulating tumor cells as “liquid biopsy”

Individual CTCs may be used as “liquid biopsy” to study tumor heterogeneity and find therapy associated markers on proteomic, transcriptomic, and genomic levels [31, 117].

One of the best established protein markers for endocrine therapy prescription and monitoring is the estrogen receptor (ER). ER expression might be defined on both proteomic and transcriptomic levels [118, 119]. The concordance of ER status between primary tumor and CTCs in metastatic breast cancer patients has been shown in 23% [39], and in 55% [41] of cases using qRT-PCR. Despite qRT-PCR is often used for determining ER status of CTCs [39-41, 120], this approach does not allow for the investigation of intra-patient CTC heterogeneity due to measurement of an average ER expression in an enriched cell population instead of single cell analysis. However, single cell mRNA analysis is challenging and cannot be combined with further DNA analysis. Taking these arguments into consideration, we have investigated the expression of ER in CTCs in blood of breast cancer patients using immunocytochemistry (ICC). With this approach, we were able to simultaneously detect and characterize CTCs with the additional possibility for downstream genetic analyses using whole genome amplification (WGA).

We detected CTCs in 16 out of 35 patient samples (45.7%), which is within the range of published reports [121]. Because EpCAM might be downregulated in tumor cells that underwent EMT [117], we have used an EpCAM-free detection method in order to overcome this limitation and investigated ER expression in the individual keratin-positive CTCs.

To our knowledge, until now only three studies have been performed in which the authors have stained ER on single CTCs using ICC [122-124]. The limited number of studies, based on ICC for the investigation of CTCs, might be explained by the technical challenges. The following challenges have to be taken into consideration: the complications of nuclei permeabilization for antibody delivery, low level of ER, difficulties in unequivocal identification of CTCs in case of CD45+/K+ cells. A study by Bock and colleagues showed lower percentage of ER-positive CTCs (30%), however, the sample size of CTC positive metastatic breast cancer patients was relatively low (n=5) [123]. In the study of Nadal *et al.*, in contrast to our study, only non-metastatic breast cancer patients before any systemic treatment were enrolled and a volume of

30 ml blood per patient was analyzed. ER-negative CTCs were detected in 38.5% of women with ER-positive primary tumors, positive for CTCs [122]. In the recently published study the authors revealed concordance of the ER status between CTCs and primary tumor in 68% of cases [124]. In our study, concordance of ER expression in CTCs with the primary tumor was demonstrated in 81.3% of the cases.

Because of the small number of patients investigated, our follow up analysis is only of exploratory nature. Nevertheless, we were able to demonstrate that the detection of  $\geq 1$  CTCs in blood of metastatic breast cancer patients was significantly associated with clinical progression of the disease ( $p < 0.001$ ). Although the cut-off of at least 5 CTCs per 7.5 ml of blood is considered to be the threshold of high risk for early progression in metastatic breast cancer patients using the CellSearch system [33], meta-analysis of Zhang *et al.*, demonstrates prognostic value of the presence of single CTCs. Moreover, the authors demonstrated that the prognostic significance of CTCs in blood does not depend on the time point of blood collection [121], which is consistent with our results where blood samples were taken during therapy. However, a larger cohort with uniform treatment and longer follow-up will be required to prove the significance and clinical relevance of our findings. Moreover, the presence of CTCs in blood does not necessarily reflect the ability of the CTCs to survive in the blood stream and to spread to distant organs. The survival and metastatic potential of CTCs need to be investigated, eventually also by identification of genetic signatures associated with the spread of CTCs and their development into metastasis.

#### **1.4.2. Estrogen receptor heterogeneity in circulating tumor cells**

We observed the presence of ER-negative CTCs in blood of women with ER-positive primary tumors during or after endocrine therapy in 71% of cases: 21% had ER-negative CTCs only and 50% had both ER-positive and ER-negative CTCs.

The presence of ER-negative CTCs in patients with ER-positive breast cancer might be explained either by the heterogeneous expression of ER in the primary tumor, leading to release of both ER-positive and ER-negative cells into circulation or by the silencing of ER expression by genomic and/or epigenomic mechanisms. It has been hypothesized, that switching from an ER-positive to ER-negative status might be one of mechanisms to evade endocrine therapy (reviewed in [125, 126]). Our findings indicate that the development of distant metastases in women with an ER-positive primary tumors during or after endocrine therapy might be related to the presence of

ER-negative CTCs, because these cells are most likely not to be affected by endocrine therapy. However, this hypothesis does not explain presence of ER-positive CTCs in blood of 78.6% patients after completion of endocrine therapy. Therefore we hypothesize that ER-positive CTCs, which can be detected in blood of patients after completion of endocrine therapy, originate from (micro)metastases, which consist of tumor cells with a dysfunctional ER pathway and, consequently, are resistant to the hormonal blockade downstream of the ER. Several mechanisms of ER-positive cells to escape anti-ER therapy are known and result in alteration of either ER expression or ER function (reviewed in [126, 127]). Loss of normal ER function, independent of cause, is the reason for inefficacy of anti-ER agents.

Several mutations are thought to lead to the inactivation of ER and/or its ligand-independent functioning [26, 127]. Mutations in *ESR1* occur in approximately 1% of the primary breast tumors [128] and in 10% of the breast cancer metastases but not in the autologous primary tumors [129]. We have performed a pilot study in which we analyzed mutations of the *ESR1* gene in both ER-negative and ER-positive CTCs. These mutations may hamper the protein's function, but not its expression [26]. We were unable to detect any mutations in the 8 single cells from 4 patients investigated.

#### **1.4.3. Intra-tumor heterogeneity as source of metastases**

Although the origin of intra-tumor heterogeneity is not fully understood yet, it seems to play a major role in a complex process of carcinogenesis and development of metastatic disease [130-132]. Intra-tumor heterogeneity and clonal diversity per se might promote cancer evolution by serving more diverse input material for Darwinian selection [133]. The newly revised “seed-and-soil” hypothesis postulates that heterogeneity of cell characteristics, survival in the circulation, and effective homing in new environment are the crucial conditions for successful metastasizing [43]. Because only very few tumor cells meet these requirements, metastasis is a biologically inefficient process (rev. in [134]). However, high amount of CTCs with heterogeneous characteristics provide extensive source for potential metastases [48]. CTCs embody an intermediate step between primary tumor and metastases. CTCs reflect the biology of the primary tumor or metastases from which they originate [28]. Furthermore, CTCs carry characteristics potentially enabling metastases' establishment. Therefore the genetic makeup of CTCs may provide a unique insight into cancer evolution.

Genome-wide studies of cancer clonality on single cells require well established, reproducible approaches for WGA and NGS analysis. The challenge of single cell genome-wide studies lies between the need of DNA amplification and the introduction of PCR artefacts during WGA and NGS and identification of objective cell-specific genomic aberrations.

#### **1.4.4. Whole genome amplification and next generation sequencing in single cell genomics**

In order to establish a reliable WGA-NGS pipeline for single cell analysis several methodological aspects of WGA and NGS techniques were investigated. More specifically, the performance of 3 single cell WGA methods in combination with subsequent whole exome sequencing on 2 different NGS platforms was evaluated.

Illumina's HiSeq platforms are widely used in human genome research due to their accuracy. Sequencing with ThermoFisher's IonProton can be faster and more cost-effective per run, however, sequencing with IonProton may result in substantial decrease of effective coverage depth due to the high abundance of PCR and optical duplicates, thereby, hampering accurate SNP and indel calling. Emulsion PCR, utilized for library preparation in IonProton technology, is thought to be the main source of PCR duplicates [135]. Moreover, the introduction of indels is a well-documented disadvantage of the semiconductor sequencing, utilized in IonProton [105]. Nevertheless, our study shows that CNA analysis was not affected by the described disadvantages of semiconductor sequencing and demonstrated comparable results for samples sequenced on both NGS platforms.

Important applications of NGS, such as SNP/mutation, indel, and CNA calling, seem to be especially hampered in single cell analysis due to incomplete genome coverage as a result of WGA [89, 96, 102]. Our data suggest that Ampli1 outperforms PicoPlex and REPLI-g WGA kits for SNP/mutations and indel calling. However, adaptor-ligation PCR, utilized in some PCR-based WGA kits (e.g., Ampli1), has certain limitations. Site-specific digestion of template DNA prior to PCR by the MseI enzyme [136] results in a wide distribution of fragment lengths. *In silico* analysis demonstrated that only 38% of  $19 \times 10^6$  fragments produced by MseI restriction of the human genome have length 100-500bp that are adequate for exome-capturing and size-selection for library preparation.

Commercially available exome enrichment kits have not been optimized for WGA products. The usage of whole genome amplified DNA as template might drastically reduce capturing efficacy due to fragmented nature of the WGA product. In order to optimize single cell sequencing, current exome capturing regions should be adapted for use with short DNA fragments.

Although samples amplified with REPLI-g WGA kit (MDA-based technique) demonstrated the highest DNA yield from a single cell, the quality of the obtained DNA was remarkably low and insufficient for appropriate SNP/mutation, indel, and CNA analyses. Based on our experience and observations of De Bourcy *et al.* [95] and Bergen *et al.* [137], we conclude that input of at least 10ng of genomic DNA and tailoring of the MDA reaction to obtain just enough DNA for further analysis is a key to optimal MDA performance. Further biases in MDA-based WGA can distort CNA analysis and have been described elsewhere, these include uneven representation and non-specific amplification of the genome, a large variability in amplification bias among the products, chimera formation, and dislocated sequences [95, 138-140].

CellSave blood preservation could be of great value in e.g., multicenter studies. Nevertheless, our results suggest that single cells from EDTA-collected blood demonstrate higher sensitivity for SNP/mutation and indel analyses, than single cells from CellSave-preserved blood.

The results of the SNP/mutation and indel analyses significantly improved in samples amplified with PicoPlex kit with an increasing number of pooled cells (increasing amount of input DNA), whereas PicoPlex performance of CNA analysis was not significantly affected by the amount of input. In contrast, the results of SNP/mutation, indel, and CNA analysis in Ampli1-amplified samples did not significantly improve with increased input material. However, Ampli1-amplified samples demonstrated sensitivity rates of the SNP and indel analyses, similar to that of PicoPlex-amplified samples. Moreover, already 3 pooled cells from CellSave-preserved blood resembled CNA pattern of unamplified DNA with strong correlation, whereas Ampli1-amplified samples reached the same correlation level with 5 or more pooled cells.

Together with blood preservatives, fixatives and DNA staining agents provide another technical challenge in SNP/mutation analysis of single cells as they may introduce mutations that are amplified during WGA. However, a recent study from our lab has demonstrated genetic heterogeneity within a cancer cell line upon sequencing

single cells [141]. Therefore, it cannot be ruled out that the low concordance of SNP/mutation calling between single cells might also be the effect of heterogeneity in addition to WGA-introduced artifacts.

It has been noted that WGA strongly affects CNA analysis due to imbalanced amplification of alleles [88, 96]. Moreover, non-linear amplification is random and is not reproducible for the same DNA template [97]. Although CNA analysis does not require exome capturing and is possible on whole genome shallow sequenced data, we performed CNA analysis on whole exome sequencing data and demonstrated that the quality of the obtained DNA by both Ampli1 and PicoPlex kits was adequate for qualitative assessment of CNA patterns.

Based on the obtained results, we conclude that CNA analysis of single cells is less hampered by WGA and NGS in comparison to SNP/mutation analysis. The best results for CNA analysis could be obtained with the use of PicoPlex WGA kit and NGS on Illumina's HiSeq2000 platform.

Intra-tumor heterogeneity and clonality of breast cancer can be investigated on genomic level with the use of SNP/mutation and CNA data. However, breast cancer is characterized by overall prevalence of CNAs over mutations [122]. This fact and our own results, implicating that CNA analysis of single cells is more robust than SNP/mutation analysis, suggest investigation of breast cancer clonality with the use of CNA analysis.

#### **1.4.5. The role of circulating tumor cells in investigation of breast cancer clonality**

Clonality and evolution of the cancer can be investigated on single cell level with the use of primary tumors, metastases, and/or CTCs. Primary tumors are removed or biopsied in the majority of cases, delivering material for investigation. Administration of systemic therapy is usually based on characteristics of the primary tumor. However, the metastases may not resemble the primary tumor anymore due to genetic progression or selection of treatment-resistant clones. The differences between primary tumor and metastases might be the reason for treatment failure. Therefore CTCs as "liquid biopsy" provide a unique, easy accessible source of tumor material [31].

CTCs that can be detected in the blood circulation many years after removal of the primary tumor are most likely coming from the metastases, because the half-life

time of CTCs in circulation is <2.4 hours [30]. It has been shown that dormant tumor cells in bone marrow may sometimes divide into micrometastases, which release CTCs, and thus cause the presence of CTCs in blood of metastases-free breast cancer patients many years after mastectomy, but in small concentrations ( $\leq 1$  CTC/ml) [30], which is in contrast with 270 and 20 CTC/ml found by us in blood of the UKE243 and UKE008 patients, respectively. These findings suggest that CTCs detected in blood of the enrolled in our study patients arise from metastases present in the body at the time point of blood sampling.

Technical obstacles can also hamper clonal analysis. Since comprehensive investigation of every single cell of the complete tumor is hardly possible, underrepresentation of certain clones as well as overrepresentation of other clones in a study cohort may lead to false reconstruction of tumor's clonal structure. Additionally, metastases are a difficult subject for clonal investigation. Distant metastases can be detected first when they reach a certain size, and are infrequently biopsied. Nevertheless, our results demonstrate the feasibility of archival material accompanied with CTCs for investigation of clonal evolution of human breast cancer. Further research is needed to obtain information about the genetic heterogeneity of the metastases and possible identification of therapy sensitive and resistant clones.

#### **1.4.6. Clonality-driven evolution of breast cancer**

Intra-tumor heterogeneity of breast cancer is a result of clonal expansion. In order to reconstruct cancer evolution and clonal organization on single cell level [142] one should assume that the tumor at any moment of the evolution contains all previous clones, or at least the most crucial ones. However, this assumption contradicts the Darwinian theory applied to carcinogenesis [143, 144]. According to the theory inter-clonal competition should lead to outcompeting of particular, not necessarily less aggressive clones [145], resulting in secondary mono- or oligoclonal structure of the primarily polyclonal tumor. As consequence the reconstructed clonal structure of the primary tumor does not necessarily reproduce cancer evolution.

Alternative look from an ecological perspective suggests that subclones within a tumor can be seen as individual units interacting with each other and their environment. This theory implies that not only competition, but other types of interaction, e.g. cooperation, are possible (rev. in [133]) and finds confirmation in cancer model systems [145-148]. Consequently, different cancer clones are not

necessarily overgrown by one dominant clone, and poly- or oligoclonal structure of cancer can be revealed.

The results obtained in our study of the breast cancer clonality demonstrated oligoclonal structure of the investigated breast tumors, indicating that both mechanisms, competition and cooperation of tumor clones, are likely being involved in cancer evolution. Additionally, the results indicate that breast cancer might utilize both linear and parallel progression ways.

#### **1.4.7. Parallel progression of breast cancer**

The results obtained from patient UKE243 suggests a parallel progression of the breast tumor. CTCs were detected in the blood of patient UKE243 12 years after the primary tumor was removed. Based on bioinformatics analysis all CTC resided to 4 out of 5 phylogenetic clusters identified in data of the primary tumor. These results suggest that metastases might have been founded by tumor cells that disseminated from multiple subclones of the primary tumor. In consideration of the time gap between primary tumor removal and detection of the first metastasis (10 years), it is likely that disseminated tumor cells underwent dormancy for a certain period before giving rise to distant metastases.

One question which may arise is whether metastases and CTCs of the patient UKE243 originate from the first primary tumor, diagnosed in 1992, or from the second contralateral primary tumor, diagnosed in 1999. The later tumor only was available for our analysis. According to the histology of both primaries, the metastases corresponded to the second primary tumor, which can be confirmed by our cluster analysis. Nevertheless, we cannot exclude the possibility that metastases and subsequent CTCs originate from the first primary tumor.

Additionally, primary tumor as well as CTCs of patient UKE243 demonstrated heterogeneous ER expression. Outgrowth of further ER-positive metastases and presence of ER-positive CTCs after the completion of endocrine therapy suggests endocrine therapy failure in this patient. Since we did not find mutations in ER-coding gene (*ESR1*) in the CTCs of the patient, endocrine therapy failure might have been caused by other mechanisms, e.g. epigenetic mechanisms or a dysfunctional ER-pathway.

Based on the observed CNA frequencies in identified clusters and CTCs, we conclude the existence of at least two lineages of tumor cells in the primary tumor of

the patient UKE243. One of the lineages, presented by clones 1-3 (cluster 1-3), is characterized by chromosome 1q and 16p gain and 16q loss. CNA profile 1q+/16p+/16q- is associated with ER-positivity, luminal A gene expression pattern, moderate to high differentiated tumors, and better outcome [149]. The second lineage, depicted by clusters 4 and 5, exhibited CNA patterns typical for basal-like breast cancer [149]. Basal-like subtype, typically ER-negative, is characterized by higher chromosomal instability than luminal subtypes [150]. Moreover, our results suggest that the clonal split happened at a very early stage of carcinogenesis. One of the lineages experienced further luminal-like differentiation, whereas the second lineage retained basal-like characteristics.

Presence of both basal- and luminal-like cell lineages in breast tumors has been demonstrated by others [69, 70, 151]. One of the possible explanations of the coexistence of basal- and luminal-like cells within a tumor can be given through the hypothesis that ER-positive cells, e.g. cells of luminal B subtype, and basal-like cells may arise from the same bipotent progenitor cell [115, 132]. Moreover, recently Cleary *et al.* demonstrated cooperation between basal- and luminal-like subclones playing a role in tumor maintenance [152].

Li *et al.* demonstrated in a mouse model that activation of Wnt signaling pathway transforms mammary progenitor cells, promoting heterogeneity of outgrowing cell lineages. The authors conclude that basal- and luminal-like lineages within the same tumor supposedly derive from a bipotent malignant progenitor cell [151]. Mammary progenitor cells are typically ER-negative, but originating lineages might undergo luminal-like differentiation and become ER-positive [115, 132, 153].

#### **1.4.8. Linear progression of breast cancer**

In contrast to patient UKE243, our data obtained from the cancer from patient UKE008, suggest linear progression to metastases. According to the linear progression model, distant metastases originated from cells, disseminated from the primary tumor at late evolutionary stage(s).

Loss of chromosome 17p and gain of 17q were observed in frequency plots of all clusters identified on data of the patient UKE243, including the CTC cluster. Gain of chromosome 17q is typical for ERBB2-positive luminal B breast cancers (rev. in [149]), which is in agreement with the pathology report (ER-positive, ERBB2-positive tumor).

Loss of chromosome 17p is a common aberration in many cancers, including breast cancer [154], due to the location of tumor suppressor gene *TP53*.

We demonstrated that at least one distant metastasis carries the genomic signatures observed in the two clones of the primary tumor. This observation might have two explanations. First possible scenario suggests collective dissemination of tumor cells from these two clones within a mixed CTC-cluster. Another explanation implies individual dissemination of the tumor cells from the two clones and subsequent cooperation at distant site. Whichever dissemination way took place, our results indicate interaction of the two tumor clones. The two genetically similar clones of the primary tumor might have interacted to obtain a selective growth advantage and/or metastatic propensity.

Evidence for cooperating clones can be found in mouse and fruitfly models. It has been shown that two cell populations can interact to promote tumorigenesis and obtain the ability to metastasize [146-148]. Moreover, interclonal cooperation contributes to tumor growth and progression [145]. Tumor cells from cooperating clones might disseminate collectively by formation of CTC-clusters. CTC clusters demonstrate an increased metastatic capacity in comparison to single CTCs [155].

A 74-fold increase of the amount of CTCs in 1 ml blood of patient UKE008 was found in comparison to baseline before therapy, 3 months after completion of chemotherapy but still under anti-ERBB2 therapy. All CTCs were found to be ER-negative, whereas the primary tumor and the metastasis were ER-positive. It has been shown that ER activity provides a way for ER-positive ERBB2-positive cells to escape ERBB2-targeted therapy [156]. Co-expression of ERBB2 and ER has been found in breast cancer patients [157-159]. The data suggests that ER-negative ERBB2-negative cells could escape therapy.

Although all CTCs identified in blood samples of the patient UKE008 were ER-negative, it cannot be excluded that ER-positive CTCs were still present in the body, but could not be detected. The cells might have been escaped to bone marrow and underwent dormancy. Alternatively, EMT-associated downregulation of epithelial markers on the cell surface might have hampered detection of these cells.

#### **1.4.9. Breast cancer heterogeneity and radiotherapy resistance**

Indirectly, it has been shown that clonality-driven heterogeneity of a tumor might play a role in development in radiotherapy resistance [160]. On molecular level

radiotherapy resistance has been correlated to CD44<sup>+</sup>/CD24<sup>-/low</sup> phenotype [161]. Radiotherapy resistance In luminal breast cancer is often observed in endocrine therapy resistant tumors [58, 59]. Unfortunately, very few data exists on the effects of radiation on endocrine therapy resistant cells. However, it has been shown that the *CD44* gene expression was elevated in tamoxifen resistance cell models [162].

CD44 is a widely expressed cell surface membrane receptor for hyaluronan (HA). HA is a glycosaminoglycan and an important component of the extracellular matrix. Elevated stromal levels of HA are associated with breast cancer progression and shorter overall survival of the patients [163]. *In vitro* HA induces breast cancer cell motility and survival [163, 164]. Notably, overexpression of CD44 was found to augment sensitivity of cancer cells to microenvironmental ligands, in particular HA [60].

Besides participation in cell-cell and cell-matrix interactions CD44 facilitates mitogenic/invasive and proliferative cellular phenotypes [165]. In particular, CD44 proteins are known to interact and modulate the activity of a diverse range of receptor tyrosine kinases including c-Met [166, 167], VEGFR-2 [168], ERBB2 [61] and EGFR [169]. The later two have been previously shown to limit endocrine response in ER-positive disease [62, 63]. A signaling loop including ERBB2, EGFR, CD44, and ER might be proposed. Interaction of ER and ERBB2 signaling pathways provides tumor cell a possibility to overcome ER or ERBB2 blockage [62, 156]. Additionally, ERBB2 signaling can be promoted by CD44-HA interaction [170]. In its turn, CD44 overexpression increases sensitivity of cells to HA [60]. This positive-feedback loop caused by cross-signaling between ER, ERBB2, and CD44 might cause cross-resistance to endocrine and radiotherapy.

Little is known about RHAMM and its role in cancer. However similarity of RHAMM and CD44 characteristics and functions allows to propose involvement of RHAMM in carcinogenesis. RHAMM and CD44 have at least three distinct characteristics in common: i) they share the same binding partner, HA, and mediate its signaling, ii) participate in growth factor-regulated signaling, and iii) both have been shown to be transcriptionally repressed by p53 [171]. Nevertheless, CD44 and RHAMM are not homologous proteins, are compartmentalized differently in the cell, and differ in the HA-binding mechanisms. Therefore they are likely to act by different mechanisms (rev. in [172]).

RHAMM participates in the ERBB2-pathway though its action with CD44. In one of the presented studies we demonstrated the presence of RHAMM being a marker for

radiotherapy sensitivity [108]. We also tested whether ER is involved in the radiosusceptibility. However, treatment of MCF-7 cells with the unspecific ER-antagonist ICI182780 did not abrogate the pro-apoptotic effect of ionizing radiation in this cell type. Possibly because of CD44-RHAMM co-operation. Conflicting functions of CD44 have been observed in experimental models of carcinogenesis and tumor progression in comparison to *in vivo* data, supposedly due to the presence or absence of RHAMM [173]. CD44 is known to co-operate with RHAMM and has been reported to compensate for loss of RHAMM.

Heterogeneity of RHAMM and CD44 expression and ligand-binding ability in cancer is not fully investigated yet, but is considered to play an important role in cancer progression [174, 175]. Recently it has been shown that RHAMM and CD44 expression do not correlate with HA-binding and demonstrate heterogeneity [75]. The authors could demonstrate heterogeneity of HA binding in phenotypically homogeneous cell lines. Moreover, 3D culture experiments revealed that different by HA-binding subpopulations of cancer cell lines demonstrate different characteristics. Low-binding cells (HA<sup>-/low</sup>) were poorly invasive/metastatic and fast-growing, whereas high-binding cells (HA<sup>high</sup>) were characterized by highly invasive/metastatic but slow-growing phenotype [75]. These observations suggest that HA-binding heterogeneity of primary tumor might play a role in clinically relevant traits and should be further investigated as providing clinically relevant models for assessing treatment efficacy. Future studies will have to address the role of CD44 and RHAMM for radiotherapy resistance of cancer cells, possibly also including cross-resistance to endocrine and radiotherapy. Additionally, our data raise the possibility that the response to radiotherapy in selected tumors may be improved by targeting RHAMM and its ligand HA.

Although heterogeneity of breast cancer itself can be utilized as prognostic marker [149, 176, 177], further research is needed to reveal a source and mechanisms of tumor heterogeneity, its role in the metastatic cascade and resistance to therapy. Further investigations might results in identification of reliable markers for therapy response.

## 1.5. CONCLUSION

Breast cancer heterogeneity is a well-known but hardly understood phenomenon, which might be responsible for metastases' development, recurrence of the disease, and resistance to therapies like endocrine, chemo- and radiotherapy.

We could demonstrate that primary tumor heterogeneity might play role in cross-resistance of luminal cancer to both endocrine and radiotherapy. The mechanism implies activation of the ERBB2 signaling pathway as a result of cells' response augmentation to microenvironmental stimuli due to overexpression of CD44 and RHAMM. Pharmacological inhibition of RHAMM resensitizes tumor cells to radiotherapy.

In order to investigate clonal heterogeneity of breast cancer on both protein expression and genomic levels, we established a multiplex immunostaining protocol, compatible with investigation by NGS of single cell DNA. This approach allows, for the first time, a simultaneous phenotype-genotype characterization of single cells for the investigation of intra-patient CTC heterogeneity and clonal evolution of cancer. This approach enables reconstruction of an approximate evolutionary pathway of the cancer disease in an individual patient.

The investigated breast cancer cases represent parallel and linear metastases progression model. Our results demonstrate that therapy resistant breast cancer metastases detected years after primary tumor removal may originate from early and more progressed clones developed in the primary tumor during carcinogenesis in one case. Analysis of the second case demonstrates that metastasis in primary metastatic breast cancer originates according to the linear progression model from the most progressed interacting clones of the primary tumor, and CTCs most probably resemble further metastases, not resembling the primary tumor anymore.

CTCs as "liquid biopsy" provide a unique, easy accessible source of tumor material. Our results demonstrate feasibility of genomic and protein expression analyses on single CTCs and underline the importance of "liquid biopsy" for companion diagnostics in metastatic breast cancer.

## 2. LIST OF ABBREVIATIONS

|         |   |  |
|---------|---|--|
| aCGH    | - | array-comparative genomic hybridization    |
| AI      | - | aromatase inhibitor                        |
| CNA     | - | copy number aberration                     |
| CSC     | - | cancer stem cell                           |
| CTC     | - | circulating tumor cell                     |
| DCIS    | - | ductal carcinoma in situ                   |
| EMT     | - | epithelial-mesenchymal transition          |
| ER      | - | estrogen receptor                          |
| ERBB2   | - | human epidermal growth factor receptor 2   |
| FACS    | - | fluorescence-activated cell sorting        |
| FFPE    | - | formalin-fixation, paraffin-embedding      |
| FISH    | - | fluorescence in situ hybridization         |
| HA      | - | hyaluronan                                 |
| ICC     | - | immunocytochemistry                        |
| IDC     | - | invasive ductal carcinoma                  |
| IHC     | - | immunohistochemistry                       |
| ILC     | - | invasive lobular carcinoma                 |
| Indel   | - | insertion-deletion                         |
| K       | - | keratin                                    |
| MDA     | - | multiple-displacement amplification        |
| MET     | - | mesenchymal–epithelial transition          |
| NGS     | - | next generation sequencing                 |
| PCR     | - | polymerase chain reaction                  |
| PR      | - | progesterone receptor                      |
| qRT-PCR | - | quantitative real-time PCR                 |
| RHAMM   | - | receptor for hyaluronan-mediated motility  |
| SERD    | - | selective estrogen receptor down-regulator |
| SERM    | - | selective estrogen receptor modulator      |
| SNP     | - | single nucleotide polymorphism             |
| SVM     | - | support vector machine                     |
| WGA     | - | whole genome amplification                 |
| 4-MU    | - | 4-methylumbelliferone                      |

### 3. REFERENCES

1. Giancotti, F.G., *Deregulation of cell signaling in cancer*. FEBS Lett, 2014. 588(16): p. 2558-70.
2. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation*. Cell, 2011. 144(5): p. 646-74.
3. Ferlay, J., et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012*. Int J Cancer, 2015. 136(5): p. E359-86.
4. Kaminska, M., et al., *Breast cancer risk factors*. Prz Menopauzalny, 2015. 14(3): p. 196-202.
5. Ehemann, C.R., et al., *The changing incidence of in situ and invasive ductal and lobular breast carcinomas: United States, 1999-2004*. Cancer Epidemiol Biomarkers Prev, 2009. 18(6): p. 1763-9.
6. Theriault, R.L., et al., *Breast cancer, version 3.2013: featured updates to the NCCN guidelines*. J Natl Compr Canc Netw, 2013. 11(7): p. 753-60; quiz 761.
7. Witherby, S., et al., *Advances in Medical Management of Early Stage and Advanced Breast Cancer: 2015*. Semin Radiat Oncol, 2016. 26(1): p. 59-70.
8. Hong, C.C., C.B. Ambrosone, and P.J. Goodwin, *Comorbidities and Their Management: Potential Impact on Breast Cancer Outcomes*. Adv Exp Med Biol, 2015. 862: p. 155-75.
9. Perou, C.M., et al., *Molecular portraits of human breast tumours*. Nature, 2000. 406(6797): p. 747-52.
10. Sorlie, T., et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications*. Proc Natl Acad Sci U S A, 2001. 98(19): p. 10869-74.
11. Eroles, P., et al., *Molecular biology in breast cancer: Intrinsic subtypes and signaling pathways*. Cancer Treat.Rev., 2012. 38(6): p. 698-707.
12. Sorlie, T., et al., *Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms*. BMC.Genomics, 2006. 7: p. 127.
13. Raica, M., et al., *From conventional pathologic diagnosis to the molecular classification of breast carcinoma: are we ready for the change?* Rom J Morphol Embryol, 2009. 50(1): p. 5-13.

14. Eroles, P., et al., *Molecular biology in breast cancer: intrinsic subtypes and signaling pathways*. *Cancer Treat Rev*, 2012. 38(6): p. 698-707.
15. Weigelt, B., et al., *Breast cancer molecular profiling with single sample predictors: a retrospective analysis*. *Lancet Oncol*, 2010. 11(4): p. 339-49.
16. Prat, A., et al., *Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer*. *Breast Cancer Res*, 2010. 12(5): p. R68.
17. Herschkowitz, J.I., et al., *Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors*. *Genome Biol*, 2007. 8(5): p. R76.
18. Hennessy, B.T., et al., *Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics*. *Cancer Res*, 2009. 69(10): p. 4116-24.
19. Anderson, E., *The role of oestrogen and progesterone receptors in human mammary development and tumorigenesis*. *Breast Cancer Res*, 2002. 4(5): p. 197-201.
20. Russo, J. and I.H. Russo, *The role of estrogen in the initiation of breast cancer*. *J Steroid Biochem Mol Biol*, 2006. 102(1-5): p. 89-96.
21. Sini, V., et al., *Endocrine therapy in post-menopausal women with metastatic breast cancer: From literature and guidelines to clinical practice*. *Crit Rev Oncol Hematol*, 2016.
22. Group, E.B.C.T.C., *Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials*. *Lancet*, 2005. 365(9472): p. 1687-1717.
23. Group, E.B.C.T.C., *Tamoxifen for early breast cancer: an overview of the randomised trials*. *Early Breast Cancer Trialists' Collaborative Group*. *Lancet*, 1998. 351(9114): p. 1451-1467.
24. Andre, F. and L. Pusztai, *Molecular classification of breast cancer: implications for selection of adjuvant chemotherapy*. *Nat.Clin.Pract.Oncol.*, 2006. 3(11): p. 621-632.
25. Loi, S., et al., *Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen*. *BMC.Genomics*, 2008. 9: p. 239.
26. Herynk, M.H. and S.A. Fuqua, *Estrogen receptor mutations in human disease*. *Endocr.Rev.*, 2004. 25(6): p. 869-898.

27. Weigelt, B., J.L. Peterse, and L.J. van 't Veer, *Breast cancer metastasis: markers and models*. *Nat Rev Cancer*, 2005. 5(8): p. 591-602.
28. Joosse, S.A., T.M. Gorges, and K. Pantel, *Biology, detection, and clinical implications of circulating tumor cells*. *EMBO Mol Med*, 2015. 7(1): p. 1-11.
29. Bednarz-Knoll, N., C. Alix-Panabieres, and K. Pantel, *Plasticity of disseminating cancer cells in patients with epithelial malignancies*. *Cancer Metastasis Rev.*, 2012. 31(3-4): p. 673-687.
30. Meng, S., et al., *Circulating tumor cells in patients with breast cancer dormancy*. *Clin Cancer Res*, 2004. 10(24): p. 8152-62.
31. Alix-Panabieres, C. and K. Pantel, *Circulating tumor cells: liquid biopsy of cancer*. *Clin.Chem.*, 2013. 59(1): p. 110-118.
32. Pantel, K., C. Alix-Panabieres, and S. Riethdorf, *Cancer micrometastases*. *Nat Rev Clin Oncol*, 2009. 6(6): p. 339-51.
33. Cristofanilli, M., et al., *Circulating tumor cells, disease progression, and survival in metastatic breast cancer*. *N.Engl.J.Med.*, 2004. 351(8): p. 781-791.
34. Fehm, T., et al., *Micrometastatic spread in breast cancer: detection, molecular characterization and clinical relevance*. *Breast Cancer Res*, 2008. 10 Suppl 1: p. S1.
35. Serrano, M.J., et al., *Circulating tumour cells in peripheral blood: potential impact on breast cancer outcome*. *Clin Transl Oncol*, 2011. 13(3): p. 204-8.
36. Wang, D. and S. Bodovitz, *Single cell analysis: the new frontier in 'omics'*. *Trends Biotechnol*, 2010. 28(6): p. 281-90.
37. Zellmer, V.R. and S. Zhang, *Evolving concepts of tumor heterogeneity*. *Cell Biosci*, 2014. 4: p. 69.
38. Hammond, M.E., et al., *American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer*. *J Clin Oncol*, 2010. 28(16): p. 2784-95.
39. Aktas, B., et al., *Comparison of estrogen and progesterone receptor status of circulating tumor cells and the primary tumor in metastatic breast cancer patients*. *Gynecol Oncol*, 2011. 122(2): p. 356-60.
40. Fehm, T., et al., *Detection and characterization of circulating tumor cells in blood of primary breast cancer patients by RT-PCR and comparison to status of bone marrow disseminated cells*. *Breast Cancer Res*, 2009. 11(4): p. R59.

41. Tewes, M., et al., *Molecular profiling and predictive value of circulating tumor cells in patients with metastatic breast cancer: an option for monitoring response to breast cancer related therapies*. Breast Cancer Res Treat, 2009. 115(3): p. 581-90.
42. Chambers, A.F., A.C. Groom, and I.C. MacDonald, *Dissemination and growth of cancer cells in metastatic sites*. Nat Rev Cancer, 2002. 2(8): p. 563-72.
43. Fidler, I.J., *The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited*. Nat Rev Cancer, 2003. 3(6): p. 453-8.
44. van 't Veer, L.J., et al., *Gene expression profiling predicts clinical outcome of breast cancer*. Nature, 2002. 415(6871): p. 530-6.
45. Ramaswamy, S., et al., *A molecular signature of metastasis in primary solid tumors*. Nat Genet, 2003. 33(1): p. 49-54.
46. Klein, C.A., *Parallel progression of primary tumours and metastases*. Nat Rev Cancer, 2009. 9(4): p. 302-12.
47. Husemann, Y., et al., *Systemic spread is an early step in breast cancer*. Cancer Cell, 2008. 13(1): p. 58-68.
48. Joosse, S.A. and K. Pantel, *Genetic traits for hematogeneous tumor cell dissemination in cancer patients*. Cancer Metastasis Rev, 2016.
49. Rossi, S., et al., *Hormone Receptor Status and HER2 Expression in Primary Breast Cancer Compared With Synchronous Axillary Metastases or Recurrent Metastatic Disease*. Clin Breast Cancer, 2015. 15(5): p. 307-12.
50. Yang, Y.F., et al., *Discordances in ER, PR and HER2 receptors between primary and recurrent/metastatic lesions and their impact on survival in breast cancer patients*. Med Oncol, 2014. 31(10): p. 214.
51. Budach, W., et al., *Adjuvant radiation therapy of regional lymph nodes in breast cancer - a meta-analysis of randomized trials- an update*. Radiat Oncol, 2015. 10(1): p. 258.
52. Fisher, B., et al., *Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer*. N Engl J Med, 2002. 347(16): p. 1233-41.
53. Early Breast Cancer Trialists' Collaborative, G., et al., *Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10,801 women in 17 randomised trials*. Lancet, 2011. 378(9804): p. 1707-16.

54. Steinauer, K., et al., *Radiotherapy in patients with distant metastatic breast cancer*. *Radiat Oncol*, 2014. 9: p. 126.
55. Kyndi, M., et al., *Estrogen receptor, progesterone receptor, HER-2, and response to postmastectomy radiotherapy in high-risk breast cancer: the Danish Breast Cancer Cooperative Group*. *J Clin Oncol*, 2008. 26(9): p. 1419-26.
56. Cecchini, M.J., et al., *Concurrent or Sequential Hormonal and Radiation Therapy in Breast Cancer: A Literature Review*. *Cureus*, 2015. 7(10): p. e364.
57. Darby, S.C., et al., *Risk of ischemic heart disease in women after radiotherapy for breast cancer*. *N Engl J Med*, 2013. 368(11): p. 987-98.
58. Paulsen, G.H., et al., *Changes in radiation sensitivity and steroid receptor content induced by hormonal agents and ionizing radiation in breast cancer cells in vitro*. *Acta Oncol*, 1996. 35(8): p. 1011-9.
59. Luzhna, L., A.E. Lykkesfeldt, and O. Kovalchuk, *Altered radiation responses of breast cancer cells resistant to hormonal therapy*. *Oncotarget*, 2015. 6(3): p. 1678-94.
60. Hiscox, S., et al., *Overexpression of CD44 accompanies acquired tamoxifen resistance in MCF7 cells and augments their sensitivity to the stromal factors, heregulin and hyaluronan*. *BMC Cancer*, 2012. 12: p. 458.
61. Bourguignon, L.Y., et al., *Hyaluronan promotes CD44v3-Vav2 interaction with Grb2-p185(HER2) and induces Rac1 and Ras signaling during ovarian tumor cell migration and growth*. *J Biol Chem*, 2001. 276(52): p. 48679-92.
62. Hiscox, S., et al., *Tamoxifen resistance in breast cancer cells is accompanied by an enhanced motile and invasive phenotype: inhibition by gefitinib ('Iressa', ZD1839)*. *Clin Exp Metastasis*, 2004. 21(3): p. 201-12.
63. Knowlden, J.M., et al., *Elevated levels of epidermal growth factor receptor/c-erbB2 heterodimers mediate an autocrine growth regulatory pathway in tamoxifen-resistant MCF-7 cells*. *Endocrinology*, 2003. 144(3): p. 1032-44.
64. Ricardo, S., et al., *Breast cancer stem cell markers CD44, CD24 and ALDH1: expression distribution within intrinsic molecular subtype*. *J Clin Pathol*, 2011. 64(11): p. 937-46.
65. Diehn, M. and M.F. Clarke, *Cancer stem cells and radiotherapy: new insights into tumor radioresistance*. *J Natl Cancer Inst*, 2006. 98(24): p. 1755-7.
66. Lagadec, C., et al., *Radiation-induced reprogramming of breast cancer cells*. *Stem Cells*, 2012. 30(5): p. 833-44.

67. Honeth, G., et al., *The CD44<sup>+</sup>/CD24<sup>-</sup> phenotype is enriched in basal-like breast tumors*. Breast Cancer Res, 2008. 10(3): p. R53.
68. Olsson, E., et al., *CD44 isoforms are heterogeneously expressed in breast cancer and correlate with tumor subtypes and cancer stem cell markers*. BMC Cancer, 2011. 11: p. 418.
69. Knox, A.J., et al., *Modeling luminal breast cancer heterogeneity: combination therapy to suppress a hormone receptor-negative, cytokeratin 5-positive subpopulation in luminal disease*. Breast Cancer Res, 2014. 16(4): p. 418.
70. Haughian, J.M., et al., *Maintenance of hormone responsiveness in luminal breast cancers by suppression of Notch*. Proc Natl Acad Sci U S A, 2012. 109(8): p. 2742-7.
71. Ito, T., et al., *Differences in stemness properties associated with the heterogeneity of luminal-type breast cancer*. Clin Breast Cancer, 2015. 15(2): p. e93-103.
72. Chang, L., et al., *PI3K/Akt/mTOR pathway inhibitors enhance radiosensitivity in radioresistant prostate cancer cells through inducing apoptosis, reducing autophagy, suppressing NHEJ and HR repair pathways*. Cell Death Dis, 2014. 5: p. e1437.
73. Kho, P.S., et al., *p53-regulated transcriptional program associated with genotoxic stress-induced apoptosis*. J Biol Chem, 2004. 279(20): p. 21183-92.
74. Sohr, S. and K. Engeland, *RHAMM is differentially expressed in the cell cycle and downregulated by the tumor suppressor p53*. Cell Cycle, 2008. 7(21): p. 3448-60.
75. Veiseh, M., et al., *Cellular heterogeneity profiling by hyaluronan probes reveals an invasive but slow-growing breast tumor subset*. Proc Natl Acad Sci U S A, 2014. 111(17): p. E1731-9.
76. Langlands, F.E., et al., *Breast cancer subtypes: response to radiotherapy and potential radiosensitisation*. Br J Radiol, 2013. 86(1023): p. 20120601.
77. Alix-Panabieres, C. and K. Pantel, *Challenges in circulating tumour cell research*. Nat Rev Cancer, 2014. 14(9): p. 623-31.
78. Pantel, K., R.H. Brakenhoff, and B. Brandt, *Detection, clinical relevance and specific biological properties of disseminating tumour cells*. Nat Rev Cancer, 2008. 8(5): p. 329-40.

79. Lianidou, E.S. and A. Markou, *Circulating tumor cells in breast cancer: detection systems, molecular characterization, and future challenges*. Clin Chem, 2011. 57(9): p. 1242-55.
80. Zhang, L., et al., *Whole genome amplification from a single cell: implications for genetic analysis*. Proc Natl Acad Sci U S A, 1992. 89(13): p. 5847-51.
81. Hubert, R., et al., *High resolution localization of recombination hot spots using sperm typing*. Nat Genet, 1994. 7(3): p. 420-4.
82. Kristjansson, K., et al., *Preimplantation single cell analyses of dystrophin gene deletions using whole genome amplification*. Nat Genet, 1994. 6(1): p. 19-23.
83. Snabes, M.C., et al., *Preimplantation single-cell analysis of multiple genetic loci by whole-genome amplification*. Proc Natl Acad Sci U S A, 1994. 91(13): p. 6181-5.
84. Dietmaier, W., et al., *Multiple mutation analyses in single tumor cells with improved whole genome amplification*. Am J Pathol, 1999. 154(1): p. 83-95.
85. Huang, L., et al., *Single-Cell Whole-Genome Amplification and Sequencing: Methodology and Applications*. Annu Rev Genomics Hum Genet, 2015.
86. Dean, F.B., et al., *Comprehensive human genome amplification using multiple displacement amplification*. Proc Natl Acad Sci U S A, 2002. 99(8): p. 5261-6.
87. Peng, W., H. Takabayashi, and K. Ikawa, *Whole genome amplification from single cells in preimplantation genetic diagnosis and prenatal diagnosis*. Eur J Obstet Gynecol Reprod Biol, 2007. 131(1): p. 13-20.
88. Barker, D.L., et al., *Two methods of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage panel*. Genome Res, 2004. 14(5): p. 901-7.
89. Macaulay, I.C. and T. Voet, *Single cell genomics: advances and future perspectives*. PLoS Genet, 2014. 10(1): p. e1004126.
90. Lovmar, L., et al., *Quantitative evaluation by minisequencing and microarrays reveals accurate multiplexed SNP genotyping of whole genome amplified DNA*. Nucleic Acids Res, 2003. 31(21): p. e129.
91. *GenomePlex Complete Whole Genome Amplification Kit. Product Information*.
92. Hanson, E.K. and J. Ballantyne, *Whole genome amplification strategy for forensic genetic analysis using single or few cell equivalents of genomic DNA*. Anal Biochem, 2005. 346(2): p. 246-57.

93. Paez, J.G., et al., *Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification*. *Nucleic Acids Res*, 2004. 32(9): p. e71.
94. Tranah, G.J., et al., *Multiple displacement amplification prior to single nucleotide polymorphism genotyping in epidemiologic studies*. *Biotechnol Lett*, 2003. 25(13): p. 1031-6.
95. de Bourcy, C.F., et al., *A quantitative comparison of single-cell whole genome amplification methods*. *PLoS One*, 2014. 9(8): p. e105585.
96. Pugh, T.J., et al., *Impact of whole genome amplification on analysis of copy number variants*. *Nucleic Acids Res*, 2008. 36(13): p. e80.
97. Talseth-Palmer, B.A., et al., *Whole genome amplification and its impact on CGH array profiles*. *BMC Res Notes*, 2008. 1: p. 56.
98. Blainey, P.C., *The future is now: single-cell genomics of bacteria and archaea*. *FEMS Microbiol Rev*, 2013. 37(3): p. 407-27.
99. Riethdorf, S., et al., *Detection of circulating tumor cells in peripheral blood of patients with metastatic breast cancer: a validation study of the CellSearch system*. *Clin Cancer Res*, 2007. 13(3): p. 920-8.
100. Qin, J., et al., *Stabilization of circulating tumor cells in blood using a collection device with a preservative reagent*. *Cancer Cell Int*, 2014. 14(1): p. 23.
101. Joosse, S.A., E.H. van Beers, and P.M. Nederlof, *Automated array-CGH optimized for archival formalin-fixed, paraffin-embedded tumor material*. *BMC Cancer*, 2007. 7: p. 43.
102. Pinard, R., et al., *Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing*. *BMC Genomics*, 2006. 7: p. 216.
103. Sermon, K. and S.p. Viville, *Textbook of human reproductive genetics*. ix, 206 pages.
104. Mohlendick, B., et al., *A robust method to analyze copy number alterations of less than 100 kb in single cells using oligonucleotide array CGH*. *PLoS One*, 2013. 8(6): p. e67031.
105. Buermans, H.P. and J.T. den Dunnen, *Next generation sequencing technology: Advances and applications*. *Biochim Biophys Acta*, 2014.
106. van Dijk, E.L., et al., *Ten years of next-generation sequencing technology*. *Trends Genet*, 2014. 30(9): p. 418-426.

107. Babayan, A., et al., *Heterogeneity of estrogen receptor expression in circulating tumor cells from metastatic breast cancer patients*. PLoS One, 2013. 8(9): p. e75038.
108. Schutze, A., et al., *RHAMM splice variants confer radiosensitivity in human breast cancer cell lines*. Oncotarget, 2016.
109. Hannemann, J., et al., *Quantitative high-resolution genomic analysis of single cancer cells*. PLoS One, 2011. 6(11): p. e26362.
110. van Beers, E.H., et al., *A multiplex PCR predictor for aCGH success of FFPE samples*. Br J Cancer, 2006. 94(2): p. 333-7.
111. Joosse, S.A., et al., *Prediction of BRCA2-association in hereditary breast carcinomas using array-CGH*. Breast Cancer Res Treat, 2012. 132(2): p. 379-89.
112. Boeva, V., et al., *Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data*. Bioinformatics, 2012. 28(3): p. 423-5.
113. Boeva, V., et al., *Control-free calling of copy number alterations in deep-sequencing data using GC-content normalization*. Bioinformatics, 2011. 27(2): p. 268-9.
114. Quah, B.J., H.S. Warren, and C.R. Parish, *Monitoring lymphocyte proliferation in vitro and in vivo with the intracellular fluorescent dye carboxyfluorescein diacetate succinimidyl ester*. Nat Protoc, 2007. 2(9): p. 2049-56.
115. Melchor, L. and J. Benitez, *An integrative hypothesis about the origin and development of sporadic and familial breast cancer subtypes*. Carcinogenesis, 2008. 29(8): p. 1475-82.
116. Forbes, S.A., et al., *COSMIC: exploring the world's knowledge of somatic mutations in human cancer*. Nucleic Acids Res, 2015. 43 (Database issue): p. D805-11.
117. Joosse, S.A. and K. Pantel, *Biologic challenges in the detection of circulating tumor cells*. Cancer Res, 2013. 73(1): p. 8-11.
118. Wilson, T.R., et al., *Development of a robust RNA-based classifier to accurately determine ER, PR, and HER2 status in breast cancer clinical samples*. Breast Cancer Res Treat, 2014. 148(2): p. 315-25.
119. Du, X., et al., *The detection of ESR1/PGR/ERBB2 mRNA levels by RT-QPCR: a better approach for subtyping breast cancer and predicting prognosis*. Breast Cancer Res Treat, 2013. 138(1): p. 59-67.

120. Onstenk, W., et al., *Gene expression profiles of circulating tumor cells versus primary tumors in metastatic breast cancer*. *Cancer Lett*, 2015. 362(1): p. 36-44.
121. Zhang, L., et al., *Meta-analysis of the prognostic value of circulating tumor cells in breast cancer*. *Clin Cancer Res*, 2012. 18(20): p. 5701-10.
122. Nadal, R., et al., *Biomarkers characterization of circulating tumour cells in breast cancer patients*. *Breast Cancer Res*, 2012. 14(3): p. R71.
123. Bock, C., et al., *Heterogeneity of ERalpha and ErbB2 Status in Cell Lines and Circulating Tumor Cells of Metastatic Breast Cancer Patients*. *Transl Oncol*, 2012. 5(6): p. 475-85.
124. Kalinsky, K., et al., *Correlation of hormone receptor status between circulating tumor cells, primary tumor, and metastasis in breast cancer patients*. *Clin Transl Oncol*, 2015. 17(7): p. 539-46.
125. Osborne, C.K. and R. Schiff, *Mechanisms of endocrine resistance in breast cancer*. *Annu Rev Med*, 2011. 62: p. 233-47.
126. Garcia-Becerra, R., et al., *Mechanisms of Resistance to Endocrine Therapy in Breast Cancer: Focus on Signaling Pathways, miRNAs and Genetically Based Resistance*. *Int.J.Mol.Sci.*, 2012. 14(1): p. 108-145.
127. Riggins, R.B., et al., *Pathways to tamoxifen resistance*. *Cancer Lett.*, 2007. 256(1): p. 1-24.
128. Roodi, N., et al., *Estrogen receptor gene analysis in estrogen receptor-positive and receptor-negative primary breast cancer*. *J.Natl.Cancer Inst.*, 1995. 87(6): p. 446-451.
129. Karnik, P.S., et al., *Estrogen receptor mutations in tamoxifen-resistant breast cancer*. *Cancer Res.*, 1994. 54(2): p. 349-353.
130. Kimbung, S., N. Loman, and I. Hedenfalk, *Clinical and molecular complexity of breast cancer metastases*. *Semin Cancer Biol*, 2015. 35: p. 85-95.
131. Koren, S. and M. Bentires-Alj, *Breast Tumor Heterogeneity: Source of Fitness, Hurdle for Therapy*. *Mol Cell*, 2015. 60(4): p. 537-46.
132. Skibinski, A. and C. Kuperwasser, *The origin of breast tumor heterogeneity*. *Oncogene*, 2015. 34(42): p. 5309-16.
133. Marusyk, A. and K. Polyak, *Tumor heterogeneity: causes and consequences*. *Biochim Biophys Acta*, 2010. 1805(1): p. 105-17.

134. Vanharanta, S. and J. Massague, *Origins of metastatic traits*. *Cancer Cell*, 2013. 24(4): p. 410-21.
135. Balzer, S., et al., *Filtering duplicate reads from 454 pyrosequencing data*. *Bioinformatics*, 2013. 29(7): p. 830-6.
136. Carpenter, E.L., et al., *Dielectrophoretic capture and genetic analysis of single neuroblastoma tumor cells*. *Front Oncol*, 2014. 4: p. 201.
137. Bergen, A.W., et al., *Effects of DNA mass on multiple displacement whole genome amplification and genotyping performance*. *BMC Biotechnol*, 2005. 5: p. 24.
138. Voet, T., et al., *Single-cell paired-end genome sequencing reveals structural variation per cell cycle*. *Nucleic Acids Res*, 2013. 41(12): p. 6119-38.
139. Iwamoto, K., et al., *Detection of chromosomal structural alterations in single cells by SNP arrays: a systematic survey of amplification bias and optimized workflow*. *PLoS One*, 2007. 2(12): p. e1306.
140. Ning, L., et al., *Quantitative assessment of single-cell whole genome amplification methods for detecting copy number variation using hippocampal neurons*. *Sci Rep*, 2015. 5: p. 11415.
141. Cayrefourcq, L., et al., *Establishment and characterization of a cell line from human circulating colon cancer cells*. *Cancer Res*, 2015. 75(5): p. 892-901.
142. Navin, N., et al., *Tumour evolution inferred by single-cell sequencing*. *Nature*, 2011. 472(7341): p. 90-4.
143. Nowell, P.C., *The clonal evolution of tumor cell populations*. *Science*, 1976. 194(4260): p. 23-8.
144. Greaves, M. and C.C. Maley, *Clonal evolution in cancer*. *Nature*, 2012. 481(7381): p. 306-13.
145. Marusyk, A., et al., *Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity*. *Nature*, 2014. 514(7520): p. 54-8.
146. Calbo, J., et al., *A functional role for tumor cell heterogeneity in a mouse model of small cell lung cancer*. *Cancer Cell*, 2011. 19(2): p. 244-56.
147. Ohsawa, S., et al., *Mitochondrial defect drives non-autonomous tumour progression through Hippo signalling in Drosophila*. *Nature*, 2012. 490(7421): p. 547-51.
148. Wu, M., J.C. Pastor-Pareja, and T. Xu, *Interaction between Ras(V12) and scribbled clones induces tumour growth and invasion*. *Nature*, 2010. 463(7280): p. 545-8.

149. Kwei, K.A., et al., *Genomic instability in breast cancer: pathogenesis and clinical implications*. Mol Oncol, 2010. 4(3): p. 255-66.
150. Pikor, L., et al., *The detection and implication of genome instability in cancer*. Cancer Metastasis Rev, 2013. 32(3-4): p. 341-52.
151. Li, Y., et al., *Evidence that transgenes encoding components of the Wnt signaling pathway preferentially induce mammary cancers from progenitor cells*. Proc Natl Acad Sci U S A, 2003. 100(26): p. 15853-8.
152. Cleary, A.S., et al., *Tumour cell heterogeneity maintained by cooperating subclones in Wnt-driven mammary cancers*. Nature, 2014. 508(7494): p. 113-7.
153. Bhagirath, D., et al., *Cell type of origin as well as genetic alterations contribute to breast cancer phenotypes*. Oncotarget, 2015. 6(11): p. 9018-30.
154. Gao, Y., et al., *Genetic changes at specific stages of breast cancer progression detected by comparative genomic hybridization*. J Mol Med (Berl), 2009. 87(2): p. 145-52.
155. Aceto, N., et al., *Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis*. Cell, 2014. 158(5): p. 1110-22.
156. Wang, Y.C., et al., *Different mechanisms for resistance to trastuzumab versus lapatinib in HER2-positive breast cancers--role of estrogen receptor and HER2 reactivation*. Breast Cancer Res, 2011. 13(6): p. R121.
157. Pinhel, I., et al., *ER and HER2 expression are positively correlated in HER2 non-overexpressing breast cancer*. Breast Cancer Res, 2012. 14(2): p. R46.
158. Konecny, G., et al., *Quantitative association between HER-2/neu and steroid hormone receptors in hormone receptor-positive primary breast cancer*. J Natl Cancer Inst, 2003. 95(2): p. 142-53.
159. Harigopal, M., et al., *Multiplexed assessment of the Southwest Oncology Group-directed Intergroup Breast Cancer Trial S9313 by AQUA shows that both high and low levels of HER2 are associated with poor outcome*. Am J Pathol, 2010. 176(4): p. 1639-47.
160. Vicini, F.A., et al., *Long-term efficacy and patterns of failure after accelerated partial breast irradiation: a molecular assay-based clonality evaluation*. Int J Radiat Oncol Biol Phys, 2007. 68(2): p. 341-6.
161. Phillips, T.M., W.H. McBride, and F. Pajonk, *The response of CD24(-/low)/CD44+ breast cancer-initiating cells to radiation*. J Natl Cancer Inst, 2006. 98(24): p. 1777-85.

162. Hiscox, S., et al., *Tamoxifen resistance in MCF7 cells promotes EMT-like behaviour and involves modulation of beta-catenin phosphorylation*. *Int J Cancer*, 2006. 118(2): p. 290-301.
163. Auvinen, P., et al., *Hyaluronan in peritumoral stroma and malignant cells associates with breast cancer spreading and predicts survival*. *Am J Pathol*, 2000. 156(2): p. 529-36.
164. Tzircotis, G., R.F. Thorne, and C.M. Isacke, *Chemotaxis towards hyaluronan is dependent on CD44 expression and modulated by cell type variation in CD44-hyaluronan binding*. *J Cell Sci*, 2005. 118(Pt 21): p. 5119-28.
165. Hiraga, T., S. Ito, and H. Nakamura, *Cancer stem-like cell marker CD44 promotes bone metastases by enhancing tumorigenicity, cell motility, and hyaluronan production*. *Cancer Res*, 2013. 73(13): p. 4112-22.
166. Gordin, M., et al., *c-Met and its ligand hepatocyte growth factor/scatter factor regulate mature B cell survival in a pathway induced by CD74*. *J Immunol*, 2010. 185(4): p. 2020-31.
167. Singleton, P.A., et al., *CD44 regulates hepatocyte growth factor-mediated vascular integrity. Role of c-Met, Tiam1/Rac1, dynamin 2, and cortactin*. *J Biol Chem*, 2007. 282(42): p. 30643-57.
168. Tremmel, M., et al., *A CD44v6 peptide reveals a role of CD44 in VEGFR-2 signaling and angiogenesis*. *Blood*, 2009. 114(25): p. 5236-44.
169. Hatano, H., et al., *RHAMM/ERK interaction induces proliferative activities of cementifying fibroma cells through a mechanism based on the CD44-EGFR*. *Lab Invest*, 2011. 91(3): p. 379-91.
170. Misra, S., S. Ghatak, and B.P. Toole, *Regulation of MDR1 expression and drug resistance by a positive feedback loop involving hyaluronan, phosphoinositide 3-kinase, and ErbB2*. *J Biol Chem*, 2005. 280(21): p. 20310-5.
171. Godar, S., et al., *Growth-inhibitory and tumor-suppressive functions of p53 depend on its repression of CD44 expression*. *Cell*, 2008. 134(1): p. 62-73.
172. Turley, E.A., P.W. Noble, and L.Y. Bourguignon, *Signaling properties of hyaluronan receptors*. *J Biol Chem*, 2002. 277(7): p. 4589-92.
173. Maxwell, C.A., J. McCarthy, and E. Turley, *Cell-surface and mitotic-spindle RHAMM: moonlighting or dual oncogenic functions?* *J Cell Sci*, 2008. 121(Pt 7): p. 925-32.

174. Misra, S., et al., *Interactions between Hyaluronan and Its Receptors (CD44, RHAMM) Regulate the Activities of Inflammation and Cancer*. *Front Immunol*, 2015. 6: p. 201.

175. Baumgarten, S.C. and J. Frasor, *Minireview: Inflammation: an instigator of more aggressive estrogen receptor (ER) positive breast cancers*. *Mol Endocrinol*, 2012. 26(3): p. 360-71.

176. Marusyk, A., V. Almendro, and K. Polyak, *Intra-tumour heterogeneity: a looking glass for cancer?* *Nat Rev Cancer*, 2012. 12(5): p. 323-34.

177. Irshad, S., P. Ellis, and A. Tutt, *Molecular heterogeneity of triple-negative breast cancer and its clinical implications*. *Curr Opin Oncol*, 2011. 23(6): p. 566-77.

#### 4. PUBLICATION 1

##### Heterogeneity of Estrogen Receptor Expression in Circulating Tumor Cells from Metastatic Breast Cancer Patients

Anna Babayan<sup>1</sup>, Juliane Hannemann<sup>1</sup>, Julia Spötter<sup>1</sup>, Volkmar Müller<sup>2</sup>, Klaus Pantel<sup>1\*</sup>, Simon A. Joosse<sup>1\*</sup>

<sup>1</sup>Department of Tumor Biology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

<sup>2</sup>Department of Gynecology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

\*Both authors contributed equally

Published: PLoS One. 2013; 8(9): e75038. Published online 2013 Sep 18.

# Heterogeneity of Estrogen Receptor Expression in Circulating Tumor Cells from Metastatic Breast Cancer Patients

Anna Babayan<sup>1</sup>, Juliane Hannemann<sup>1</sup>, Julia Spötter<sup>1</sup>, Volkmar Müller<sup>2</sup>, Klaus Pantel<sup>1\*</sup>, Simon A. Joosse<sup>1,2,3</sup>

**1** Department of Tumor Biology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, **2** Department of Gynecology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

## Abstract

**Background:** Endocrine treatment is the most preferable systemic treatment in metastatic breast cancer patients that have had an estrogen receptor (ER) positive primary tumor or metastatic lesions, however, approximately 20% of these patients do not benefit from the therapy and demonstrate further metastatic progress. One reason for failure of endocrine therapy might be the heterogeneity of ER expression in tumor cells spreading from the primary tumor to distant sites which is reflected in detectable circulating tumor cells (CTCs).

**Methods:** A sensitive and specific staining protocol for ER, keratin 8/18/19, CD45 was established. Peripheral blood from 35 metastatic breast cancer patients with ER-positive primary tumors was tested for the presence of CTCs. Keratin 8/18/19 and DAPI positive but CD45 negative cells were classified as CTCs and evaluated for ER staining. Subsequently, eight individual CTCs from four index patients (2 CTCs per patient) were isolated and underwent whole genome amplification and *ESR1* gene mutation analysis.

**Results:** CTCs were detected in blood of 16 from 35 analyzed patients (46%), with a median of 3 CTCs/7.5 ml. In total, ER-negative CTCs were detected in 11/16 (69%) of the CTC positive cases, including blood samples with only ER-negative CTCs (19%) and samples with both ER-positive and ER-negative CTCs (50%). No correlation was found between the intensity and/or percentage of ER staining in the primary tumor with the number and ER status of CTCs of the same patient. *ESR1* gene mutations were not found.

**Conclusion:** CTCs frequently lack ER expression in metastatic breast cancer patients with ER-positive primary tumors and show a considerable intra-patient heterogeneity, which may reflect a mechanism to escape endocrine therapy. Provided single cell analysis did not support a role of *ESR1* mutations in this process.

**Citation:** Babayan A, Hannemann J, Spötter J, Müller V, Pantel K, et al. (2013) Heterogeneity of Estrogen Receptor Expression in Circulating Tumor Cells from Metastatic Breast Cancer Patients. PLoS ONE 8(9): e75038. doi:10.1371/journal.pone.0075038

**Editor:** Syed A. Aziz, Health Canada and University of Ottawa, Canada

**Received:** May 3, 2013; **Accepted:** August 6, 2013; **Published:** September 18, 2013

**Copyright:** © 2013 Babayan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This research was supported by the Roggenbuck-Stiftung. KP was supported by ERC advanced Investigator grant: ERC-2010-AdG\_2010317, grant agreement number 269081, DISSECT; and BMBF grant 01ES0724, Metfinder. SJ was supported by the Research Promotion Fund of the Medical Faculty UKE (FFM). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: s.joosse@uke.de (SAJ); pantel@uke.de (KP)

‡ These authors contributed equally to this work.

## Introduction

Breast cancer is the most common malignancy among women, accounting for approximately 23% of all cancer cases. Furthermore, breast cancer represents the most frequent cause of cancer related death in women worldwide [1]. On the molecular level, breast cancer is a heterogeneous disease and several molecular subtypes have been described based on gene expression profiles and immunohistochemistry [2–4] that might be explained by their cell of origin [5]. The most common subtype is the luminal A type, presenting up to 50–60% of all breast cancer cases [2,6]. These tumors are characterized by high estrogen receptor alpha (ER) expression and are - due to their low proliferation rate - associated with a relatively good prognosis [6,7]. The luminal B subtype

represents 10–20% of all breast tumors and is characterized by a mixed expression of ER $\alpha$ , PR, and/or ERBB2. It is often represented by an more aggressive phenotype of breast cancer with higher tumor grade [8].

A breast tumor's ER expression is normally assessed by immunohistochemistry and the definition of ER "positive" status is based on the presence of 1% or more ER positive tumor cells [9]. Expression of ER often mediates sensitivity of these tumors to hormonal treatment with either selective estrogen receptor modulators, such as tamoxifen, or aromatase inhibitors. Although the therapeutic efficacy of endocrine treatment for women with ER $\alpha$ -positive primary or metastatic disease has been clearly demonstrated [10,11], failure of therapy is observed in 20–25% of

patients [12,13]. More importantly, these patients demonstrate endocrine therapy "experienced progression" [12], meaning either *de novo* or acquired resistance to endocrine therapy. Resistance to endocrine therapy has been correlated to both ER-dependent [14] and ER-independent reasons [13]. To ER-dependent mechanisms belong genetic and/or epigenetic changes of the ER $\alpha$  gene, causing either lack of ER $\alpha$  protein expression or a dysfunctional ER $\alpha$  pathway [14] (e.g., due to *ESR1* promoter hypermethylation, expression of truncated isoforms of ER $\alpha$ , post-translational modifications, and other genetic changes of ER $\alpha$  [15]). ER-independent ways of acquired endocrine resistance include alteration in cell cycle and cell survival signaling molecules, activation of escape pathways [13]. Failure of systemic therapy may eventually lead to outgrowth of metastases in distant organs and cancer-related death.

The putative precursors of distant metastases are circulating tumor cells (CTCs). These cells have detached from the primary tumor, circulate in the bloodstream, and may finally extravasate to metastasize [16–20]. CTC analysis hold great promise to be used to monitor adjuvant therapy efficacy, as a prognostic marker, for early detection of minimal residual disease [19,21], and as a predictive marker for individualized cancer treatment [22]. Easy accessibility and possibility of sequential blood analyses make CTC analysis a promising new blood-based biomarker [22–25].

Several techniques have been developed for the enrichment and detection of CTCs, including assays based on cell size, immunological properties, and physical properties of the tumor cells (reviewed in [22,23]). CTCs might be discriminated from leukocytes with high precision using their origin specific markers. CTCs, originating from carcinomas, normally express epithelial markers such as EpCAM (epithelial cell adhesion molecule) and keratins, on the other hand, CD45 molecules, also known as leukocyte common antigen, are expressed on the surface of white blood cells only (reviewed in [26]). Thus, the use of differently labeled antibodies against these specific markers allows to distinguish between CTCs and leukocytes.

Recently it was shown that the presence of CTCs after completion of adjuvant therapy is a predictor of metastatic relapse and poor survival [19,27–30]. Additionally, prognostic information provided by CTCs might not be limited to the amount of CTCs only. CTCs might reflect the primary tumor's biology, including intratumoral heterogeneity. Breast tumors are considered being ER $\alpha$  positive if 1% of the cells show nuclear reactivity of any intensity by immunohistochemical investigation [9]. Therefore, CTCs arising from primary ER $\alpha$ -positive breast tumors are not necessarily expected to be ER $\alpha$ -positive. ER-negative CTCs might originate from ER-negative cells of ER-mosaic primary tumor [31], or ER-negative clones might be selected and get growth superiority under the pressure of anti-ER therapy [32]. Appearance of genomic or epigenomic aberrations might also result in the appearance of ER-negative CTCs [33]. Since endocrine treatment is dependent on the hormone receptor status and targets ER $\alpha$ -positive cancer cells only, CTC heterogeneity might be one reason for treatment failure and metastasis development in patients with ER $\alpha$ -positive tumors.

Both ER-positive and ER-negative cells can be identified in therapy naïve primary tumors. Moreover, it was shown that ER status changes from positive to negative in 2.5–17.0% of the cases after therapy [34] and changing is possible in both directions [35]. In metastatic breast cancer, a change of ER status in comparison to the primary tumor was found in 17% of the cases [36]. Moreover, it is proposed, that the change from ER-positivity to ER-negativity might be one of the mechanisms to evade hormonal treatment (reviewed in [13,33]).

Recent studies could show that divergence of hormone receptor status between primary tumor and CTCs is not a rare event [37–39]. However, in all of these studies polymerase chain reaction (PCR) assays were conducted based on the measurement of mRNA expression levels in a total CTC population. Using such an approach, intra-patient heterogeneity between single CTCs cannot be seen. Investigating ER $\alpha$  status of single CTCs might shed light on the cause of endocrine therapy resistance in individuals and ultimately lead to treatment optimization.

Therefore, in this study we present a highly sensitive approach to detect CTCs and simultaneously investigate their ER expression in blood samples from 35 metastatic breast cancer patients with ER-positive luminal primary tumors. Moreover, our method allows further genetic analyses of these single CTCs which is not possible in most of the commonly used immunostaining protocols due to fixation and crosslinking of the DNA.

## Materials and Methods

### Cell culture

Two ER-negative (BT-20 and MDA-MB-231) and two ER-positive (BT474 and MCF7) breast cancer cell lines were used. All cell lines were acquired from American Type Culture Collection and cultured under the prescribed conditions: MCF7, BT-20, and MDA-MB-231 cells were cultivated in DMEM (catalog no. E15-011, PAA Laboratories) at 37°C at 10% CO<sub>2</sub>; BT474 cells were cultivated in RPMI (catalog no. E15-039, PAA Laboratories) at 37°C at 5% CO<sub>2</sub>. Both media were supplemented with 10% fetal bovine serum (catalog no. E15-105, PAA Laboratories). Cells were grown in a 75 cm<sup>2</sup> flask until confluency was reached. Cells were harvested using trypsin/EDTA (catalog no. R001100; Gibco), washed with PBS (catalog no. 14190-094; Gibco), and resuspended in 1×PBS for either spiking experiments or cytospin preparation for direct staining.

### Patients and blood sampling

Thirty five metastatic breast cancer patients with ER-positive primary tumors were included into the study. Average time between primary diagnosis and diagnosis of metastases was 7.2 years (range: 0.5–17.0 years). Median follow up was 13.1 months (range 1–30 month) starting from the time point of blood analysis. Patient details are described in Table S1. Patients were treated for metastatic breast cancer at the University Medical Center Hamburg-Eppendorf and received therapy, according to international guidelines.

Blood from five apparently healthy women of age 25–35 years was included into the study to function as negative control for the establishment of our protocol. All patients and healthy volunteers gave written informed consent to be included into the study. The examination of blood samples in this study was carried out anonymously and was approved by the local ethics review board Ärztenkammer Hamburg under number OB/V/03.

Four to fourteen milliliters of blood were collected from each patient in EDTA tubes and processed within 24 hours. The density gradient Ficoll was used for mononuclear cell enrichment: full blood was transferred to a 50 ml tube containing 30 ml HBSS (catalog no. L2045; Biochrom) and centrifuged at 400×g for 10 minutes at 4°C. Supernatant was removed by pipetting and the cell pellet was resuspended in 30 ml 1×PBS. Cell suspension was added to 20 ml Ficoll (catalog no. 17-1440-03; GE Healthcare). The mixture was spun at 400×g for 30 minutes at 4°C without acceleration and deceleration. The interface and supernatant, containing the mononuclear cells (i.e., leukocytes and tumor cells), were transferred to a new 50 ml tube. The tube was filled with

1×PBS and centrifuged at 400×g for 10 minutes at 4°C. Supernatant was discarded and cell pellet was resuspended in 1 ml 1×H-Lysis buffer (catalog no. W1.1000; R&D Systems) and incubated for 3 minutes with gentle shaking at room temperature. Thirty milliliters of PBS was added and sample was centrifuged again at 400×g for 10 minutes at 4°C. Supernatant was discarded and pellet was resuspended in 5 to 10 ml 1×PBS for cytospin preparation. Cell count was determined by a Neubauer counting chamber. Approximately 700,000 cells were applied to each slide.

Five milliliters of healthy volunteers' blood were spiked with 500, 100, and 40 cell line cells. The density gradient Ficoll was used for mononuclear and spiked cells enrichment as described above. Prepared cytospins were stained with mouse IgG1 A43-B/B3-Cy3 labeled anti-human keratins 8/18/19 (Micromet) 1:500 for 30 min and counterstained with 4',6-diamidino-2-phenylindole (DAPI) to evaluate a recovery rate of spiked cells.

Cytospins for the CTC model system and controls were obtained by spiking 500 breast cancer cell line cells into 5 ml blood from healthy volunteers followed by Ficoll gradient mononuclear cell enrichment as described above.

### Antibody detection systems and the establishment of the triple staining protocol

Cytospins of MCF7 breast cancer cells spiked into blood of healthy volunteers were stained for keratins 8/18/19 using different detection methods using anti-keratin 8/18/19 primary antibodies (Table S2). Tested methods included horseradish peroxidase-, alkaline phosphatase- and beta-galactosidase based systems, as well as fluorescence. Slides were stained according to the protocols described in Table S2, part 1 (steps 1–6). Three single cells, positive for keratin staining, for each tested system were picked by micromanipulation and whole genome amplification (WGA) was performed. Table S2 demonstrates the whole procedure of the antibody detection including staining protocols (part 1), description of staining results (part 2), and possibility of WGA on single cells (part 3). Only two out of seven tested approaches could be used in the establishment of triple ER/K/CD45 staining: fluorescence and nitro-blue tetrazolium 5-bromo-4-chloro-3'-indolylphosphate (NBT/BCIP) based visualization provided clear staining without background and were compatible with micromanipulation and WGA of single cells. The remaining five systems demonstrated either inadequate staining and/or inhibited WGA. In detail, horseradish peroxidase substrates 3,3'-diaminobenzidine (DAB) and nickel/catachol-based DAB enhancement (HistoMark Orange from KPL) resulted in a strong background staining of leukocytes. TrueBlue substrate provided highly sensitive staining, but is stable in alcohols and water only and therefore hampering subsequent staining or single cell manipulation. New Fuchsin containing substrate of alkaline phosphatase showed high auto-fluorescence and was not compatible with WGA and therefore, was not suitable for the triple ER/K/CD45 staining establishment. Beta-galactosidase substrate X-Gal provides a clear turquoise stain without background. Unfortunately, staining results were not reproducible, and thus this method was also not suitable for the triple staining.

A single staining protocol for estrogen receptor (ER) was established on breast cancer cell line cytospins. The protocol was considered optimal when all cells were positive for ER staining on ER-positive breast cancer cell line cytospins (MCF-7 and BT474) and all cells were negative in case of ER-negative breast cancer cell lines (BT-20 and SKBR3). Figures of Data S1 show a clear positive staining for ER in MCF7 and BT474 cells (row A and C), whereas no signal could be detected in the ER negative BT-20 and MDA-

231 cells (row B and D) and no background was detected in both experiments.

A double staining protocol for ER and keratin (K) was established on cytospins of MCF7 breast cancer cells. Different fluorescent visualization systems in different combinations were tested, *i.e.*, Alexa Fluor 350, 488, 546, 555, 594, Cy3, and Cy5 (data not shown). The best results were obtained in combination of ER staining visualized with Alexa Fluor 488 dye and direct keratin-Cy3 staining. Figure A of Data S2 demonstrates clear distinguishable ER (green) and K (red) staining, allowing for easy localization of signals even in all channels merged.

Next, we established a CTC model system by spiking breast cancer cell line cells in blood of healthy volunteers. This model system was used for the optimization of double ER/K staining (Figure B of Data S2) in the natural context of blood cells mimicking the clinical situation, establishment of CD45 single staining, a combination of the protocols, and the adjustment of a final triple ER/K/CD45 staining protocol.

The single staining and final triple staining protocol were both validated for unspecific binding of the primary and secondary antibodies. Rabbit normal IgG was applied instead of specific primary antibodies in order to proof specific binding of anti-ER antibodies and unspecific binding of the secondary antibodies. Figure C Figure of Data S2 shows that in absence of specific primary antibody no green staining could be detected, meaning that anti-ER and secondary antibodies demonstrate specific binding only.

### Cytospin triple staining

Slides were dried overnight at room temperature and stained according to the protocol steps described in Table 1 with 3×3 min washing in TBS between each step. After staining, slides were mounted with cover slips and Dako Glycergel Mounting Medium (Dako, C0563).

Positive and negative staining controls were included for each procedure. Slides with MCF7 breast cancer cells spiked into blood of healthy volunteers were used as control. Positive control slide was stained according to the protocol; for negative (isotype) control, mouse normal IgG was applied instead of anti-ER antibodies.

The estimation of the ER staining intensity was based on the principle of the standard IRS scoring system [40,41] and included following grades: no staining (negative); a weak staining (positive); a moderate staining (positive); a strong staining (positive). ER-negative cell line cells were used as standard of negative staining.

### Micromanipulation and whole genome amplification of single cells

Picking and transfer of single cells was done according to the previously established protocol by *Hannemann et al.* [42]. Briefly, each cell was picked individually by the use of a micromanipulator (the microinjector CellTram Vario and micromanipulator TransferMan NKII, Eppendorf Instruments, Hamburg, Germany), transferred in a drop of PBS onto a silanized glass stick. The stick was immediately transferred into a 200 µl PCR reaction tube. Individual single cells in 200 µl PCR tubes can be stored at -80°C for further analysis.

Whole genome amplification was performed using the PicoPlex WGA Kit for single cells (Rubicon Genomics, R30050) according to the manufacturer's recommendations. The WGA product was cleaned up with NucleoSEQ spin columns (Macherrey-Nagel, Germany). DNA concentration of WGA products was measured with Nanodrop 1000 (Pecqlab, Erlangen, Germany). The total yield was 1.4–5.2 µg of DNA per sample.

**Table 1.** The established triple staining protocol for detection and characterization of ER expression on CTC.

| Step | Substance and antibodies                                      | Concentration          | Application time | Diluent                                | Manufacturer                         |
|------|---|------------------------|------------------|--|--------------------------------------|
| 1    | Paraformaldehyde  | 0.5%                   | 10 min           | PBS                                    | Merck, 1040051000                    |
| 2    | Triton X-100  | 0.1%                   | 10 min           | TBS                                    | Sigma, 110K01792                     |
| 3    | AB serum  | 10%                    | 20 min           | TBS                                    | Bio-Rad Medical Diagnostics, 805135  |
| 4    | Rabbit anti-human estrogen receptor SP-1                      | 1:50                   | 90 min at 37°C   | TBS + 0.005% Triton X-100              | Abcam, ab16660                       |
| 5    | Alexa Fluor 488 goat anti-rabbit                              | 1:500                  | 45 min           | TBS + 0.005% Triton X-100              | Invitrogen, A11008                   |
| 6    | Mouse anti-human CD45, clone HI30                             | 1:400                  | 45 min           | TBS + 0.005% Triton X-100              | BioLegend, 304002                    |
| 7    | Donkey anti-mouse IgG alkaline phosphatase labeled            | 1:35                   | 30 min           | TBS + 0.005% Triton X-100              | Abnova, PAB10741                     |
| 8    | Normal mouse IgG  | 1:250                  | 30 min           | TBS + 0.005% Triton X-100              | Millipore, 12-371                    |
| 9    | NBT/BCIP substrate  | According to datasheet | 15 min           | TBS + 0.005% Triton X-100              | Bio-Rad, 1706432                     |
| 10   | Mouse IgG1 A45-B/93 - Cy3 labeled anti-human keratins 8/18/19 | 1:500                  | 30 min           | TBS + 0.005% Triton X-100 + DAPI 1:500 | Micromet, commercially not available |

ER - estrogen receptor; CTC - circulating tumor cell; NBT/BCIP - nitro-blue tetrazolium and 5-bromo-4-chloro-3'-indolylphosphate; PBS - phosphate buffered saline; TBS - tris buffered saline.

doi:10.1371/journal.pone.0075038.t001

Quality control was done by multiplex PCR as described elsewhere [43]. Briefly, four primer set were used to amplify of 100, 200, 300, and 400bp non-overlapping fragments of GAPDH gene. One hundred fifty nanogram of genomic DNA of each single cell was taken into the PCR reaction. PCR products were analyzed in a 2% agarose TAE gel. Human leukocyte DNA was used as positive control for the multiplex PCR. Negative control probe did not contain any DNA.

#### ESR1 mutation analysis

Exons 4, 6, and 8 of the gene *ESR1* (estrogen receptor 1) were amplified and sequenced. PCR was performed using AmpliTaq Gold DNA Polymerase (Applied Biosystems, N808-0240) under following conditions for each individual probe: 0.2 mM each of ATP, GTP, CTP, TTP; 2 pmol each primer; 1.25 U of Taq polymerase; 10 ng of DNA. Concentration of MgCl<sub>2</sub> required was established experimentally and represented 3 mM for amplification of exons 4, 6, and 8. Oligonucleotide primers 5'-3', used for *ESR1* mutational analysis of exon 4: forward ACATGAGAGCTGCCAACCTT, reverse CCCCACTATTTCTCC-CATGA; exon 6: forward CCCTTTCATGTCTTGTGGAAG, reverse ATGCCTTTGGAGTGGGTAGA; exon 8: forward GCTCGGGTTGGCTCTAAAGT, reverse ATGCGATGAAG-TAGAGCCCG.

PCR products were analyzed in 2% agarose gel. Sequencing PCR was performed using the BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, 4336774) and 40 ng of the PCR product. The sequencing was performed in a Genetic Analyzer 3130 (Applied Biosystems).

#### Protocol validation on patient material

All slides, obtained after the processing of blood samples (4–20 slides per blood sample) were stained according to the established protocol and reviewed by fluorescence or light microscopy, respectively.

The most suitable approach combines chromogenic and fluorescent staining: dark blue chromogenic substrate NBT/BCIP was used for the detection of CD45, while ER and keratin were stained using Alexa 488 and Cy-3 dyes, respectively; counter staining was performed using DAPI. The staining of keratins (K8/

18/19), CD45, and ER, allowed detection of CTCs in blood and simultaneous determination of the ER status of the detected CTCs. Cells were identified as CTCs if they were positive for keratin and nuclear staining and negative for CD45 staining (K+/CD45-/DAPI+). CTC status was additionally confirmed by light microscopy using the criteria: nearly round or oval shape and high nuclear/cytoplasm ratio. The CTC number variation in blood volume collected from each patient was normalized as number of CTCs detected per one milliliter of analyzed blood.

Eight CTCs from 4 patient samples (2 CTCs per patient) were picked by micromanipulation and underwent WGA as proof of principal for the feasibility of subsequent genomic analysis. Subsequent multiplex PCR of one housekeeping gene (GAPDH) was performed. Detection of expected PCR products confirms that the quality of single cell DNA after the established staining is sufficient for further genetic analysis. The *ESR1* mutation analysis of exons 4, 6, and 8 on single cells was performed.

#### Statistical analyses

Statistical significance between the groups of CTC+ and CTC- patients depending on clinical disease status was calculated by Fisher's exact test. Survival analysis of patients tested for CTCs was done using the log-rank test after dividing the patient cohorts into CTC-positive and CTC-negative groups; HRs and 95% CI were calculated using Cox proportional hazards model. Survival data is estimated from the time point of blood collection. H score of the ER staining was calculated for each CTC-positive patient and normalized in respect to the volume of analyzed blood according to the formula  $\frac{\sum P_i \times i}{V}$ , where  $P_i$  - % of cells of each intensity level,  $i$  - intensity level (from 0 to 3),  $V$  - blood volume in mL. Statistical significance between the groups of patients who received endocrine therapy vs. chemotherapy at the time of blood collection was calculated by Mann-Whitney U-test.

#### Results

##### Spiking experiment and recovery rate

Using blood of healthy volunteers spiked with 500, 100, and 40 cell line cells we demonstrated recovery rate of 79% ± 4% for the

density gradient Ficoll centrifugation as a method for mononuclear cell enrichment.

### CTC detection and evaluation of ER expression

We have established a triple immunostaining protocol for the simultaneous investigation of estrogen receptor (ER), keratins (K) 8/18/19, and CD45 expression on our CTC model system (blood of healthy volunteers spiked with breast cancer cell line cells) with the possibility of further single cell *ESR1* gene mutation analysis. The protocol was used for the detection and characterization of CTCs on blood samples obtained from metastatic breast cancer patients diagnosed for metastases on average 7.2 years (range: 0.5–17.0 years) after initial primary tumor resection. In total, 35 blood samples were screened by non-automated microscopy and CTCs were detected in 16 out of 35 samples (45.7%). ER staining intensity was estimated based on the following grading: no staining (negative); a weak staining (positive); a moderate staining (positive); a strong staining (positive). Samples with a weak, moderate or strong staining will be referred to as being positive for ER expression.

The number of detected CTCs and their ER status are presented in Table 2 (for more detailed results with grades of ER staining see Table S1). All patients had ER-positive primary tumors. ER-positive CTCs were detected in 13/16 patients totally (81.3%). Figure 1 shows a representative example of a single ER-positive CTC (Figure 1A) and a single ER-negative CTC (Figure 1B). Both cells show expression of keratins, but no expression of CD45, indicating that these are tumor cells were of epithelial origin.

Among all 16 CTC positive cases, 8 samples (50.0%) demonstrated homogeneity of ER status: 3 samples (18.7%) with ER-negative CTCs only and 5 cases (31.3%) with ER-positive CTCs only. Eight out of 16 samples (50.0%) displayed both ER-negative and ER-positive CTCs. The distribution of CTC-positive samples according to their ER status is presented in Table 3. Thus, ER-negative CTCs are present in 11/16 cases (68.7%). The average fraction of ER-negative and ER-positive CTCs in samples with mixed population was 36.8% and 63.2%, respectively.

No significant correlation was found between the intensity and/or percentage of ER staining in the primary tumor with the number and ER status of CTCs of the same patient.

### *ESR1* mutation analysis

In subsequent experiments we investigated whether the DNA isolated from CTCs could still be used for genetic downstream analysis after triple staining and micromanipulation. The efficiency of WGA was validated with a single multiplex PCR that amplifies DNA fragments of 100, 200, 300, and 400bp from the housekeeping gene *GAPDH*. All four bands could be produced in the eight CTCs that we investigated (Data S3). Successful amplification of all these four fragments demonstrate that fragments of at least 400bp were specifically produced by the WGA for further genetic analyses [43].

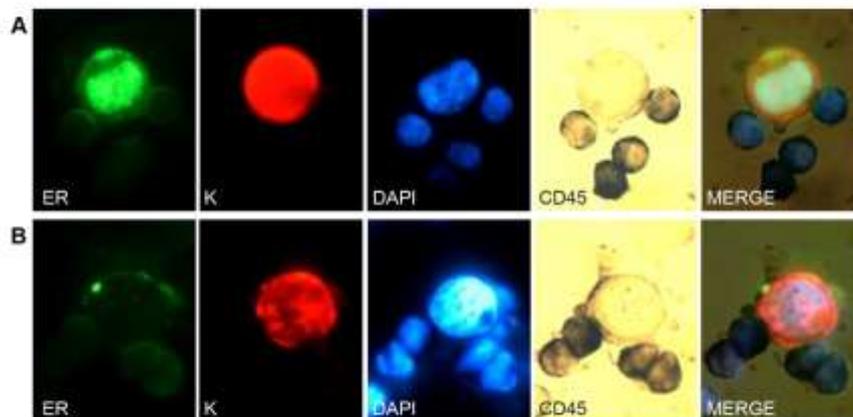
Therefore, we performed mutation analysis of exons 4, 6, and 8 of the *ESR1* gene in 8 individual cells from 4 patients. Figure of Data S4 shows fragments of the high quality sequences that could be produced from all cells in the three exons. However, no mutations were found.

### CTC analysis and clinical outcome

At the time of blood sampling the disease was progressing in 15 patients out of the 16 CTC positive cases and one patient was in remission. In the CTC negative cases, 11 patients were in remission and 3 patients was in progression at the time point of the blood analysis; for 5 patients the clinical status was not evaluated at the time blood was drawn. Number of detected CTCs in respect to clinical status of the patients is presented in Table 4. Thus, the detection of CTCs was significantly associated with clinical progression of the disease ( $p < 0.0001$ , two-sided Fisher's exact test).

Survival analysis starting from the time point of blood analysis until the end of this study (median follow up: 13.1 months, range 1–30 month), demonstrated significant correlation of CTC presence in the blood with shorter disease-free survival ( $p = 0.0381$ ), as depicted by the Kaplan-Meier curves in Figure 2.

Among all 16 CTC positive cases, 14 women received endocrine therapy (87.5%), two (12.5%) did not receive endocrine



**Figure 1. Triple immunostaining of a metastatic breast cancer patient blood sample.** From left to right: estrogen receptor (ER) stained with AlexaFluor 488 (green), keratins 8/18/19 (K) stained with Cy3 (red), DAPI (fluorescent blue) for counter staining, CD45, stained with NBT/BCIP (visible dark blue), and all channels merged. Magnification x100. Row A. Images of ER-positive CTC. A cell with phenotype ER+ (green)/ K+ (red)/CD45- (dark blue)/DAPI+ (fluorescent blue) is considered to be ER-positive CTC. CTC is surrounded with leukocytes (phenotype ER-/K-/CD45+/DAPI+). Row B. Images of ER-negative CTC. A single CTC demonstrating no specific nuclear ER staining. The phenotype is ER-/K+/CD45-/DAPI+. Leukocytes present ER-/K-/CD45+/DAPI+ phenotype.  
doi:10.1371/journal.pone.0075038.g001

**Table 2.** Number of detected CTCs and corresponding ER status.

| Patient ID | Volume of analyzed blood, ml | Normalized number of detected CTCs (per 1 mL of analyzed blood) | Total number of detected CTCs (in total volume of analyzed blood) | ER-negative CTCs | ER-positive* CTCs | H score normalized (per 1 mL of analyzed blood) |
|------------|------------------------------|---|---|------------------|-------------------|---|
| 069        | 7.2                          | 2.78  | 20  | 17               | 3                 | 3.5   |
| 072        | 6.3                          | 0.16  | 1   | 1                | 0                 | 0   |
| 074        | 7.8                          | 0.38  | 3   | 3                | 0                 | 0   |
| 076        | 8.7                          | 2.53  | 22  | 10               | 12                | 12.1  |
| 241        | 10.9                         | 0.73  | 8   | 5                | 3                 | 8.1   |
| 243        | 14.0                         | 270.0   | 270/1 ml  | 98/1 ml          | 172/1 ml          | 109   |
| 250        | 4.5                          | 2.67  | 12  | 3                | 9                 | 33.3  |
| 253        | 10.8                         | 0.19  | 2   | 0                | 2                 | 18.5  |
| 256        | 8.4                          | 0.24  | 2   | 0                | 2                 | 17.9  |
| 259        | 9.6                          | 0.42  | 4   | 3                | 1                 | 2.6   |
| 260        | 7.4                          | 0.27  | 2   | 2                | 0                 | 0   |
| 261        | 8.2                          | 0.12  | 1   | 0                | 1                 | 24.4  |
| 262        | 7.8                          | 1.15  | 9   | 8                | 1                 | 2.8   |
| 280        | 11.5                         | 0.26  | 3   | 0                | 3                 | 23.5  |
| 340        | 7.5                          | 0.13  | 1   | 0                | 1                 | 26.7  |
| 354        | 5.2                          | 0.96  | 5   | 2                | 3                 | 30.8  |

\*ER positive group includes CTCs with weak, moderate, and strong uniform ER staining. For more detailed information see Table S1.  
ER – estrogen receptor; CTC – circulating tumor cell.  
doi:10.1371/journal.pone.0075038.t002

therapy (Table 3). In the blood samples of women with ER-positive primary tumors that received endocrine therapy, ER-negative CTCs were found in 3/14 cases (21.47%), ER-positive CTCs in 4/14 cases (28.6%), and both ER-positive and ER-negative CTCs were detected in 7/14 patients (50.0%). Thus, presence of ER-positive CTCs in patients that received endocrine therapy was detected in 11/14 cases in total (78.6%) and ER-negative CTCs could be found in 10/14 cases (71.4%). Among the three patients in which only ER-negative CTCs were detected, two had progression of disease and therefore received chemotherapy by the time of blood analyses. One patient that developed distant metastases during endocrine therapy was switched to chemotherapy after which remission of the disease was documented.

We analyzed the normalized H score to investigate the clinical relevance of the ER intra-patient heterogeneity (Table 2 and Table S1). The groups of patients who received endocrine therapy vs. chemotherapy at the time of blood collection were compared in respect to the normalized H score for each patient (Mann-Whitney U test), but no significant correlation was found ( $P > 0.05$ ).

No significant correlations were found between ER status of CTCs and the following parameters: progression/remission of the disease, survival, number of detected CTCs, initial therapy, therapy by the time of blood analysis, and time to the metastases diagnosis.

## Discussion

CTCs might serve a “liquid biopsy” to investigate therapeutic targets [25]. One of the techniques often used for determining ER status of CTCs is qRT-PCR [37–39]; however, this approach does not allow for the investigation of intra-patient CTC heterogeneity. Therefore, in the study presented here we have investigated the expression of ER in CTCs in breast cancer patients using immunocytochemistry (ICC). With this approach, we were able to simultaneously detect and characterize CTCs with the additional possibility for downstream genetic analyses of the ER gene using whole genomic amplification (WGA).

In our study we were able to detect CTCs in 16 of 35 patient samples (45.7%), which is within the range of published reports [44]. Because EpCAM might be down regulated in tumor cells that underwent epithelial-mesenchymal transition [45], we have

**Table 3.** The distribution of CTC-positive samples according to their ER status and received therapy.

| CTC status and received therapy | CTC-positive cases       |                            |                            |
|---------------------------------|--------------------------|----------------------------|----------------------------|
|                                 | all kinds of therapy, 16 | women ever received ET, 14 | women never received ET, 2 |
| ER-positive only                | 5                        | 4                          | 1                          |
| ER-positive and ER-negative     | 8                        | 7                          | 1                          |
| ER-negative only                | 3                        | 3                          | 0                          |

ER – estrogen receptor; ET – endocrine therapy; CTC – circulating tumor cell.  
doi:10.1371/journal.pone.0075038.t003

**Table 4.** CTC status in respect to clinical status of the patients.

| Disease status          | Total, 35        |                  |
|-------------------------|------------------|------------------|
|                         | CTC positive, 16 | CTC negative, 19 |
| Progression             | 15               | 3                |
| Remission               | 1                | 11               |
| no data                 | 0                | 5                |
| p (Fisher's exact test) | 0.0001           |                  |

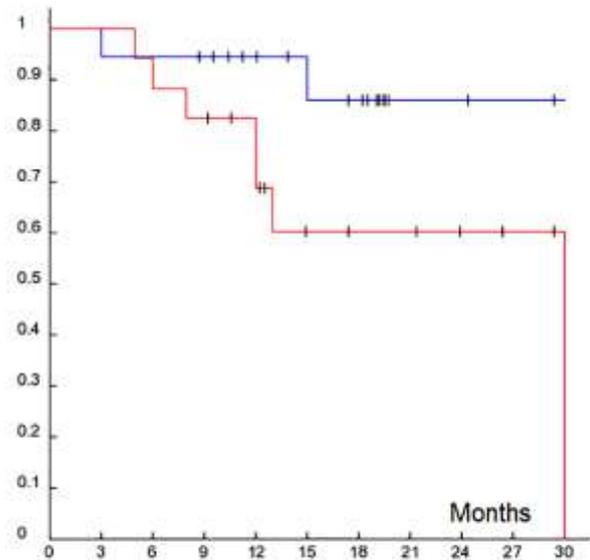
CTC – circulating tumor cell.

doi:10.1371/journal.pone.0075038.t004

used an EpCAM-free detection method in order to capture as many CTCs as possible. Furthermore, we investigated ER expression in the individual keratin-positive CTCs. ER-positive CTCs were detected in 13/16 cases totally (81.3%). Primary tumors of all the patients were positive for ER with the range of ER-positive cells from 10% up to >80% of the cells. No correlation was found between the intensity of ER staining of the primary tumor and the number and/or ER status of CTCs in blood.

Others have found a concordance of ER $\alpha$  status between primary tumor and CTCs in metastatic breast cancer patients in 23% [37], and in 55% [39] of cases using RT-PCR approach (Table 5), which was substantially lower than our results (81.3%), obtained with ICC approach. This might be explained by the low correlation of mRNA and protein expression of ER [46]. To our knowledge, only two studies have been performed in which the authors have stained ER on single CTCs using ICC [47,48]. Limited number of studies, based on ICC for the investigation of CTCs, might be explained by the technical challenges. These challenges had to be taken into consideration: the complications of nuclei permeabilization for antibody delivery, low level of ER presence, difficulties in unequivocal identification of CTCs in case of CD45+/K+ cells presence. A recent study by Bock and colleagues showed a higher percentage of ER-negative CTCs, however, the sample size of CTC positive metastatic breast cancer patients was relatively low ( $n = 5$ ) [48]. In the study of *Nadal et al.*, in contrast to our study, only non-metastatic breast cancer patients before any systemic treatment were enrolled and volume of 30 ml blood per patient was analyzed. ER-negative CTCs were detected in 38.5% of women with ER-positive primary tumors, positive for CTCs [47].

Because of the small number of patients investigated in our study, our follow up analysis is only of exploratory character. Nevertheless, we were able to demonstrate that the detection of CTCs in blood of metastatic breast cancer patients was significantly associated with clinical progression of the disease ( $p < 0.0001$ ). Although the cut-off of at least 5 CTCs per 7.5 ml of blood is considered to be the threshold of high risk of early progression in metastatic breast cancer patients using the CellSearch system [16], recent meta-analysis of *Zhang et al.*, demonstrates prognostic value of the presence of single CTCs [14]. In our study 7 patients had more than 5 CTCs in 7.5 ml of blood; nevertheless, we could demonstrate that the presence vs. absence of CTCs in blood is significantly associated with clinical progression of the disease. Moreover, it has been proposed that level of CTCs at baseline, before a new treatment for the metastatic disease starts, correlate with prognosis and outcome and might be used as independent prognostic marker of progression-free and overall survival [16]. The meta-analysis of *Zhang et al.*



**Figure 2. Kaplan-Meier estimate of survival function.** Kaplan-Meier estimate of survival function of metastatic breast cancer patients separated on CTC-positive (red line) and CTC-negative (blue line) groups. The survival period in month of the corresponding patient. Censored patients are indicated by vertical bars (|). Statistical significance determined by log-rank test. Shorter survival correlates with presence of CTCs in blood ( $P = 0.0332$ , HR: 7.38, CI = 0.84-64.09). doi:10.1371/journal.pone.0075038.g002

demonstrates that prognostic significance of CTCs' presence does not depend on the time point of blood collection [44], which is consistent with our results where blood samples were taken during therapy.

A larger cohort with uniform treatment and longer follow-up will be required to prove the significance and clinical relevance of our findings.

Despite the considered prognostic impact of the presence of CTCs in blood, detection of CTCs in blood does not necessarily reflect the ability of CTCs to survive in the blood stream and to spread to distant organs. The survival and metastatic potential of CTCs need to be investigated.

We hypothesize that distant metastases development in women with ER-positive primary tumors during or after endocrine therapy might be related to the presence of ER-negative CTCs because these cells are most likely to be not affected by endocrine therapy.

Presence of ER-negative CTCs in patients with ER-positive breast cancer might be explained either by heterogeneity of primary tumor, leading to release of both ER-positive and ER-negative cells in circulation or by the switch of ER expression by genomic and/or epigenomic changes (Figure 3). It is proposed, that switching from an ER-positive to ER-negative status might be one of mechanisms to evade hormonal treatment (reviewed in [13,33]).

We observed the presence of ER-negative CTCs in blood of women with ER-positive primary tumors during or after endocrine therapy in 10/14 cases (71.4%); 3/14 had ER-negative CTCs only (21.4%), 7/14 had ER-positive and ER-negative CTCs (50.0%). Interestingly, three of them had disease progression, receiving chemotherapy during the time of blood analyses. Further studies on larger cohorts of patients are required to determine the

**Table 5.** Overview of studies on ER status of CTC in metastatic breast cancer patients.

| Assay / Approach (Ref.)                      | CTC enrichment  | Detection of CTCs                              | Detection of ER | Patient cohort | Distribution of ER in CTCs of the patients with ER-positive primary tumors |                  |                  |                                  |
|--|---|--|-----------------|----------------|--|------------------|------------------|----------------------------------|
|  |   |  |                 |                | CTC detection rate, %  | ER-positive CTCs | ER-negative CTCs | ER-positive and ER-negative CTCs |
| Multiplex RT-PCR Adna Test BreastCancer [39] | anti-EpCAM and MUC1 antibodies coupled with ferrofluidics | Multiplex PCR for mucin-1, ERBB2, actin, EPCAM | RT-PCR          | 42             | 52   | 6/11 (55%)       | 5/11 (45%)       | n.a.*                            |
| Multiplex RT-PCR Adna Test BreastCancer [37] | anti-EpCAM and MUC1 antibodies coupled with ferrofluidics | Multiplex PCR for mucin-1, ERBB2, actin, EPCAM | RT-PCR          | 193            | 45   | 14/62 (23%)      | 48/62 (77%)      | n.a.*                            |
| IF [48]                                      | Ficoll density gradient                                   | IF for K8/18/19                                | IF              | 26             | 38.5   | 0                | 3/5 (60%)        | 2/5 (40%)                        |
| IF (present study)                           | Ficoll density gradient                                   | IF for K8/18/19                                | IF              | 35             | 45.7   | 5/16 (31.3%)     | 3/16 (18.7%)     | 8/16 (50.0%)                     |

CTC – circulating tumor cell; ER – estrogen receptor; IF – immunofluorescence; RT-PCR – real-time PCR.  
 \*RT-PCR approach does not allow to assess intrapatient heterogeneity of ER-status of CTCs.  
 doi:10.1371/journal.pone.0075038.t005

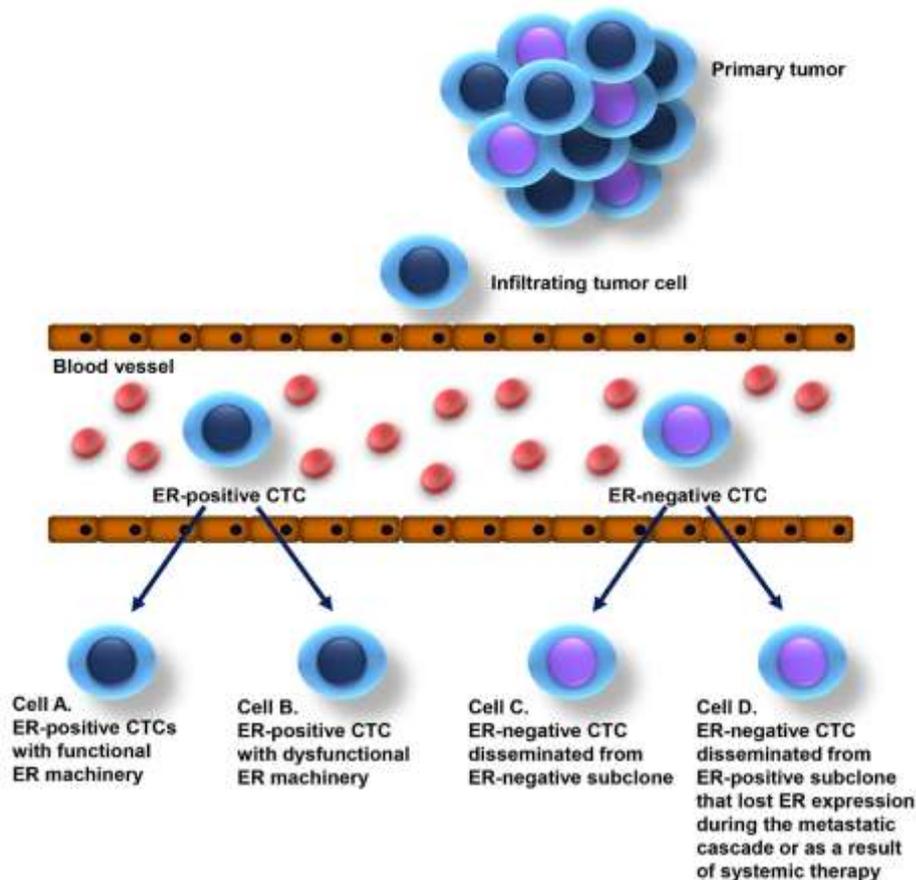
relevance of these findings in the context of specific endocrine therapies.

Another hypothesis, based on our observation of ER-positive CTCs in 11/14 patients after endocrine therapy, is that these cells, which are still present in blood of patients after completion of endocrine therapy, might have a dysfunctional ER pathway and, consequently, resist the hormonal ER blockade. Several mechanisms of ER-positive cells to escape anti-ER therapy have been proposed and include altered crosstalk between ER and signal transduction pathways, growth factor receptors, co-regulatory proteins of ER, and altered expression of specific microRNAs [reviewed in [33,49]]. All these mechanisms potentially lead to the loss of normal ER function and, therefore, inefficacy of anti-ER agents. Several mutations are thought to lead to the inactivation of ER and/or its ligand-independent functioning [14,49]. Therefore, we have performed mutation analysis of the *ESR1* gene in both ER-negative and ER-positive CTCs as certain mutations hamper the protein's function but not its expression [14]. Mutations in *ESR1* occur in approximately 1% of primary breast tumors [50], however were found in 10% of breast cancer metastases but not in the autologous primary tumors [51]. Although further investigation is required, so far we were unable to detect any mutations in the 8 single cells from 4 patients investigated in our study. However, our proof-of-principle study showed that the established immunostaining protocol is compatible with subsequent genomic analysis of CTCs, which allows for the first time a genotype-phenotype correlation at the single cell level with potential implication for future clinical studies using this information to stratify breast cancer patients to endocrine therapies, and to estimate the efficacy of endocrine therapy.

Although the intra-patient CTC heterogeneity is now a fact, a uniform scoring system for its estimation is still missing and estimation of ER expression on CTCs remains subjective. The establishment of such a system would allow for the comparison of ER heterogeneity between patients in respect to therapy as well as monitoring for intra-patient heterogeneity during/after therapy. Different approaches have been reported, nevertheless many of them base on scoring systems suggested for the estimation of IHC staining results of paraffin embedded tissue blocks. *Puumoie et al.* used a scoring system that was originally proposed by *McCarthy*, but modified it for CTCs by using the sum of the positive cell percentage at each intensity level, multiplied by the weighted intensity of staining [52,53].

Another approach was suggested by *Lighthart et al.* The authors used the mean intensity of leukocytes stained as internal threshold for each sample to quantify the intensity of HER2 expression with the use of an automated algorithm [54]. Such approach excludes subjective estimation by the investigator.

For this study, the H score system proposed by *Puumoie et al.* was used with the modification that the obtained H score was normalized to the volume of analyzed blood. This additional normalization allowed for the comparison of samples of different blood volumes. The normalized H scores for CTC-positive patients are presented in Table 2 and Table S1. We compared two groups of patients: those receiving endocrine therapy at the time of blood collection and those receiving chemotherapy, using the Mann-Whitney U-Test. It can be expected that patients who received endocrine therapy by the time of blood collection and still were in progression of the disease would demonstrate higher rates of normalized H score, than those receiving chemotherapy. In our study, the difference between the two groups was not statistically significant. Nevertheless, a larger cohort of patients is needed to study the clinical relevance of this scoring system and its impact on survival.



**Figure 3. Occurrence of ER-positive and ER-negative CTCs in the peripheral blood of patients with breast carcinomas classified as ER-positive.** Circulating tumor cells disseminating from an ER-positive breast tumor can be ER-positive or ER-negative. ER-positive CTCs can have normal functional ER machinery and be sensitive to endocrine therapy (cell A) or have dysfunctional ER machinery and therefore be resistant to endocrine therapy (cell B). ER-negative CTCs might disseminate from ER-negative subclones in tumors classified as ER-positive (diagnostic cut-off value: 1% of ER-stained tumor cells) (cell C) or disseminate from ER-positive subclones that lost ER expression during the metastatic cascade or as a result of systemic therapy (cell D).  
doi:10.1371/journal.pone.0075038.g003

## Conclusion

We established a multiplex immunostaining protocol for the detection and investigation of intra-patient CTC heterogeneity, based on triple staining for keratins, ER and CD45 molecules on blood cytopins, which allows further genetic analyses of single CTCs including mutations in the *ESR1* gene. Our results demonstrate that CTCs in individual metastatic breast cancer patients with ER-positive primary tumors are frequently both ER-positive and ER-negative. ER-negative CTCs may escape ER-targeted endocrine therapy and are, therefore, a potential source of metastatic growth in breast cancer patients with ER-positive primary tumors or metastases. The investigation of CTCs for ER expression and gene status might gain future clinical utility for monitoring and optimization of breast cancer treatment.

## Supporting Information

**Table S1** Patient data.  
(DOCX)

**Table S2** The protocol of testing, staining results and WGA compatibility of different visualization systems.  
(DOCX)

**Data S1** Immunofluorescent staining of estrogen receptor on breast cancer cell line cytopins using Alexa Fluor 488 dye (green) and DAPI nuclei counter staining (blue). Magnification  $\times 100$ . A. MCF7 breast cancer cell line cytopin demonstrating ER staining. B. BT20 breast cancer cell line cytopin demonstrating no ER staining. C. BT174 breast cancer cell line cytopin demonstrating ER staining. D. MDA-MB-231 breast cancer cell line cytopin demonstrating no ER staining.  
(DOCX)

**Data S2** Double immunofluorescent staining of estrogen receptor (ER), stained with AlexaFluor 488 (green) and keratins 8/18/19 (K) stained with Cy3 (red) and DAPI (blue) for nuclei counter staining. A. MCF7 breast cancer cell line cells demonstrating positivity for both ER and keratin staining. B. Cytopin of MCF7 breast cancer cell line cells spiked into blood from healthy volunteer. MCF7 single cell is positive for ER and keratin staining, leukocytes are negative for ER and keratin staining. C. Negative



- alternative hypothesis. *Breast Cancer Res* 7: R1153-R1158. [bcr1342](#) [pii]; doi:10.1186/bcr1342.
33. Garcia-Becerra R, Santos N, Diaz I, Camacho J (2012) Mechanisms of Resistance to Endocrine Therapy in Breast Cancer: Focus on Signaling Pathways, miRNAs and Genetically Based Resistance. *Int J Mol Sci* 14: 198–145. [ijms14010198](#) [pii]; doi:10.3390/ijms14010198.
  34. van de Ven S, Smit VT, Dekker LJ, Noetier JW, Kroep JR (2011) Discordances in ER, PR and HER2 receptors after neoadjuvant chemotherapy in breast cancer. *Cancer Treat Rev* 37: 422–430. [S0305-7372\(10\)00207-0](#) [pii]; doi:10.1016/j.ctrv.2010.11.006.
  35. Thompson AM, Jordan LB, Quinlan P, Anderson L, Skene A, et al. (2010) Prospective comparison of switches in biomarker status between primary and recurrent breast cancer: the Breast Recurrence In Tissues Study (BRITS). *Breast Cancer Res* 12: R92. [bcr2771](#) [pii]; doi:10.1186/bcr2771.
  36. Curitt E, Nerich V, Mansi L, Chaigneau L, Cuis L, et al. (2013) Discordances in Estrogen Receptor Status, Progesterone Receptor Status, and HER2 Status Between Primary Breast Cancer and Metastasis. *Oncologist*. [theoncologist.2012-0250](#) [pii]; doi:10.1634/theoncologist.2012-0250.
  37. Aktas B, Muller V, Trowes M, Zeitz J, Kasimir-Bauer S, et al. (2011) Comparison of estrogen and progesterone receptor status of circulating tumor cells and the primary tumor in metastatic breast cancer patients. *Gynecol Oncol* 122: 356–360. [S0090-8259\(11\)00340-4](#) [pii]; doi:10.1016/j.ygyno.2011.04.039.
  38. Fehm T, Hoffmann O, Aktas B, Becher S, Soltau M, et al. (2009) Detection and characterization of circulating tumor cells in blood of primary breast cancer patients by RT-PCR and comparison to status of bone marrow disseminated cells. *Breast Cancer Res* 11: R59. [bcr2349](#) [pii]; doi:10.1186/bcr2349.
  39. Trowes M, Aktas B, Welt A, Mueller S, Hauch S, et al. (2009) Molecular profiling and predictive value of circulating tumor cells in patients with metastatic breast cancer: an option for monitoring response to breast cancer related therapies. *Breast Cancer Res Treat* 115: 581–590. doi:10.1007/s10549-008-0143-x.
  40. Kohlberger PD, Breininger F, Kaider A, Losch A, Gitsch G, et al. (1999) Modified true-color computer-assisted image analysis versus subjective scoring of estrogen receptor expression in breast cancer: a comparison. *Anticancer Res* 19: 2189–2193.
  41. Remacle W, Stegner HE (1987) [Recommendation for uniform definition of an immunoreactive score (IRS) for immunohistochemical estrogen receptor detection (ER-ICN) in breast cancer tissue]. *Pathologie* 8: 138–140.
  42. Hansmann J, Meyer-Stackling S, Kemning D, Alpers J, Joosse SA, et al. (2011) Quantitative high-resolution genomic analysis of single cancer cells. *PLoS One* 6: e26362. [PONE-D-11-16097](#) [pii]; doi:10.1371/journal.pone.0026362.
  43. van Beers ELJ, Joosse SA, Ligenberg MJ, Eijs R, Hogervorst FB, et al. (2006) A multiplex PCR predictor for aCGH success of FFPE samples. *Br J Cancer* 94: 333–337. [6602889](#) [pii]; doi:10.1038/sj.bjc.6602889.
  44. Zhang L, Riedel S, Wu G, Wang T, Yang K, et al. (2012) Meta-analysis of the prognostic value of circulating tumor cells in breast cancer. *Clin Cancer Res* 18: 5701–5710. [1078-0432.CCR-12-1587](#) [pii]; doi:10.1158/1078-0432.CCR-12-1587.
  45. Joosse SA, Paudel K (2013) Biologic challenges in the detection of circulating tumor cells. *Cancer Res* 73; 8: 11. [0008-5472.CAN-12-3422](#) [pii]; doi:10.1158/0008-5472.CAN-12-3422.
  46. Bonleaux JM, Cheng H, Welsh AW, Haffly BG, Lamm DR, et al. (2012) Quantitative in situ measurement of estrogen receptor mRNA predicts response to tamoxifen. *PLoS One* 7: e36339. [PONE-D-12-02493](#) [pii]; doi:10.1371/journal.pone.0036339.
  47. Nadal RM, Fernandez A, Sanchez-Rovira P, Salido M, Rodriguez M, et al. (2012) Biomarkers Characterization of Circulating Tumour Cells in Breast Cancer Patients. *Breast Cancer Res* 14: R71. [bcr3180](#) [pii]; doi:10.1186/bcr3180.
  48. Bock C, Rack B, Kuhn C, Hoffmann S, Finkenzeller C, et al. (2012) Heterogeneity of ERalpha and ErbB2 Status in Cell Lines and Circulating Tumor Cells of Metastatic Breast Cancer Patients. *Transl Oncol* 5: 175–183.
  49. Riggins RB, Schreengost RS, Guerrero MS, Bouton AL (2007) Pathways to tamoxifen resistance. *Cancer Lett* 236: 1–24. [S0304-3835\(07\)00154-1](#) [pii]; doi:10.1016/j.canlet.2007.03.016.
  50. Roodi N, Bailey LR, Kao WY, Verrier CS, Yee CJ, et al. (1995) Estrogen receptor gene analysis in estrogen receptor-positive and receptor-negative primary breast cancer. *J Natl Cancer Inst* 87: 446–451.
  51. Karnik PS, Kulkarni S, Liu XP, Baddi GI, Bukowski RM (1994) Estrogen receptor mutations in tamoxifen-resistant breast cancer. *Cancer Res* 54: 349–353.
  52. McCarty KS Jr, Szabo E, Flowers JL, Cox EB, Leigh GS, et al. (1988) Use of a monoclonal anti-estrogen receptor antibody in the immunohistochemical evaluation of human tumors. *Cancer Res* 48: 4244a–4248a.
  53. Punnoose FA, Atwal SK, Spoeckel JM, Savage H, Pandita A, et al. (2010) Molecular biomarker analyses using circulating tumor cells. *PLoS One* 5: e12317. doi:10.1371/journal.pone.0012317.
  54. Lighart ST, Bidari FC, Derraine C, Bachelot T, Delaloge S, et al. (2013) Unbiased quantitative assessment of Her-2 expression of circulating tumor cells in patients with metastatic and non-metastatic breast cancer. *Ann Oncol* 24: 1231–1238. [mdc625](#) [pii]; doi:10.1093/annonc/mdc625.

## 5. PUBLICATION 2

Comparative study of whole genome amplification and next generation sequencing performance of single cancer cells

Anna Babayan<sup>1</sup>, Malik Alawi<sup>2,5</sup>, Michael Gormley<sup>3</sup>, Volkmar Müller<sup>4</sup>, Harriet Wikman<sup>1</sup>, Ryan P. McMullin<sup>3</sup>, Denis A. Smirnov<sup>3</sup>, Weimin Li<sup>3</sup>, Deborah S. Ricci<sup>3</sup>, Maria Geffken<sup>6</sup>, Klaus Pantel<sup>1\*</sup>, Simon A. Joosse<sup>1\*</sup>

<sup>1</sup>Department of Tumor Biology, <sup>2</sup>Bioinformatics single cell Core, <sup>4</sup>Department of Gynecology and <sup>6</sup>Department of Transfusion Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

<sup>3</sup>Janssen, Pharmaceutical Companies of Johnson and Johnson, Spring House, PA, USA

<sup>5</sup>Heinrich-Pette-Institute, Leibniz-Institute for Experimental Virology (HPI), Hamburg, Germany

\*shared senior authorship

Submitted: 2016

Published 19.07.2016 in Oncotarget

## **Abstract**

**BACKGROUND:** Single cell genotyping may provide the next step towards individualized medicine in the management of cancer patients. However, whole genome amplification (WGA) is required in order to study the genome of a single cell. Different technologies for WGA are currently available but their effectiveness in combination with next generation sequencing (NGS) and material preservation is currently unknown.

**METHODS:** We analyzed the performance of WGA kits Ampli1, PicoPlex, and REPLI-g on single and pooled tumor cells obtained from EDTA- and CellSave-preserved blood and from formalin-fixed, paraffin-embedded material. Amplified DNA was investigated with exome-Seq with the Illumina HiSeq2000 and ThermoFisher IonProton platforms.

**RESULTS:** In respect to the accuracy of SNP/mutation, indel, and copy number aberrations (CNA) calling, the HiSeq2000 platform outperformed IonProton in all aspects. Furthermore, more accurate SNP/mutation and indel calling was demonstrated using single tumor cells obtained from EDTA-collected blood in respect to CellSave-preserved blood, whereas CNA analysis in our study was not detectably affected by fixation. Although REPLI-g WGA kit yielded the highest DNA amount, DNA quality was not adequate for downstream analysis. Ampli1 WGA kit demonstrates superiority over PicoPlex for SNP and indel analysis in single cells. However, PicoPlex SNP calling performance improves with increasing amount of input DNA whereas CNA analysis does not. Ampli1 performance did not significantly improve with increase of input material. CNA profiles of single cells, amplified with PicoPlex kit and sequenced on both HiSeq2000 and IonProton platforms, resembled unamplified DNA the most.

**CONCLUSION:** Single cell genomic analysis can provide valuable information. Our study shows the feasibility of genomic analysis of single cells isolated from differently preserved material, nevertheless, WGA and NGS approaches have to be chosen carefully depending on the study aims.

## INTRODUCTION

Introduction of single cell analysis led to paradigm shifts in almost all fields of biology and medical sciences as it allows for an accurate representation of the cell-to-cell heterogeneity instead an average measure of an entire cell population [1]. In cancer research, single cell analysis empowers characterization of tumor heterogeneity, and most notably has potential for clinical impact through characterization of circulating tumor cells (CTCs).

CTCs are tumor cells that have separated from primary tumor or current metastases and have infiltrated the systemic blood circulation [2]. Quantification and characterization of CTCs in blood of cancer patients was introduced as a concept of “liquid biopsy”. Enumeration of CTCs as a validated clinical biomarker has been utilized for disease prognosis, diagnosis of minimal residual disease, and monitoring of therapy effectiveness for breast, prostate, and colon cancer [3, 4]. Genomic characterization of CTCs provides insights into genetic heterogeneity of the cancer and metastases and might aid clinical management of cancer patients due to identification of therapy sensitive and resistant clones. Herewith, investigation of single cell genomics may provide the next step towards individualized medicine.

Individual CTCs can be investigated using a combination of whole genome amplification (WGA) and next generation sequencing (NGS) to determine copy number aberrations (CNAs) and gene mutations. However, single cell genomics is associated with certain technical challenges, such as introduction of WGA- and NGS-associated errors. Different technologies for WGA and NGS are currently available but their effectiveness in combination is currently unknown, as well as influence of material preservation on downstream analysis. Suitability of a certain WGA-NGS combination for a particular downstream analysis should be extensively investigated in order to establish a powerful and reliable tool for single cell genomics.

WGA is required for molecular profiling of CTCs since a single cell does not contain enough DNA for direct biomolecular investigation. WGA can be performed by different techniques, such as PCR-based, multiple-displacement amplification (MDA)-based, and a combination of MDA pre-amplification and PCR-amplification. Unlike exponential gain in the first two WGA methods, combined MDA-PCR provides quasi-linear amplification [5-7]. The amplification approach has to be chosen carefully depending on its specific characteristics and the subsequent analysis [8].

An important factor influencing WGA is material preservation, in particular blood preservation. EDTA-preserved blood requires processing as soon as possible [9]. Circulating tumor cells in blood may be preserved in special preservation tubes (CellSave) in order to overcome this requirement. These tubes contain a cell preservative, that stabilizes the sample and maintain cell morphology and cell-surface antigens for up to 96 hours at room temperature, allowing for shipment of the samples. However, fixatives may inhibit DNA amplification, hampering downstream analysis [9, 10]. Most tissue samples are conserved by formalin-fixation, and paraffin-embedding (FFPE), which is difficult to handle in biomolecular analysis due to formalin-induced cross-links [11]. Therefore, it is essential to have WGA methods compatible with these types of preservation.

Downstream analysis of amplified DNA can be performed by massive parallel sequencing using NGS in order to identify SNPs (single nucleotide polymorphisms), indels (insertions-deletions), loss of heterozygosity, structural variations, and CNAs.

Single cell analysis of genomic aberrations by array-CGH is hampered. The necessity of the pre-selected targets' analysis on template, obtained by random and incomplete genome amplification during WGA [12-14] results in high noise and misinterpretation of the results [15]. Moreover, array-CGH provides limited resolution. The highest resolution for whole genome analysis by array-CGH is 56 kb [16]. In contrast, NGS provides the possibility to examine each nucleotide of the entire amplified product with single base resolution.

Existing NGS platforms differ by library preparation and signal detection methods. Illumina's HiSeq machines exploit sequencing-by-synthesis approach [17, 18]. Currently, HiSeq platforms offer the highest throughput per run, although a sequencing run lasts multiple days [18]. Thermofisher's IonProton sequencers utilize semiconductor sequencing technology, allowing to complete a sequencing run within 4 hours, but homopolymer stretches might be called incorrectly [17].

In this study, we evaluated different protocols, including different methods of preservation, WGA and sequencing to identify an optimal process for single cell sequencing. We compared our findings against unamplified DNA from bulk cell pellets to quantitatively define the impact of different protocols on single cell sequencing. In order to determine the impact of WGA method, we evaluated three different commercially available WGA kits and measured DNA quality and yield. In order to investigate the performance and compatibility of NGS platforms with whole exome

sequencing of WGA single cell DNA, we compared the detection of genomic variants (SNPs, indels and CNAs) from single SK-BR-3 cells spiked and re-captured from EDTA-preserved blood. In order to investigate the influence of material fixatives, we evaluated detection of genomic variants from single SK-BR-3 cells spiked and re-captured from EDTA-preserved vs. CellSave preserved blood. Next, we evaluated the limit of detection and consistency of genomic variant detection with increasing amounts of starting material (i.e. increasing numbers of pooled cells). Finally, we demonstrate proof of principle by evaluating genomic variants detected from CTCs collected from breast cancer patients. Our findings indicate the technical and biological variability in genomic variant detection from single cell sequencing and suggest optimized protocols dependent on starting material and objective (i.e. SNP calling vs. CNA calling).

## MATERIALS AND METHODS

### **Experimental design**

First, we investigated performance of 3 WGA kits, representing 3 WGA methods, in 4 groups of source material, differing by origin and preservation method. The 4 sources of material included: A) individual SK-BR-3 cells obtained from EDTA-preserved blood; B) individual SK-BR-3 cells obtained from CellSave-preserved blood; C) single SK-BR-3 cells picked from FFPE SK-BR-3 cells; and D) individual CTCs obtained from EDTA-preserved blood from a breast cancer patient.

After DNA yield and quality per WGA kit were estimated, DNA of single cells from each WGA group was used for whole exome NGS on 2 platforms. Briefly, 3 SK-BR-3 cells, obtained from EDTA-preserved blood and amplified with Ampli1, PicoPlex, and REPLI-g kit, were analyzed with both HiSeq200 and IonProton platforms.

Based on results obtained from initial pilot experiments, the IonProton platform and Repli-G WGA kit were excluded from further experiments. The second round of experiments included WGA of single and pooled cells in duplicates and NGS of obtained DNA in order to investigate the performance and the limit of detection with increasing amounts of material. Duplicates of 1, 3, 5, and 10 pooled SK-BR-3 cells obtained from CellSave-preserved blood and amplified with Ampli1 and PicoPlex kits were sequenced on HiSeq2000.

Subsequently, a proof of principle experiment was performed on 2 individual CTCs obtained from EDTA-collected blood of a breast cancer patient. The cells were individually amplified with PicoPlex WGA kit and sequenced on HiSeq2000

(Supplementary Figure 1). In total, 120 single cells and 72 pooled cells were processed.

### **Cell culture**

The breast cancer cell line SK-BR-3 was acquired from ATCC and cultivated under prescribed conditions. The cells were harvested using trypsin/EDTA (R001100; Gibco), washed and resuspended in PBS (14190-094; Gibco) for further experiments. Genomic DNA was extracted using the Blood&Cell Culture DNA Mini Kit (13323, Qiagen). The same cell line was previously formalin-fixed, paraffin-embedded, and stored for over 3 years to simulate archival material.

### **Blood sampling**

Blood from healthy individuals and metastatic breast cancer patients was obtained from the Department of Transfusion Medicine and Department of Gynecology at the University Medical Center Hamburg-Eppendorf, respectively. All study participants gave written informed consent. The examination of blood from breast cancer patients was approved by the local ethics review board Aerztekammer Hamburg (OB/V/03). Breast cancer patients' blood was sampled in EDTA collection tubes (01.1605.001, Sarstedt). Blood from healthy donors was collected either in EDTA or CellSave tubes (7900005, Janssen Diagnostics) and spiked with SK-BR-3 cells to simulate CTCs.

### **Sample preparation**

Blood samples collected in EDTA tubes were processed within 2 hours. Blood samples collected in CellSave tubes were stored for 24-30 hours at room temperature before being processed. Mononuclear cells from both cancer patients' and healthy donors' blood spiked with SK-BR-3 cells were enriched by Ficoll density gradient centrifugation as previously described [19], fixed with 0.5% paraformaldehyde for 10min, and stained for keratins as described elsewhere [20].

Single cells were picked by micromanipulation (micro injector CellTramVario and micromanipulator TransferManNKII, Eppendorf Instruments, Hamburg, Germany). Each individual cell was transferred in 1µl of PBS into the cap of a 200µl PCR tube and stored at -80°C overnight.

FFPE SK-BR-3 material was cut in 5µm thin sections and preprocessed as described before [21, 22]. Cross-links were removed by incubation of the slides in 1M NaSCN at 56°C overnight. Subsequently, the slides were washed 3x3min with TBS, stained with hematoxylin for 30s, rinsed with water, single cells were picked by micromanipulation.

### **Whole genome amplification**

WGA was performed according to the manufacturers' recommendations using 3 different kits: PCR-based Ampli1 (WG-001-050-R02, Silicon Biosystems), combined MDA-PCR PicoPlex (E2620L, New England Biolabs,), and MDA-based REPLI-g (150343, Qiagen) WGA kits. The WGA products after Ampli1 and PicoPlex underwent cleanup with NucleoSpin Gel and PCR Clean-up kit (740609, Macherey-Nagel). REPLI-g WGA products were cleaned according to the QIAGEN recommendations with ethanol for the FFPE samples and spin columns (51304, Qiagen) for the blood samples.

DNA concentration was measured with a Nanodrop1000 (Peqlab, Erlangen, Germany). The quality control of the WGA products was assessed by a multiplex PCR of the *GAPDH* gene as described elsewhere [21] with minor adaptations (Supplementary Material 1). Samples were considered of sufficient quality for further analyses if at least one of 200-400bp bands was detectable.

### **Next generation sequencing**

Amplified DNA was investigated with whole exome sequencing on HiSeq2000 and IonProton platforms, unamplified DNA of the SK-BR-3 cells was sequenced with HiSeq2000. Sequence data are available at <http://www.ebi.ac.uk/ena/data/view/PRJEB11307>.

### **Data analysis**

Raw data from the Ampli1- and PicoPlex-amplified samples underwent adapter clipping. PCR adapters of the Ampli1 kit are ligated to DNA sticky ends after MseI restriction of T<sup>A</sup>TAA sites [23], therefore adapters can be identified as oligonucleotide sequences framing TAA...(N)...T fragments. For PicoPlex-amplified samples we trimmed the first/last 14 bases as suggested by the manufacturer. Random hexamer

primers of REPLI-g are complementary to the DNA and therefore did not need to be trimmed.

Further data analysis was done according to the GATK Best Practices recommendations [24, 25], detailed available as Supplementary Material 2. SNP/mutation and indel discovery was limited to protein coding exons only (downloaded from the CCDS Project database [26]).

Sensitivity, specificity, positive and negative predictive values of the variant calling analysis were evaluated based on the schema presented on Supplementary Figure 2. Calls from single cells (analyzed samples) were compared to calls made from unamplified DNA from bulk cell pellets (reference). Analyses were limited to SNP positions and alleles as defined in the dbSNP (Version 138.hg19) to minimize discrepancies from random error between samples. Truepositives (TP) are defined as known SNPs found in both the reference and analyzed samples. Falsepositives (FP) are known SNPs identified in analyzed samples but not present in reference. Conversely, known SNPs identified in the reference sample but not in the analyzed sample are false negatives (FN). Based on this definition, sensitivity (S), specificity (Sp), positive predictive value (PPV) and negative predictive value (NPV) were calculated as follows:  $S = TP/(TP+FN)$ ,  $Sp = TN/(TN+FP)$ ,  $PPV = TP/(TP+FP)$ ,  $NPV = TN/(TN+FN)$ . Indel calling statistics were calculated similarly.

Venn diagrams were created with the use of BioVenn web application [27].

Allelic dropout (ADO) rate was calculated as follows: heterozygous SNPs in sample, present in reference, divided by their sum with homozygous SNPs in sample, present in reference as heterozygous SNPs.

CNAs were evaluated using Control-FREEC [28] with a window size of 30kb, visualized and further analyzed using custom scripts (MATLAB R2015a, The MathWorks Inc.). Correlation among CNA profiles was calculated using Spearman correlation test.

## RESULTS

### **Whole genome amplification of single cells**

Three WGA kits (i.e. Ampli1, PicoPlex, REPLI-g) were used to amplify single cell samples of 4 groups: A) 10 individual SK-BR-3 cells spiked and picked from EDTA-preserved blood; B) 10 individual SK-BR-3 cells spiked and picked from CellSave-preserved blood; C) 10 single SK-BR-3 cells picked from FFPE SK-BR-3 cells; and D)

10 individual CTCs picked from EDTA-collected blood of breast cancer patients. In total, 120 single cells were individually processed by WGA. DNA yield and success rate, as measured by multi-plex PCR of *GAPDH*, of the tested WGA kits are presented in Table 1 and on Supplementary Figure 3.

WGA with the Ampli1 kit demonstrated an average DNA yield of 7.07 µg, 5.86 µg, 6.74 µg and 4.69 µg for the 4 different 10-sample sets respectively with the average DNA yield 6.09 µg (Table 1a). The *GAPDH* multiplex-PCR demonstrated a 100% success rate for the experiment with EDTA tubes, CellSave tubes, and FFPE experiments, whereas the amplification of the patients' CTCs demonstrated a success of 70% for CTCs (Table 1b). The average DNA yield for PicoPlex kit was 2.86 µg, 3.39 µg, 4.71 µg and 4.01 µg for the 4 different 10-sample sets respectively and 3.74 µg on average for all 40 samples. Quality control PCR demonstrated 100% success rate in all groups except single SK-BR-3 cells picked from EDTA blood (80% success rate). The REPLI-g kit demonstrated the highest DNA yield: 15.39 µg, 11.37 µg, 77.97 µg and 31.41 µg in the same 4 experimental groups respectively. The average DNA output was 34.04 µg for all 40 samples. Quality control PCR demonstrated 70% success rate in cases of single SK-BR-3 picked from EDTA and CellSave tubes and 30% in cases of FFPE SK-BR-3 cells as well as patient CTCs. Among all tested WGA kits REPLI-g demonstrated the highest DNA yield along all sample group, however with the lowest success rate (50% average). Ampli1 and PicoPlex kits demonstrated comparable success rates (on average 93 and 95%, respectively) with DNA yield prevalence of Ampli1 over PicoPlex-processed samples in all compared groups (on average 6.09 and 3.74µg, respectively).

### **SNP/mutation, indel, and CNA analyses of SK-BR-3 cells, obtained from EDTA-preserved blood**

Genomic variants detected from single cells recovered from EDTA-preserved blood were analyzed to compare sequencing platforms and WGA methods. Variants detected in single cell analyses were compared to variants detected in bulk cell pellets without WGA as a gold standard. We report sequencing quality statistics (e.g. read depth), the total number of SNPs and indels detected, including both previously reported SNPs and indels and novel variants, the allelic dropout rate and the sensitivity and positive predictive value of detection compared against unamplified DNA as metrics to compare different protocols. Sequencing with HiSeq2000 platform produced

more reads and provided higher depth and breadth of target base coverage, higher mapping rates, and lower duplicate rates compared to IonProton. Comparing the applied WGA procedures, the highest numbers of clean reads, mapping and duplicate rates were observed for REPLI-g WGA kit. The complete characteristics of NGS data are presented in Supplementary Table 1. The number of total and known SNPs identified with HiSeq2000 platform was higher than for IonProton regardless of the WGA method used (Figure 1A). Sequencing with the HiSeq2000 platform resulted in 7125, 4680, 173 known SNPs detected with Ampli1, PicoPlex, and REPLI-g kits, respectively, and concordant with known SNPs detected in bulk unamplified NA. Sequencing with the IonProton platform resulted in the detection of 1525, 1073, and 30 concordant known SNPs with respective WGA kits. Sensitivity, the probability of detecting a known SNP found in the reference sample in the single cell samples, was also higher in samples sequenced with HiSeq2000 with 41.3, 27.1% and 1.0% for Ampli1, PicoPlex, and REPLI-g WGA experiments, respectively (Table 2). Novel single nucleotide variants (SNVs) were identified in single cells, as well as in genomic DNA, which might be sequencing or amplification errors. Higher numbers of novel SNVs were observed for HiSeq2000 over IonProton sequenced samples (265 vs 50, 4711 vs 538, and 203 vs 33 for Ampli1, PicoPlex, and REPLI-g WGA kits and HiSeq2000 vs IonProton NGS, respectively). The highest number of novel SNPs was observed for the PicoPlex-amplified and HiSeq2000-sequenced cell (4711 SNPs). Among the samples sequenced with the same NGS platform, more known indels were identified in Ampli1-amplified samples (176 vs 23, 82 vs 14, and 3 vs 1 for Ampli1, PicoPlex, and REPLI-g WGA kits compared for HiSeq2000 vs IonProton, respectively). The fraction of known indels was the highest for Ampli1-HiSeq2000 analysis (15.3%) (Table 2). CNA profiles from single cells were compared with the CNA profile of genomic SK-BR-3 DNA using Spearman correlation (Figure 2A-G). Correlation between whole genome amplified single cells and genomic DNA did not depend on NGS platform, but was dependent on WGA kit. Ampli1, PicoPlex, and REPLI-g amplified cells demonstrated median ( $r < 0.7$ ), strong ( $r > 0.8$ ) and weak ( $r < 0.3$ ) correlation with genomic DNA, respectively (Table 2).

ADO rates demonstrated dependence on both WGA and sequencing platform (Table 2). ADO rates were lower in HiSeq2000-sequenced samples in comparison to IonProton with outperformance of the Ampli1 kit within the same NGS platform (9, 24, and 100% for cells, amplified with Ampli1, PicoPlex and REPLI-g WGA kits and

sequenced on Hiseq200 vs 20, 42, and 100% in IonProton group, respectively). Based on the obtained results of the first NGS experiments, we excluded REPLI-g WGA kit (MDA-based technique) and IonProton platform from further analyses.

### **SNP/mutation, indel, and CNA analyses of single and pooled SK-BR-3 cells, obtained from CellSave-preserved blood**

To investigate the detection limit with increasing amount of starting material for WGA, as well as the influence of CellSave preservative on WGA and NGS performance we analyzed duplicates of 1, 3, 5, and 10 pooled SK-BR-3 cells amplified with Ampli1 and PicoPlex WGA kits and sequenced on Illumina's Hiseq2000. The whole obtained data is presented in Supplementary Table 1.

Comparison between single cells obtained from EDTA- and CellSave-collected blood revealed lower numbers of the total and known identified SNPs and indels, and higher number of novel SNPs and indels in cells from CellSave-preserved blood. Sensitivity of SNP and indel calling was lower for single cells from CellSave tubes in comparison to single cells obtained from EDTA-preserved blood (Table 3). The overlap in known SNPs detected from single cells in EDTA and CellSave preserved blood was similar to the overlap detected from technical replicates of single cells in CellSave preserved blood (Figure 1B), indicating that technical bias from other sources is greater than variation from the preservation method. As described above, comparison of findings obtained by different WGA kits demonstrates superiority of Ampli1 WGA over PicoPlex for SNP and indel analysis in single cells, as indicated by the higher sensitivity of Ampli1 WGA (Table 3).

Analyses of pooled cells demonstrated that the numbers of the identified total and known SNPs/mutations increased with increasing number of pooled cells, however statistically significant between different groups of pooled cells for PicoPlex only. Moreover, the rate of change of detection of total and known SNPs with increasing number of cells was found to be different with Ampli1 and PicoPlex kits. Ampli1 kits appear to have more variability in performance as indicated by the variance in the number of total SNPs detected (Figure 3A). Variance is smaller in the percentage of known SNPs detected (Figure 3A). In addition, the increase in the percentage of known SNPs detected with increasing numbers of pooled cells is greater for the PicoPlex kit (Figure 3A, 3B).

Sensitivity of SNP and indel analyses increased with increasing number of pooled cells. The effect was statistically significant for PicoPlex, but not Ampli1 kit as indicated by significant correlation. However, the differences in kits' performance were not significantly different by these metrics (Figure 3C, 3D).

Samples analyzed with either kit showed similar ADO rates (3-79 and 2-74%, respectively) (Supplementary Table 1). With each kit, ADO rates decreased with increasing numbers of pooled cells, indicating that the high ADO rate with single cells is largely attributed to WGA. This effect was significantly different for PicoPlex kit only (Figure 3E).

Correlation between CNA profiles of genomic DNA and analyzed samples increased along with the number of pooled cells for both WGA kits, however, this effect was not statistically significant. There was no significant difference in the rate of change of performance between the two kits (Figure 3F).

The obtained results suggest that the Ampli1 WGA kit is superior to the PicoPlex kit for SNP and indel analysis in single cells (Table 3). Notably, detection of SNPs by PicoPlex significantly improves with the number of pooled cells (i.e. increasing amount of input DNA). Ampli1 performance did not significantly improve with increase of input material in any case. This suggests a greater effect of WGA for MDA-PCR amplification in comparison to PCR-based amplification.

### **Genomic characterization of patient tumor cells**

As proof of principle, two CTCs from a metastatic breast cancer patient were isolated from 10ml of blood obtained in an EDTA tube. The PicoPlex WGA kit was used to amplify the genomes of the individual cells, followed by exome sequencing using HiSeq2000. CNA analysis demonstrated two genetically different profiles (Figure 2H, 2I), suggesting cancer genetic heterogeneity of this patient's disease. Both CTCs carry gain of chromosome 1q, which has been identified previously as an universal genomic feature of breast cancer [29]. Additionally, CTC-1 demonstrates copy number variation typical for luminal breast cancer including chromosome 16p gain and chromosome 16q loss. In contrast, CTC-2 is strongly characterized by chromosome 9p loss. SNP calling analysis revealed 1135 SNPs and 15 indels common in both cells (Figure 1C). Mutation analysis revealed 5 missense mutations annotated in COSMIC database [30]. Mutations in genes *CHEK2*, *PRAME*, and *KIT* were present in both CTCs, mutation in gene *FGFR2* was detected in CTC-1 only and in gene *TP53* – in CTC-2 only (Table 4).

## DISCUSSION

In the study presented here, the performance of single cell WGA and subsequent whole exome sequencing were investigated on 2 different NGS platforms. Illumina's HiSeq platforms are widely used in human genome research due to their accuracy. Sequencing with ThermoFisher's IonProton can be faster and more cost-effective per run, however, sequencing with IonProton may result in substantial decrease of effective coverage depth due to the high abundance of PCR and optical duplicates. Emulsion PCR, utilized for library preparation in IonProton technology, is thought to be the main source of PCR duplicates [31]. Moreover, the introduction of indels is a well-documented disadvantage of the semiconductor sequencing utilized in IonProton [17]. Nevertheless, our study shows that CNA analysis was not affected by the described disadvantages of semiconductor sequencing and demonstrated comparable results for samples sequenced on both NGS platforms.

Important applications of NGS such as SNP/mutation, indel, and CNA calling seem to be especially hampered in single cell analysis due to relatively high variance in amplification efficiency across the genome as a result of WGA [6, 12, 13]. Allelic dropout (ADO), defined as the complete absence of one allele of heterozygous loci, is one of the major concerns associated with WGA, leading to false interpretation of SNP/mutation and indel calling results. In our study, Ampli1 WGA kit demonstrated lower ADO rates on the single cell level and consequently more accurate SNP/mutation and indel calling independent of sequencing platform, blood preservative, and number of pooled cells. These data suggest that Ampli1 (PCR-based WGA) outperforms PicoPlex (MDA-PCR combining WGA technique) and REPLI-g (MDA-based) WGA kits for SNP/mutations and indel calling. However, adaptor-ligation PCR, utilized in some PCR-based WGA kits (e.g., Ampli1), has certain limitations. Site-specific digestion of template DNA prior to PCR by the MseI enzyme [32] results in a wide distribution of fragment lengths. In silico analysis (data not shown) demonstrates that only 38% of  $19 \times 10^6$  fragments produced by MseI restriction of the human genome have length 100-500bp and therefore sufficient for exome-capturing and size-selection for library preparation. In order to optimize single cell sequencing, revision of the current exome capturing regions is required.

Commercially available exome enrichment kits have not been optimized for WGA products. The usage of fragmented WGA DNA as template might drastically reduce capturing efficacy. Moreover, a significant fraction of template DNA can be

nonspecifically enriched outside target regions, varying from kit to kit [33-35], causing identification of thousands of high quality SNPs outside the target regions [33]. A limitation of this study is that only one exome capturing kit has been tested and thus, it cannot be ruled out that other capturing kits may have different results. Exome capturing in which smaller regions are targeted might outperform capturing of larger genomic regions. Another limitation of this study is that we used SK-BR-3 bulk DNA, sequenced on HiSeq2000 as reference for SNP/mutation, indel and CNA analysis for SK-BR-3 cells, sequenced on both HiSeq2000 and IonProton platforms. IonProton sequenced bulk SK-BR-3 DNA used as reference might improve results of IonProton-sequenced samples.

Although REPLI-g amplified samples (MDA-based technique) demonstrated the highest DNA yield from a single cell, the quality of the obtained DNA was remarkably low and insufficient for appropriate SNP/mutation, indel, and CNA analyses. Based on our experience and observations of de Bourcy et al. [8] and Bergen et al. [36], we conclude that input of at least 10ng of genomic DNA and tailoring of the MDA reaction to obtain just enough DNA for further analysis is a key to optimal MDA performance. Further biases in MDA-based WGA can distort CNA analysis and have been described elsewhere, these include uneven representation and non-specific amplification of the genome, a large variability in amplification bias among the products, chimera formation, and dislocated sequences [8, 37-39].

Single cells from EDTA-collected blood demonstrated higher sensitivity for SNP/mutation and indel analyses, than single cells from CellSave-preserved blood. Since EDTA-collected blood requires timely processing after collection [9], CellSave blood preservation could be of great value in e.g., multicenter studies. In this study, we examined the consistency of SNP/mutation and indel calling performance in 1, 3, 5, and 10 pooled cells in comparison to unamplified genomic DNA and the influence of WGA technique on the results. SNP/mutation and indel analyses of single and pooled cells revealed high variability in results for Ampli1 and PicoPlex WGA kits on single cell level, decreasing with the number of pooled cells.

The concordance of the identified SNPs/mutations in 1, 3, 5, and 10 cells from CellSave-preserved blood with the reference was invariant to the WGA technique used and improved with increasing number of pooled cells. For PicoPlex amplified DNA, the sensitivity  $\left(\frac{\text{true-positives}}{\text{true-positives} + \text{false-negatives}}\right)$  of the SNP/mutation analysis increases with increasing number of pooled cells, in association with a decrease in ADO rates. A

similar trend was detected with the Ampli1 kit amplified DNA, although the effects were not significant. The effects may have been obscured by the relatively high variance observed with Ampli1 amplification. Similarly, CNAs detected from single or pooled cells demonstrates a trend for increasing correlation with calls made from unamplified DNA. These effects are not significant which may be due to the relatively high variance in correlation with low number of starting cells (Figure 3). Nevertheless, CNA profiles of even a single cell from EDTA-collected blood, amplified with PicoPlex WGA kit and sequenced on both HiSeq2000 and IonProton, demonstrated strong correlation ( $r \geq 0.8$ ) with unamplified DNA in contrast to a moderate correlation observed for Ampli1 kit for the same experiment ( $r < 0.7$ ) (Table 2). Moreover, as few as 3 pooled cells from CellSave-preserved blood resembled CNA pattern of unamplified DNA with strong correlation, whereas Ampli1-amplified samples reached the same correlation level with 5 pooled cells.

A recent study from our lab has demonstrated genetic heterogeneity within a cancer cell line upon sequencing single cells [40]. Therefore, it cannot be ruled out that the low concordance of SNP/mutation calling between single cells might also be the effect of heterogeneity in addition to WGA artifacts. However, the strong correlations of CNA of the SK-BR-3 cell-line between different lineages published in the past [11] suggests that its overall genome is relatively stable. Further research entailing deep sequencing of unamplified genomic DNA will reveal the genetic heterogeneity of this cell line. It has been noted that WGA strongly affects CNA analysis due to imbalanced amplification of alleles [5, 13]. Moreover, non-linear amplification is random and is not reproducible for the same DNA template [14]. Although CNA analysis does not require exome capturing and is possible on whole genome shallow sequenced data, we performed CNA analysis on whole exome sequencing data and demonstrated that the quality of the obtained DNA by both Ampli1 and PicoPlex kits was adequate for qualitative assessment of CNA patterns. Deeper exome sequencing may compensate imbalanced allele amplification, crucial for CNA analysis of shallow sequenced whole genome data.

Sequencing CTCs from cancer patients has been suggested as a “liquid biopsy” that could be used to study tumor heterogeneity and find therapy associated markers [41]. In our study, we identified 3 cancer-associated mutations, 1135 SNPs, and 15 indels common in two CTCs from a single breast cancer patient, however their CNA profiles were not similar, reflecting intra-patient heterogeneity. Given the findings

presented from our benchmarking analyses, it is difficult to separate true biological variants from variation introduced by WGA or sequencing artifacts. However, identification of non-overlapping mutations in *FGFR2* and *TP53* genes might indicate clonal evolution of the tumor. Further single cell genomic research and improved WGA methods may enable us to investigate cancer evolution during tumor development and under therapy pressure leading to treatment resistance using CTC sequencing.

## CONCLUSION

We comprehensively tested the effectiveness of WGA of single cells for exome sequencing by NGS. As an aspect of testing, we evaluated 3 WGA techniques, 2 NGS platforms, and the influence of material fixation for long term preservation. Although REPLI-g WGA kit yielded the highest DNA amount, DNA quality was not adequate for SNP/mutation, indel, and CNA analysis.

Ampli1 WGA kit combined with Illumina's HiSeq2000 platform demonstrated the best concordance with unamplified DNA for SNP/mutation and indel calling, both for EDTA- and CellSave-preserved cells with ADO rates 9-79%, mostly dependent on the amount of starting material. However, PicoPlex performance significantly improves with the number of pooled cells (increasing amount of input DNA), whereas Ampli1 performance did not significantly improve with increase of input material in any case.

The CNA profiles produced with PicoPlex kit on both HiSeq2000 and IonProton, independent of blood preservative, resembled unamplified DNA the most. PicoPlex performance of CNA analysis is not affected by input amount.

Our study shows the feasibility of genomic analysis of single cells isolated from differently preserved material, enabling advanced diagnostics such as on CTCs during cancer treatment for companion diagnostics.

## ACKNOWLEDGEMENTS

The authors receive support from CANCER-ID, an Innovative Medicines Initiative Joint Undertaking under grant agreement n° 115749, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution. This work was further supported by a contract grant from Janssen R&D and the European Research Council Advanced Investigator grant 269081 DISSECT.

## REFERENCES

1. Wang, D. and S. Bodovitz, *Single cell analysis: the new frontier in 'omics'*. Trends Biotechnol, 2010. 28(6): p. 281-90.
2. Joosse, S.A., T.M. Gorges, and K. Pantel, *Biology, detection, and clinical implications of circulating tumor cells*. EMBO Mol Med, 2015. 7(1): p. 1-11.
3. Cristofanilli, M., et al., *Circulating tumor cells, disease progression, and survival in metastatic breast cancer*. N.Engl.J.Med., 2004. 351(8): p. 781-791.
4. Alix-Panabieres, C. and K. Pantel, *Challenges in circulating tumour cell research*. Nat Rev Cancer, 2014. 14(9): p. 623-31.
5. Barker, D.L., et al., *Two methods of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage panel*. Genome Res, 2004. 14(5): p. 901-7.
6. Macaulay, I.C. and T. Voet, *Single cell genomics: advances and future perspectives*. PLoS Genet, 2014. 10(1): p. e1004126.
7. Zong, C., et al., *Genome-wide detection of single-nucleotide and copy-number variations of a single human cell*. Science, 2012. 338(6114): p. 1622-6.
8. de Bourcy, C.F., et al., *A quantitative comparison of single-cell whole genome amplification methods*. PLoS One, 2014. 9(8): p. e105585.
9. Qin, J., et al., *Stabilization of circulating tumor cells in blood using a collection device with a preservative reagent*. Cancer Cell Int, 2014. 14(1): p. 23.
10. Riethdorf, S., et al., *Detection of circulating tumor cells in peripheral blood of patients with metastatic breast cancer: a validation study of the CellSearch system*. Clin Cancer Res, 2007. 13(3): p. 920-8.
11. Joosse, S.A., E.H. van Beers, and P.M. Nederlof, *Automated array-CGH optimized for archival formalin-fixed, paraffin-embedded tumor material*. BMC Cancer, 2007. 7: p. 43.
12. Pinard, R., et al., *Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing*. BMC Genomics, 2006. 7: p. 216.
13. Pugh, T.J., et al., *Impact of whole genome amplification on analysis of copy number variants*. Nucleic Acids Res, 2008. 36(13): p. e80.
14. Talseth-Palmer, B.A., et al., *Whole genome amplification and its impact on CGH array profiles*. BMC Res Notes, 2008. 1: p. 56.

15. Sermon, K. and S.p. Viville, *Textbook of human reproductive genetics*. ix, 206 pages.
16. Mohlendick, B., et al., *A robust method to analyze copy number alterations of less than 100 kb in single cells using oligonucleotide array CGH*. PLoS One, 2013. 8(6): p. e67031.
17. Buermans, H.P. and J.T. den Dunnen, *Next generation sequencing technology: Advances and applications*. Biochim Biophys Acta, 2014.
18. van Dijk, E.L., et al., *Ten years of next-generation sequencing technology*. Trends Genet, 2014. 30(9): p. 418-426.
19. Babayan, A., et al., *Heterogeneity of estrogen receptor expression in circulating tumor cells from metastatic breast cancer patients*. PLoS One, 2013. 8(9): p. e75038.
20. Joosse, S.A., et al., *Changes in keratin expression during metastatic progression of breast cancer: impact on the detection of circulating tumor cells*. Clin Cancer Res, 2012. 18(4): p. 993-1003.
21. van Beers, E.H., et al., *A multiplex PCR predictor for aCGH success of FFPE samples*. Br J Cancer, 2006. 94(2): p. 333-7.
22. Joosse, S.A., et al., *Prediction of BRCA2-association in hereditary breast carcinomas using array-CGH*. Breast Cancer Res Treat, 2012. 132(2): p. 379-89.
23. Klein, C.A., et al., *Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells*. Proc Natl Acad Sci U S A, 1999. 96(8): p. 4494-9.
24. DePristo, M.A., et al., *A framework for variation discovery and genotyping using next-generation DNA sequencing data*. Nat Genet, 2011. 43(5): p. 491-8.
25. Van der Auwera, G.A., et al., *From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline*. Curr Protoc Bioinformatics, 2013. 11(1110): p. 11 10 1-11 10 33.
26. Farrell, C.M., et al., *Current status and new features of the Consensus Coding Sequence database*. Nucleic Acids Res, 2014. 42(Database issue): p. D865-72.
27. Hulsen, T., J. de Vlieg, and W. Alkema, *BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams*. BMC Genomics, 2008. 9: p. 488.

28. Boeva, V., et al., *Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data*. *Bioinformatics*, 2012. 28(3): p. 423-5.
29. Melchor, L. and J. Benitez, *An integrative hypothesis about the origin and development of sporadic and familial breast cancer subtypes*. *Carcinogenesis*, 2008. 29(8): p. 1475-82.
30. Forbes, S.A., et al., *COSMIC: exploring the world's knowledge of somatic mutations in human cancer*. *Nucleic Acids Res*, 2015. 43(Database issue): p. D805-11.
31. Balzer, S., et al., *Filtering duplicate reads from 454 pyrosequencing data*. *Bioinformatics*, 2013. 29(7): p. 830-6.
32. Carpenter, E.L., et al., *Dielectrophoretic capture and genetic analysis of single neuroblastoma tumor cells*. *Front Oncol*, 2014. 4: p. 201.
33. Guo, Y., et al., *Exome sequencing generates high quality data in non-target regions*. *BMC Genomics*, 2012. 13: p. 194.
34. Ng, S.B., et al., *Exome sequencing identifies the cause of a mendelian disorder*. *Nat Genet*, 2010. 42(1): p. 30-5.
35. Liu, Q., et al., *Steps to ensure accuracy in genotype and SNP calling from Illumina sequencing data*. *BMC Genomics*, 2012. 13 Suppl 8: p. S8.
36. Bergen, A.W., et al., *Effects of DNA mass on multiple displacement whole genome amplification and genotyping performance*. *BMC Biotechnol*, 2005. 5: p. 24.
37. Voet, T., et al., *Single-cell paired-end genome sequencing reveals structural variation per cell cycle*. *Nucleic Acids Res*, 2013. 41(12): p. 6119-38.
38. Iwamoto, K., et al., *Detection of chromosomal structural alterations in single cells by SNP arrays: a systematic survey of amplification bias and optimized workflow*. *PLoS One*, 2007. 2(12): p. e1306.
39. Ning, L., et al., *Quantitative assessment of single-cell whole genome amplification methods for detecting copy number variation using hippocampal neurons*. *Sci Rep*, 2015. 5: p. 11415.
40. Cayrefourcq, L., et al., *Establishment and characterization of a cell line from human circulating colon cancer cells*. *Cancer Res*, 2015. 75(5): p. 892-901.
41. Joosse, S.A. and K. Pantel, *Biologic challenges in the detection of circulating tumor cells*. *Cancer Res*, 2013. 73(1): p. 8-11.

| WGA kit         | WGA output, mean $\pm$ st. dev., $\mu$ g |                   |                      |                   |                   |
|-----------------|--|-------------------|----------------------|-------------------|-------------------|
|                 | SKBR3<br>EDTA                            | SKBR3<br>CellSave | SKBR3<br>FFPE        | CTC<br>EDTA       | Average           |
| <b>Ampli1</b>   | 7.07 $\pm$ 1.08                          | 5.86 $\pm$ 2.23   | 6.74 $\pm$ 1.61      | 4.69 $\pm$ 3.19   | 6.09 $\pm$ 2.29   |
| <b>PicoPlex</b> | 2.86 $\pm$ 1.14                          | 3.39 $\pm$ 2.32   | 4.71 $\pm$ 0.41      | 4.01 $\pm$ 1.24   | 3.74 $\pm$ 1.56   |
| <b>REPLI-g</b>  | 15.39 $\pm$ 1.35                         | 11.37 $\pm$ 1.25  | 77.97 $\pm$<br>30.82 | 31.41 $\pm$ 12.84 | 34.04 $\pm$ 31.23 |

Table 1a. DNA output,  $\mu$ g.

| WGA kit         | PCR quality control success rate, % |                   |               |          |         |
|-----------------|-------------------------------------|-------------------|---------------|----------|---------|
|                 | SKBR3<br>EDTA                       | SKBR3<br>CellSave | SKBR3<br>FFPE | CTC EDTA | Average |
| <b>Ampli1</b>   | 100                                 | 100               | 100           | 70       | 93      |
| <b>PicoPlex</b> | 80                                  | 100               | 100           | 100      | 95      |
| <b>REPLI-g</b>  | 70                                  | 70                | 30            | 30       | 50      |

Table 1b. PCR quality control success rate, %.

Table 1. Mean DNA yield (Table 1a) and PCR quality control success rate (Table 1b) for single SK-BR-3 cells and CTCs extracted from EDTA and CellSave preservation tubes, and FFPE material, after amplification with Ampli1, PicoPlex, and REPLI-g WGA kits.

CTC – circulating tumor cell; st.dev – standard deviation.

| <b>Groups</b>                                | <b>WGA kit</b>                           | <b>Ampli1</b> |           | <b>PicoPlex</b> |           | <b>REPLI-g</b> |           | <b>SK-BR-3 genomic DNA</b> |
|--|--|---------------|-----------|-----------------|-----------|----------------|-----------|----------------------------|
|  | <b>NGS platform</b>                      | HiSeq2000     | IonProton | HiSeq2000       | IonProton | HiSeq2000      | IonProton | HiSeq2000                  |
| <b>SNP statistics</b>                        | Total SNPs                               | 9944          | 1986      | 9948            | 1695      | 403            | 64        | 17659                      |
|  | Known SNPs                               | 9679          | 1936      | 5237            | 1157      | 200            | 31        | 17251                      |
|  | Known SNPs, %                            | 97.3          | 97.5      | 52.6            | 68.3      | 49.6           | 48.4      | 97.7                       |
|  | Novel SNPs                               | 265           | 50        | 4711            | 538       | 203            | 33        | 408                        |
|  | ADO, %                                   | 9.0           | 19.8      | 24.0            | 42.4      | 100.0          | 100.0     | na                         |
|  | Common SNPs with known SNPs in reference | 7125          | 1525      | 4680            | 1073      | 173            | 30        | 17251                      |
|  | Sensitivity, %                           | 41.3          | 8.8       | 27.1            | 6.2       | 1.0            | 0.2       | 100.0                      |
|  | PPV, %                                   | 73.6          | 78.8      | 89.4            | 92.7      | 86.5           | 96.8      | na                         |
|  | <b>Indel statistics</b>                  | Total indels  | 1148      | 2688            | 2469      | 1688           | 140       | 52                         |
| Known indels                                 |  | 176           | 23        | 82              | 14        | 3              | 1         | 310                        |
| Known indels, %                              |  | 15.3          | 0.9       | 3.3             | 0.8       | 2.1            | 1.9       | 61.8                       |
| Novel indels                                 |  | 972           | 2665      | 2387            | 1674      | 137            | 51        | 192                        |
| Common indels with known indels in reference |  | 116           | 16        | 71              | 11        | 2              | 1         | 310                        |
| Sensitivity, %                               |  | 37.4          | 5.2       | 22.9            | 3.6       | 0.7            | 0.3       | 100.0                      |
| PPV, %                                       |  | 65.9          | 69.6      | 86.6            | 78.6      | 66.7           | 100.0     | na                         |
| <b>CNA analysis</b>                          | Spearman correlation coefficient (r)     | 0.66          | 0.63      | 0.81            | 0.80      | 0.25           | 0.25      | na                         |
|  | P-value                                  | <0.01         | <0.01     | <0.01           | <0.01     | <0.01          | <0.01     | na                         |

Table 2. The counts and statistics of SNP and indel calling in SK-BR-3 individual cells, obtained from EDTA-collected blood, amplified with Ampli1, PicoPlex, and REPLI-g WGA kits and sequenced with Illumina HiSeq2000 and ThermoFisher IonProton NGS platforms.

Total SNPs/indels – number of all identified SNPs/indels. Known SNP – fraction of SNPs/indels, present in SNP database. Novel SNPs/indels – number of SNPs/indels, not present in SNP database. ADO – allelic dropout. PPV – positive predictive value.

| Groups of experiments                        | WGA kit                                  | Ampli1       |          |          | PicoPlex  |          |          | SK-BR-3 genomic DNA |
|--|--|--------------|----------|----------|-----------|----------|----------|---------------------|
|  | NGS                                      | HiSeq2000    |          |          | HiSeq2000 |          |          | HiSeq2000           |
|  | Material preservation                    | EDTA         | CellSave | CellSave | EDTA      | CellSave | CellSave | Not applied         |
|  | Number of cells                          | 1            | 1        | 1        | 1         | 1        | 1        | .8x10 <sup>6</sup>  |
| SNP statistics                               | Total SNPs                               | 9944         | 7826     | 4088     | 9948      | 9738     | 9821     | 17659               |
|  | Known SNPs                               | 9676         | 6189     | 2857     | 5237      | 2522     | 4457     | 17251               |
|  | Known SNPs, %                            | 97.3         | 79.1     | 69.9     | 52.6      | 25.9     | 45.4     | 97.7                |
|  | Novel SNPs                               | 265          | 1637     | 1231     | 4711      | 7216     | 5364     | 408                 |
|  | ADO rate, %                              | 9.0          | 36.4     | 78.5     | 54.0      | 74.3     | 66.4     | na                  |
|  | Common SNPs with known SNPs in reference | 7125         | 5680     | 2381     | 4680      | 2088     | 2885     | 17251               |
|  | Sensitivity, %                           | 41.3         | 32.9     | 13.8     | 27.1      | 12.1     | 16.7     | na                  |
|  | PPV, %                                   | 73.6         | 91.8     | 83.3     | 89.4      | 82.8     | 64.7     | na                  |
|  | Indel statistics                         | Total indels | 1148     | 723      | 165       | 2469     | 790      | 914                 |
| Known indels                                 |  | 176          | 89       | 36       | 82        | 24       | 63       | 310                 |
| Known indels, %                              |  | 15.3         | 12.3     | 21.8     | 3.3       | 3.0      | 6.9      | 61.8                |
| Novel indels                                 |  | 972          | 634      | 129      | 2387      | 766      | 851      | 192                 |
| Common indels with known indels in reference |  | 116          | 76       | 32       | 71        | 19       | 42       | 310                 |
| Sensitivity, %                               |  | 37.4         | 24.5     | 10.3     | 22.9      | 6.1      | 13.6     | na                  |
| PPV, %                                       |  | 65.9         | 85.4     | 88.9     | 86.6      | 79.2     | 66.7     | na                  |
| CNA analysis                                 | Spearman correlation coefficient (r)     | 0.64         | 0.84     | 0.25     | 0.81      | 0.69     | 0.09     | na                  |
|  | P-value                                  | <0.01        | <0.01    | <0.01    | <0.01     | <0.01    | <0.01    | na                  |

Table 3. The counts and statistics of SNP and indel calling in single SK-BR-3 cells, analyzed in duplicates, obtained from CellSave-preserved blood, in comparison to single SK-BR-3 cells, obtained from EDTA-collected blood.

Total SNPs/indels – number of all identified SNPs/indels. Known SNP – fraction of SNPs/indels, present in SNP database. Novel SNPs/indels – number of SNPs/indels, not present in SNP database. ADO – allelic dropout. PPV – positive predictive value.

| Groups  | Cell   | <b>CTC-1</b>         | <b>CTC-2</b>        |
|---|--|----------------------|---------------------|
|   | WGA kit  | PicoPlex             |                     |
|   | NGS  | HiSeq2000            |                     |
|   | Blood preservative   | EDTA                 |                     |
| SNP statistics and concordance between CTCs   | Total SNPs   | 34994                | 14658               |
|   | Known SNPs   | 4304                 | 6030                |
|   | Known SNPs, %  | 12.3                 | 41.1                |
|   | Novel SNPs   | 30690                | 8628                |
|   | Known SNPs common in both datasets                           | 1135                 |                     |
|   | Fraction of common known from known identified in dataset, % | 26.4                 | 18.8                |
| Indel statistics and concordance between CTCs | Total indels   | 4383                 | 4103                |
|   | Known indels   | 42                   | 81                  |
|   | Known indels, %  | 1.0                  | 2.0                 |
|   | Novel indels   | 4341                 | 4022                |
|   | Known indels common in both datasets                         | 15                   |                     |
|   | Fraction of common known from known identified in dataset, % | 37.7                 | 18.5                |
| CNA analysis                                  | Correlation between CTCs, r                                  | 0.10                 |                     |
| Mutation analysis (amino acid change)         | Present in both CTCs   | <i>CHEK2</i> (K373E) |                     |
|   |  | <i>KIT</i> (M541L)   |                     |
|   |  | <i>PRAME</i> (W7R)   |                     |
|   | Present in only one CTC                                      | <i>FGFR2</i> (Y376C) | wt                  |
|   |  | wt                   | <i>TP53</i> (E285K) |

Table 4. The counts and statistics of SNP and indel calls in CTCs.

Total SNPs/indels – number of all identified SNPs/indels. Known SNP – fraction of SNPs/indels, present in SNP database. Novel SNPs – number of SNPs, not present in SNP database. ADO – allelic dropout. r – Spearman correlation coefficient. wt – wild type.

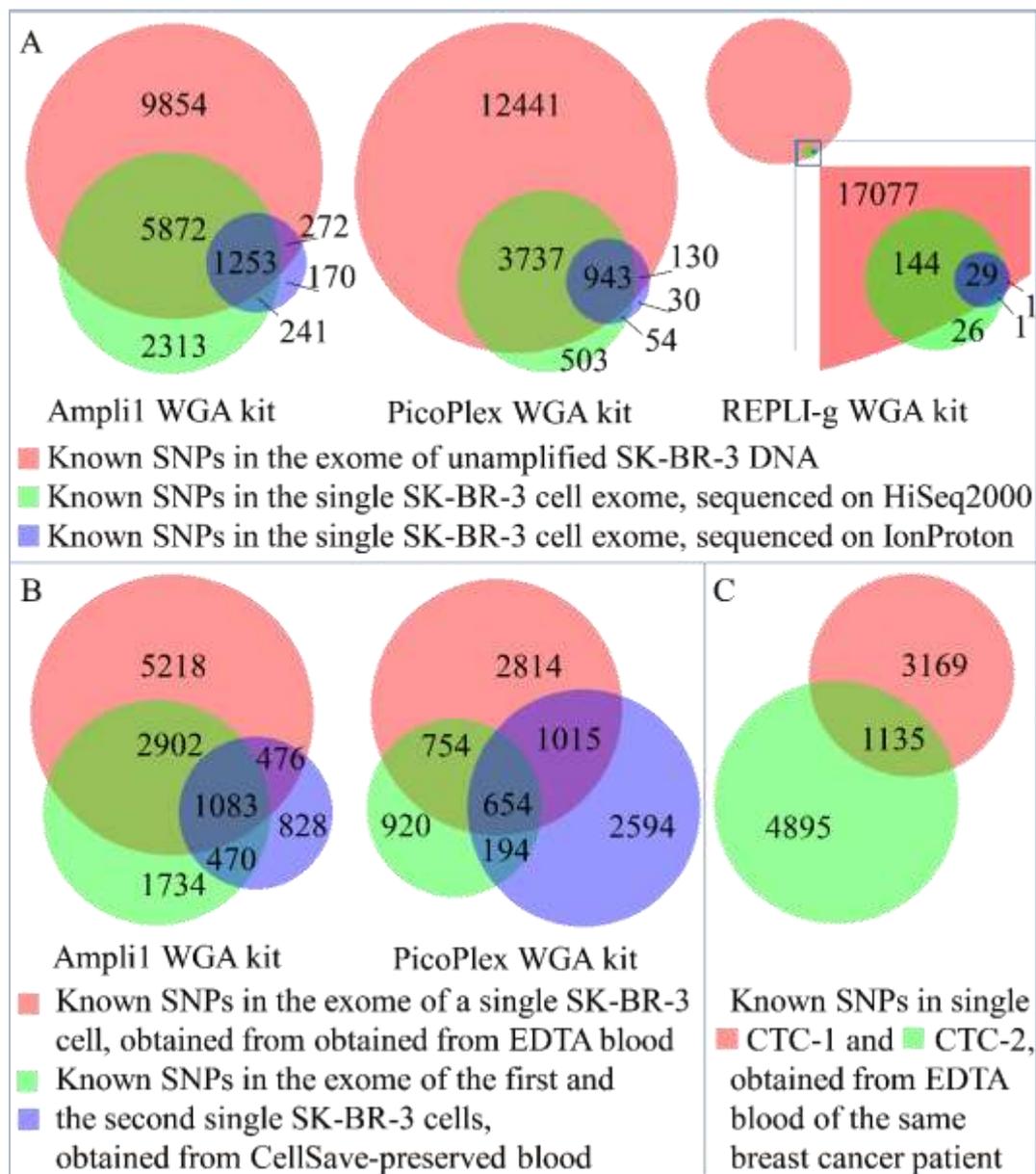


Figure 1. Distribution of identified known SNPs between datasets.

- A. Known SNPs identified in single cells, amplified with Ampli1, PicoPlex, and REPLI-g WGA kits and obtained from EDTA-preserved blood in comparison to unamplified DNA.
- B. Known SNPs identified in single cells, amplified with Ampli1 or PicoPlex and obtained from EDTA- and CellSave-preserved blood in comparison to unamplified DNA from unfixed cells.
- C. Known SNPs identified in single CTCs, amplified with PicoPlex in comparison to each other.

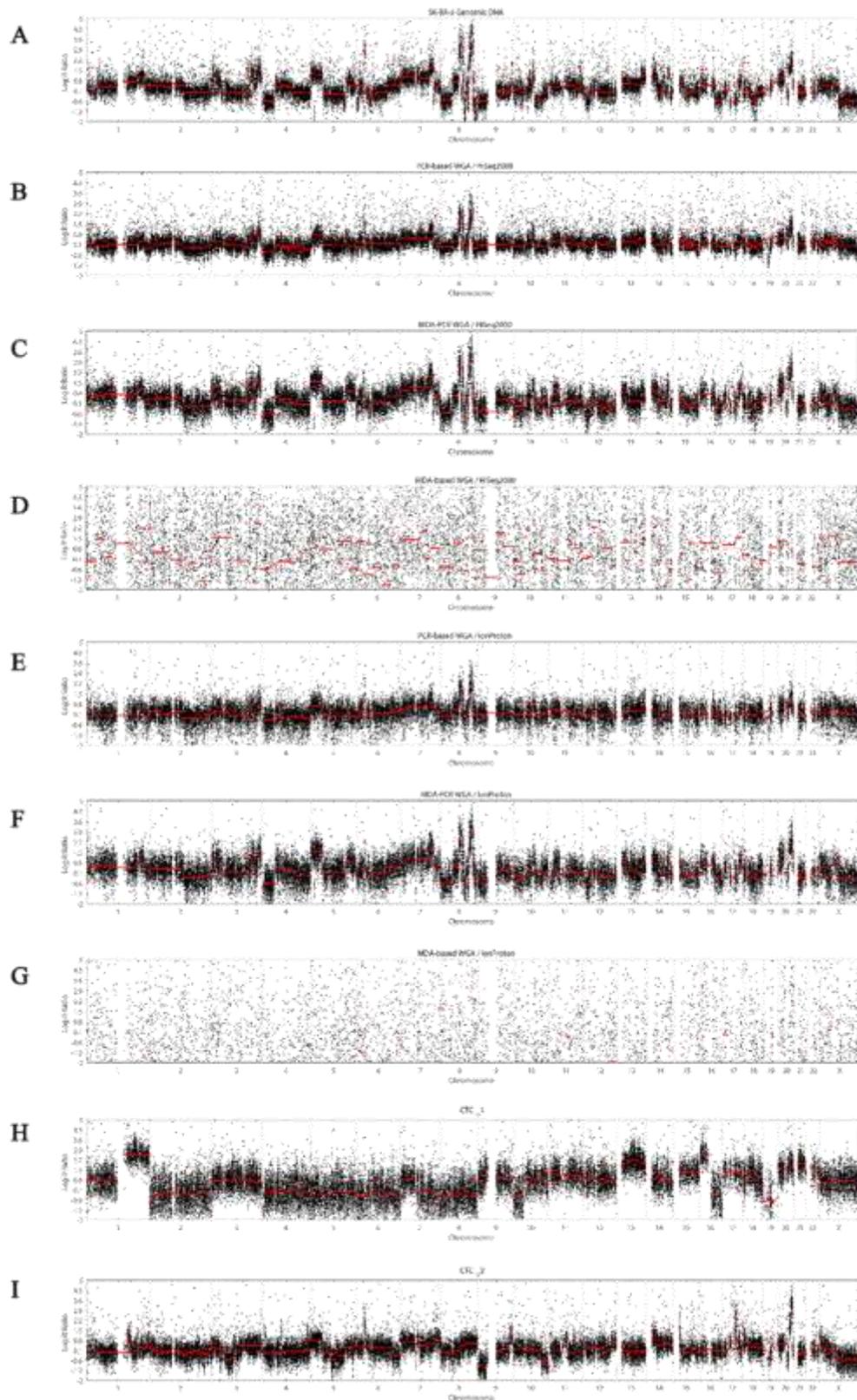


Figure 2. Plots of CNA profiles along the whole genome (x axis).

A – CNA profile of unamplified DNA from unfixed cells. B-G – plots of CNAs in single SK-BR-3 cells, obtained from EDTA-preserved blood. H, I – CNA profiles of individual CTCs, obtained from EDTA-preserved blood of the same breast cancer patient. WGA kits: B, E – Ampli1; C, F, H, I – PicoPlex; D, G – REPLI-g.

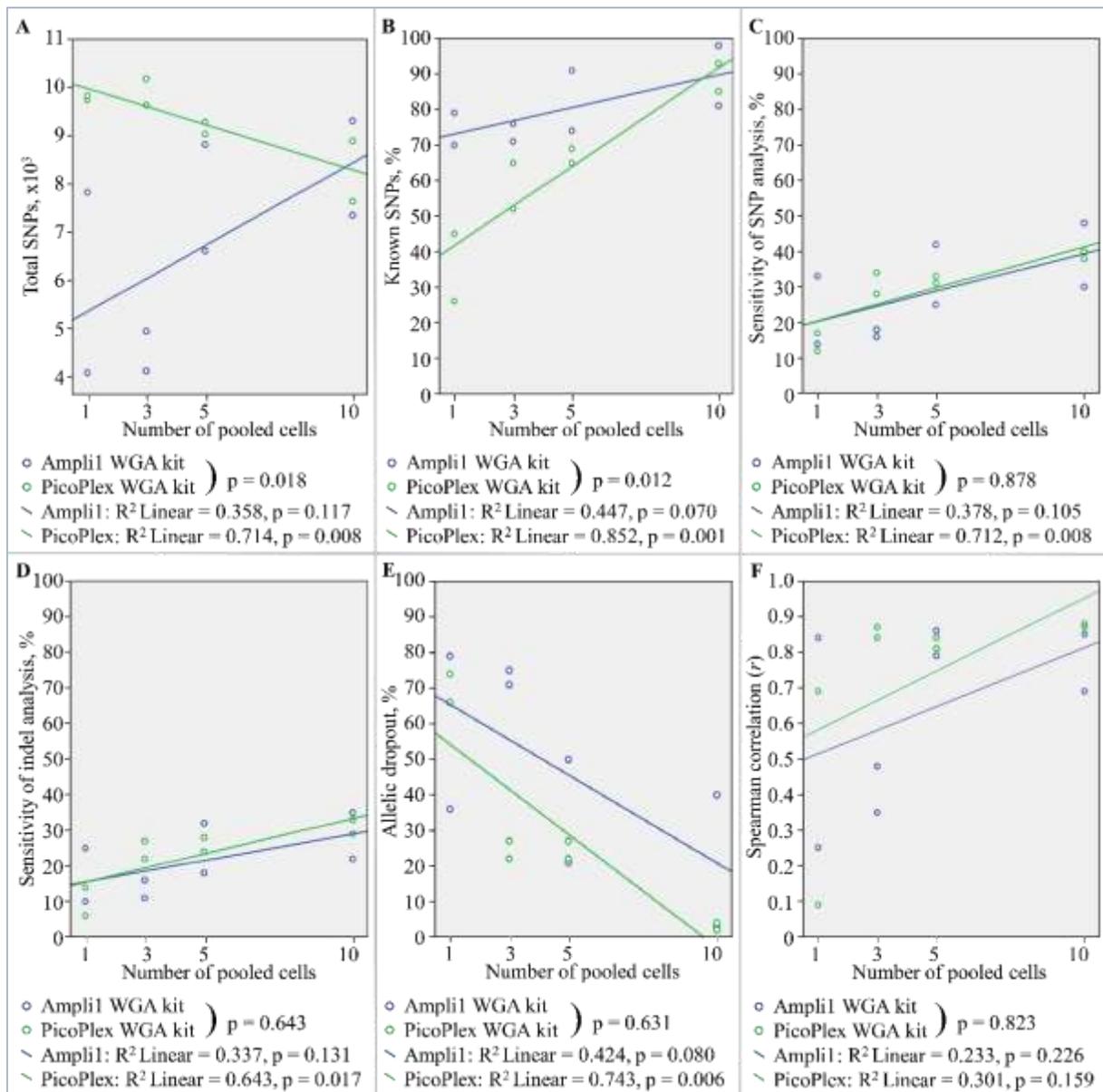
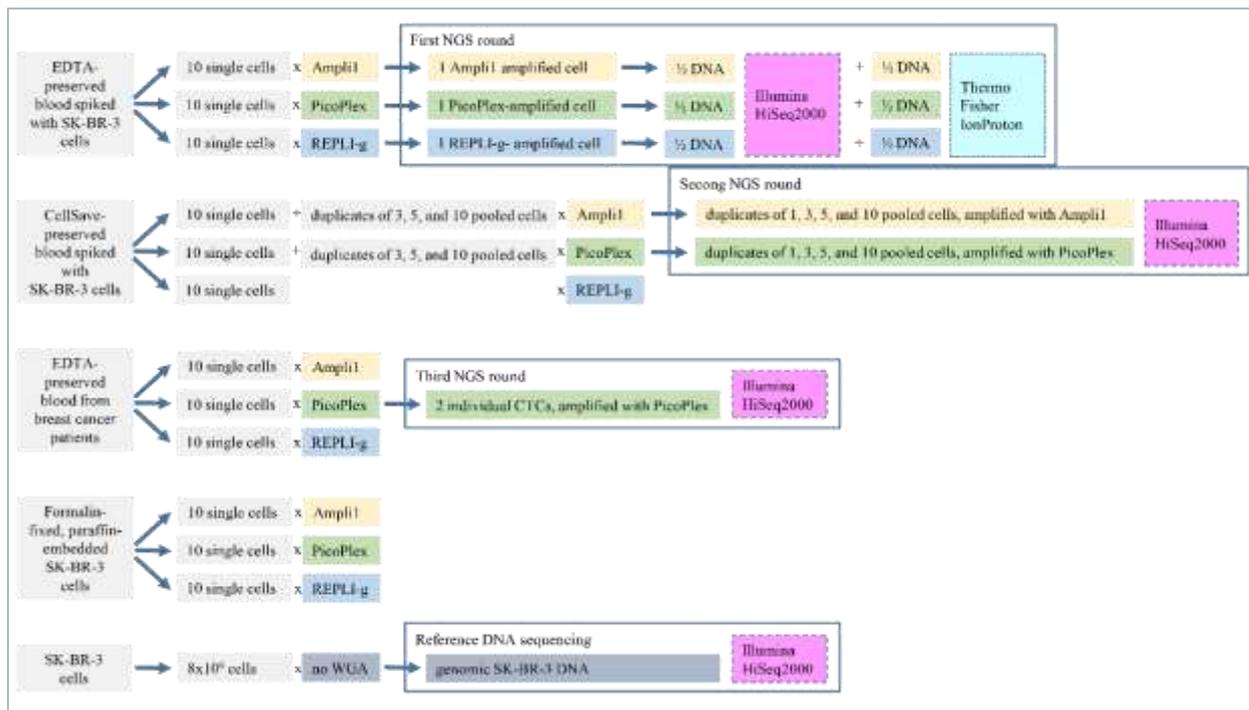


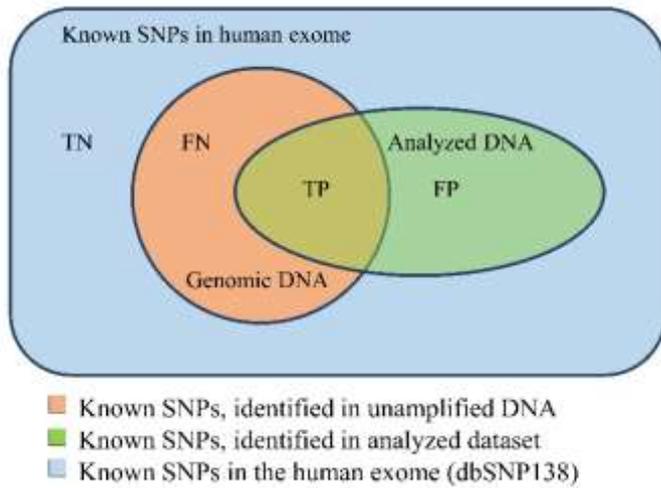
Figure 3. Characteristics of single and pooled 3, 5, and 10 SK-BR-3 cells, obtained from CellSave-preserved blood, amplified with Ampli1 and PicoPlex WGA kits, and sequenced with HiSeq2000 NGS platform.

- Total identified SNPs.
- Fraction of known identified SNPs.
- Sensitivity of SNP calling analysis.
- Sensitivity of indel calling analysis.
- Allelic dropout.
- Correlation of CNA profiles of the analyzed samples with CNA profile of unamplified DNA.



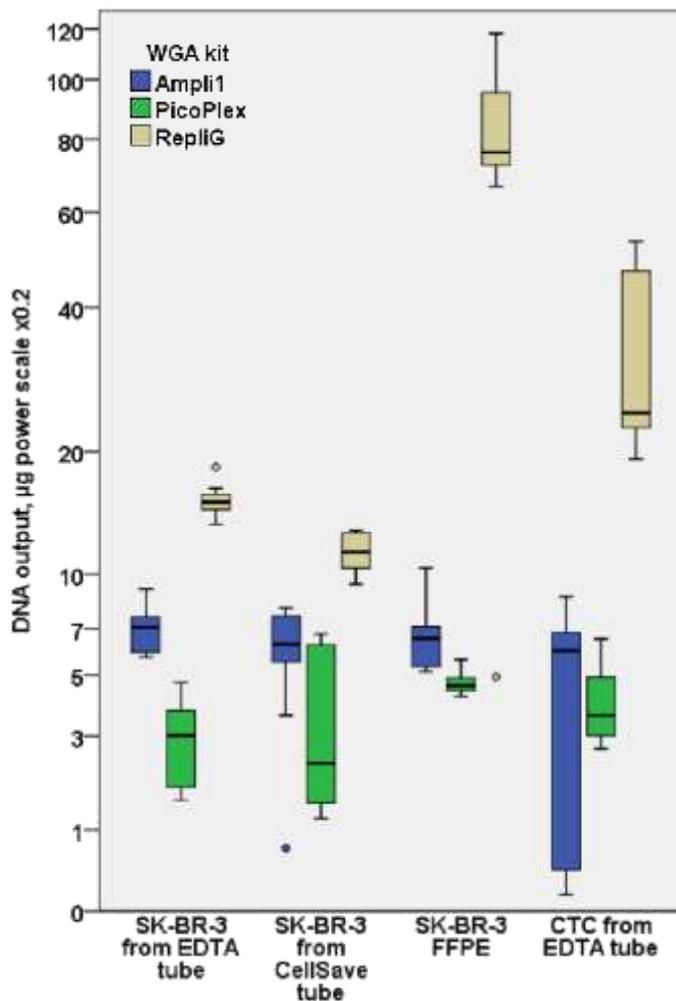
Supplementary Figure 1. Experimental design.

SK-BR-3 breast cancer cell line cells were spiked into blood donors' blood, collected into EDTA- and CellSave tubes. Previously the same cell line cells were used to prepare formalin-fixed, paraffin-embedded material (FFPE). Blood from breast cancer patients was drawn into EDTA-tubes. Blood and FFPE samples were processed and used for picking of individual tumor cells: A) 10 individual SK-BR-3 cells spiked and picked from EDTA-preserved blood; B) 10 individual SK-BR-3 cells spiked and picked from CellSave-preserved blood; C) 10 single SK-BR-3 cells picked from FFPE SK-BR-3 cells; and D) 10 individual CTCs from breast cancer patients, from each group of samples. Collected cells were used for WGA with Ampli1, PicoPlex, and REPLI-g WGA kits. For the EDTA-preserved SK-BR-3 cells, 3 representative whole genome amplified cells, one per WGA kit, underwent NGS on both HiSeq200 and IonProton platforms. Taking the results of SNP, indel, and CNA analyses into consideration, next NGS round on HiSeq2000 included duplicates of 1, 3, 5, and 10 CellSave-preserved SK-BR-3 cells, amplified with Ampli1 and PicoPlex, and 2 patient CTCs, amplified with PicoPlex. Unamplified SK-BR-3 DNA from unfixed cells was sequenced on HiSeq2000.



Supplementary Figure 2. Definition of SNP calls, identified in analyzed dataset in comparison to reference as true-positive, false-positive, true-negative, and false-negative SNPs.

TP – true-positive calls; FP – false-positive calls; TN – true-negative calls; FN – false-negative calls.



Supplementary Figure 3. DNA yield in respect to WGA kit in groups of single SK-BR-3 cells, picked from EDTA- and CellSave preserved blood, FFPE material, and CTCs, picked from EDTA-preserved blood of breast cancer patients.

### Supplementary Material 1.

The quality of the WGA products was assessed by a multiplex PCR of the *GAPDH* gene producing fragments of 100, 200, 300, and 400 bp fragments from non-overlapping target sites as described elsewhere [1]. Since the original 200 bp fragment is not amplified by the Ampli1 WGA kit, we used the following primers to produce a 200 bp fragment: fw: 5'-AAGATCATCAGGTGAGGAAGGC-3' rev: 5'-CCCCAGCTCTCATACCATGAGTC-3'.

PCR conditions were optimized for a reaction of 15 µl total volume with input of 100 ng DNA as follows: 0.75 U AmpliTaq Gold DNA Polymerase (Applied Biosystems, 4486226), 0.2 mM of each ATP, GTP, CTP, TTP; 0.136 µM of each primer, and 2 mM MgCl<sub>2</sub> (Applied Biosystems, R01911). Human leukocyte DNA was used as positive control for the multiplex PCR. The PCR program was as follows: 95°C for 5 min; 35 cycles of 94°C for 30 sec, 64°C for 30 sec, 72°C for 45 sec; final elongation at 72°C for 7 min. PCR products were analyzed in a 2% agarose TAE ethidium bromide-stained gel. Samples were considered to be of sufficient quality for further analyses if at least one of the 200, 300, and 400 bp bands was detected.

1. van Beers EH, Joesse SA, Ligtenberg MJ, Fles R, Hogervorst FB, Verhoef S and Nederlof PM. A multiplex PCR predictor for aCGH success of FFPE samples. *British journal of cancer*. 2006; 94(2):333-337.

## Supplementary Material 2.

Data analysis was performed according to the GATK Best Practices recommendations [1, 2]. Exome capturing was performed with “BGI Exome Enrichment Kit (59M) and Capture” for sequencing on HiSeq2000 and “Ion AmpliSeq exome RDY kit” for sequencing on IonProton. The corresponding exome regions were used respectively for calculation of descriptive statistics over target regions and during post-alignment data processing. To ensure the location of made calls within the exome and to unify results of SNP and indel calling between the datasets, SNP/mutation and indel discovery was limited to protein coding exons only (downloaded from the CCDS Project database [3, 4]).

### *Reference datasets used for the analysis*

|  |          |
|--|----------|
| Human genome UCSC hg19   | [5]      |
| dbsnp_138.hg19.vcf   | [6]      |
| Mills_and_1000G_gold_standard.indels.hg19.sites.vcf  | [7]      |
| UCSC_CCDS_per_exon.bed   | [4]      |
| HG19 snpEff database   | [8]      |
| Control file for FREEC was generated out of alignment of 185 reference European female genomes, obtained from 1000 Genome database | [9]      |
| GEM_mapp_hg19/out100m1_hg19.gem  | [10]     |
| COSMIC database  | [11, 12] |

### *Programs*

|               |          |
|---------------|----------|
| bwa mem       | [13]     |
| gatk          | [14]     |
| picard        | [15]     |
| samtools      | [16]     |
| trimmomatic   | [17]     |
| snpEff        | [18]     |
| snpSift       | [19]     |
| Control-FREEC | [20, 21] |

|   | Program and command                 | Specifications or differing from default parameters  | References   |
|---|-------------------------------------|--|--|
| <b>1A. PREPROCESSING AND ALIGNMENT FOR PAIRED-END HISEQ2000 READS</b> |                                     |  |  |
| Clip WGA adapters if present  |                                     | For Ampli1 and PicoPlex amplified samples  |  |
| Trim  | trimmomatic                         | PE ILLUMINACLIP:2:30:10<br>LEADING:3 TRAILING:3<br>SLIDINGWINDOW:4:10<br>MINLEN:25 TOPHRED33 |  |
| Align to the genome   | bwa mem                             | -t 30 -v 0 -M -R   | UCSC hg19  |
| <b>1B. PREPROCESSING AND ALIGNMENT FOR SINGLE-END IONPROTON READS</b> |                                     |  |  |
| Sort and convert bam file to fastq                                    | samtools sort<br>samtools<br>bam2fq | -n<br>-n -O -s   |  |
| Clip WGA adapters if present  |                                     | For Ampli1 and PicoPlex amplified samples  |  |
| Trim  | trimmomatic                         | SE LEADING:3 TRAILING:3<br>SLIDINGWINDOW:4:10<br>MINLEN:25 TOPHRED33                         |  |
| Align to the genome   | bwa mem                             | -t 30 -v 0 -M -R   | UCSC hg19  |
| <b>2. POSTALIGNMENT PROCESSING</b>                                    |                                     |  |  |
| Sort sam file and convert to bam                                      | picard<br>SortSam                   | SORT_ORDER=coordinate<br>VERBOSITY=ERROR<br>COMPRESSION_LEVEL=0                              |  |
| Mark duplicates   | picard<br>MarkDuplicates            | VERBOSITY=ERROR<br>COMPRESSION_LEVEL=0   |  |
| Index bam file  | samtools index                      |  |  |
| Realign indels  | gatk Realigner<br>Target Creator    | -nt 24   | UCSC hg19<br>Mills_and_1000G_gold_standard.inde<br>ls.hg19.sites.vcf                           |
|   | gatk<br>IndelRealigner              | -compress 0 -model<br>USE_READS -LOD 0.4   | UCSC hg19<br>Mills_and_1000G_gold_standard.inde<br>ls.hg19.sites.vcf                           |
| Recalibrate bases   | gatk Base<br>Recalibrator           | -nct 24  | UCSC hg19<br>dbsnp_138.hg19.v<br>cf<br>Mills_and_1000G_gold_standard.inde<br>ls.hg19.sites.vcf |
|   | gatk<br>PrintReads                  | -BQSR -compress 0  | UCSC hg19  |

| 3. DISCOVER SNPS AND INDELS                     |                       |   |  |
|---|-----------------------|---|--|
| SNP and indel calling                           | gatk HaplotypeCaller  | -stand_call_conf 30 -stand_emit_conf 30 -gt_mode DISCOVERY -out_mode EMIT_ALL_CONFIDENT_SITES -ploidy 3 --annotation FisherStrand --annotation QualByDepth --annotation HaplotypeScore --annotation HomopolymerRun --annotation RMSMappingQuality --annotation ReadPosRankSumTest | UCSC hg19 dbsnp_138.hg19.vcf UCSC_CCDS_per_exon.bed  |
| Select for SNPs                                 | gatk SelectVariants   | -selectType SNP   | UCSC hg19  |
| Annotate HRun                                   | gatk VariantAnnotator | --annotation HomopolymerRun   | UCSC hg19 dbsnp_138.hg19.vcf   |
| Filter for quality and GQ                       | snpSift filter        | ( QD >= 5 ) & ( MQ > 25 ) & ( QUAL > 30 ) & ( FS < 60 ) & ( SOR < 4 ) & ( HRun < 5 ) & ( GEN[*].GQ >= 20 )  |  |
| Annotate with snpEff                            | snpEff                |   | HG19 snpEff database   |
| Select for INDELS                               | gatk SelectVariants   | -selectType INDEL   | UCSC hg19  |
| Filter for quality and GQ                       | snpSift filter        | ( QD >= 2 ) & ( MQ > 25 ) & ( QUAL > 20 ) & ( FS < 200 ) & ( SOR < 10 ) & ( GEN[*].GQ >= 20 )   |  |
| Annotate with snpEff                            | snpEff                |   | HG19 snpEff database   |
| investigate mutations (for patient's data only) | gatk VariantAnnotator | #Annotate with COSMIC data -comp: COSMIC #{Cosmic} -resource #{Cosmic}  | hg19_cosmic_v54_120711 (#{Cosmic})   |
| 4. COPY NUMBER ANALYSIS                         |                       |   |  |
| Create mpileup file                             | samtools mpileup      | -E  |  |
| Run Control-FREEC                               | freec                 | breakPointType = 4, forceGCcontentNormalization = 2, noisyData = TRUE, ploidy = 3 (for SK-BR-3, and ploidy = 2 for CTCs), printNA = FALSE, readCountThreshold = 50, sex = XX, window = 30000, uniqueMatch = TRUE  | UCSC hg19 GEM_mapp_hg19/out100m1_hg19.gem control file for FREEC for SK-BR-3 analysis, no control for CTCs |

1. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*. 2011; 43(5):491-498.
2. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S and DePristo MA. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current protocols in bioinformatics / editorial board, Andreas D Baxevanis [et al]*. 2013; 11(1110):11 10 11-11 10 33.
3. Farrell CM, O'Leary NA, Harte RA, Loveland JE, Wilming LG, Wallin C, Diekhans M, Barrell D, Searle SM, Aken B, Hiatt SM, Frankish A, Suner MM, Rajput B, Steward CA, Brown GR, et al. Current status and new features of the Consensus Coding Sequence database. *Nucleic acids research*. 2014; 42(Database issue):D865-872.
4. CCDS Project.
5. Human genome HG19
6. Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda (MD): National Center for Biotechnology Information, National Library of Medicine. (dbSNP Build ID: 137). Available from: <http://www.ncbi.nlm.nih.gov/SNP/>.
7. Mills and 1000G gold standard indels.
8. HG19 snpEff database.
9. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT and McVean GA. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491(7422):56-65.
10. GEM\_mapp\_hg19/out100m1\_hg19.gem for Control-FREEC.
11. COSMIC database.
12. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, Ding M, Bamford S, Cole C, Ward S, Kok CY, Jia M, De T, Teague JW, Stratton MR, McDermott U, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic acids research*. 2015; 43(Database issue):D805-811.
13. Li H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

14. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M and DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*. 2010; 20(9):1297-1303.
15. Picard tools.
16. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R and Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25(16):2078-2079.
17. Bolger AM, Lohse M and Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014; 30(15):2114-2120.
18. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X and Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*. 2012; 6(2):80-92.
19. Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM and Lu X. Using *Drosophila melanogaster* as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Frontiers in genetics*. 2012; 3:35.
20. Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, Janoueix-Lerosey I, Delattre O and Barillot E. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics*. 2012; 28(3):423-425.
21. Boeva V, Zinovyev A, Bleakley K, Vert JP, Janoueix-Lerosey I, Delattre O and Barillot E. Control-free calling of copy number alterations in deep-sequencing data using GC-content normalization. *Bioinformatics*. 2011; 27(2):268-269.

## 6. PUBLICATION 3

Clonal evolution of metastatic breast cancer:  
two cases, two progression models

Anna Babayan<sup>1</sup>, Katharina Prieske<sup>2</sup>, Daniela Indenbirken<sup>3</sup>, Malik Alawi<sup>3,4</sup>, Eike Burandt<sup>5</sup>, Adam Grundhoff<sup>3</sup>, Volkmar Müller<sup>2</sup>, Klaus Pantel<sup>1\*</sup>, Simon A. Joosse<sup>1\*</sup>

<sup>1</sup>Department of Tumor Biology, <sup>2</sup>Department of Gynecology, <sup>4</sup>Bioinformatics Core, <sup>5</sup>Department of Pathology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

<sup>3</sup>Heinrich-Pette-Institute, Leibniz-Institute for Experimental Virology (HPI), Hamburg, Germany,

\* shared senior authorship

manuscript in preparation

## **Abstract**

**BACKGROUND.** Early dissemination of circulating tumor cells (CTCs) into the blood circulation of cancer patients allows for parallel genetic evolution of the primary tumor and metastases. Anti-cancer treatment failure is caused by the genetic and phenotypic heterogeneous properties of the disease. This study investigated the genomic make-up of CTCs originating from metastasis as a so called “liquid biopsy” and compared these with the different clones of the archived autologous primary tumor on single cell level.

**METHODS.** From two breast cancer patients, individual CTCs were isolated from blood using Ficoll density gradient followed by micromanipulation of keratin positive cells. Single cells from archived primary breast tumors from the same patients were captured by laser microdissection. DNA was isolated and amplified by whole genome amplification and copy number alterations (CNA) were obtained by shallow, whole genome next generations sequencing (NGS).

**RESULTS.** From the first breast cancer patient, the genomes from 50 single cells from the primary tumor were sequenced. An unsupervised phylogenetic cluster analysis based on CNA revealed five distinct clusters with increasing chromosomal instability. Using support vector machine (SVM) learning, the CNA profiles of 42 CTCs were residing mostly to the first three clusters, whereas only one CTC resided to the fourth cluster and none to the fifth. The tumor from a second breast cancer patient displayed the presence of three genetically distinct clones with increasing chromosomal instability after sequencing 11 single cells. CTCs (n=12) from this patient at metastatic disease were classified to the last branch using SVM, however with low probability. Repeating unsupervised clustering on the genomes of the primary tumor tissue and CTCs, a fourth branch was formed with CTCs only.

**CONCLUSION.** Our results suggest that therapy resistant metastases in breast cancer patients can originate from tumor clones from early stages of tumor evolution and may genetically still be similar (patient 1). On the other hand, further genetic progression may also take place (patient 2) where after the genetic landscape of the metastasis does not resemble the primary tumor anymore. These results underline the importance of “liquid biopsy” in the diagnosis of metastatic cancer.

## INTRODUCTION

Intra-tumor heterogeneity is a well-known phenomenon in human cancers and may be caused by clonal evolution of the tumor. Current screening technologies allow for the investigation of cancer heterogeneity on all levels of molecular organization: genomic, epigenomic, transcriptomic, metabolomic, and proteomic [1]. Intra-tumor heterogeneity on functional level, such as transcriptome, metabolome, and proteome, might be caused by niche adaptation mechanisms and varies through cell cycle dynamics, and thus does not necessarily reflect clonality of the cancer. Genetic intra-tumor heterogeneity caused by clonal evolution of cancer and reflecting clonal origin of a cell lineage, is supposedly more stable and thereby providing accessible information about clonal evolution of cancer. Genomic intra-tumor heterogeneity might be investigated on single cell level by using for instance next generation sequencing (NGS).

It has been long discussed whether metastatic dissemination is an early or late event in cancer evolution, resulting in development of two progression models.

The first model, the linear progression model, postulates that metastasis-initiating cells originate from most progressed clone(s) of the primary tumor, which were developed during evolution of the primary tumor with selection for clones with high metastatic proclivity [2, 3].

On the other hand, data showing the metastatic potential of primary tumors at early stages, led to the coinage of the parallel progression model [4, 5]. This model proposes the presence of metastatic potential already in the early disease progression, leading to early dissemination of circulating tumor cells (CTCs) into circulatory system with subsequent parallel and independent evolution of the primary tumor and metastases [6, 7]. The fact that CTCs can be found in blood of both late and early stage cancer patients suggests the parallel progression model is more likely than the linear progression model [8].

An alternative scenario of cancer metastasis, proposed in our institute, suggests continuous dissemination of tumor cells from a primary tumor developing higher metastatic potential over the time during further evolution of the primary tumor [9].

Understanding tumor progression and the metastatic cascade in breast cancer is of tremendous value because distant metastases development is the most challenging issue in clinical management of cancer. Investigation of progression mechanisms and clonal evolution in cancer could identify molecular signatures,

involved in progression and metastatic process. Parallel genetic evolution of the primary tumor and distant metastases might explain failure of systemic endocrine therapy, which prescription is based on ER-positivity of the primary tumor. ER-positive primary breast cancers, treated with endocrine therapy, often demonstrate presence of ER-negative metastases, insensitive to anti-estrogen therapy [10, 11].

Although the origin of intra-tumor heterogeneity is not fully understood yet, it seems to play a major role in a complex process of carcinogenesis and development of metastatic disease [12-14]. Intra-tumor heterogeneity and clonal diversity per se might promote cancer evolution by serving more diverse input material for Darwinian selection [15]. The newly revised “seed-and-soil” hypothesis postulates that heterogeneity of cell characteristics, survival in the circulation, and effective homing in new environment are the crucial conditions for successful metastasizing [3]. Because only very few tumor cells meet these requirements, metastasis is a biologically inefficient process (rev. in [16]). However, high amount of CTCs with heterogeneous characteristics provide extensive source for potential metastases [9]. CTCs embody an intermediate step between primary tumor and metastases. CTCs reflect the biology of the primary tumor or metastases from which they originate [17]. Furthermore, CTCs carry characteristics potentially enabling metastases’ establishment. Therefore the genetic makeup of CTCs may provide a unique insight into cancer evolution.

All current models of carcinogenesis emphasize genetic changes as one of the initiating conditions for cancer development [18-20]. Such genetic changes include point mutations, copy number aberrations (CNA), and copy number-neutral rearrangements of genetic material. NGS in combination with WGA provides a powerful tool for the investigation of single cell genetics. Despite mutation analysis of single cells is possible, breast cancer is characterized by overall prevalence of CNAs over mutations [21]. Genome-wide studies of cancer clonality on single cells require well established, reproducible approaches for WGA and NGS analysis. The challenge of single cell genome-wide studies lies between the need of DNA amplification and the introduction of PCR artefacts during WGA and NGS and identification of objective cell-specific genomic aberrations. Technical aspects of CNA analysis on tissue and CTC material have been covered in our previous paper [Babayan et al., 2016]. In the study present here, we investigated clonal evolution of human breast cancer on primary tissue and CTCs from two metastatic breast cancer patients.

## MATERIALS AND METHODS

### **Blood samples**

Two metastatic breast cancer patients were enrolled into this study during their treatment at the University Medical Center Hamburg-Eppendorf after giving informed consent (ethics review board Ärztenkammer Hamburg approval number OB/V/03). Ten ml blood of each patient was drawn into EDTA tubes (01.1605.001, Sarstedt) and processed within 2 hours as previously described [Babayan et al., 2016, [22]. Mononuclear cell fraction was stained for protein expression of keratins, CD45, and ER. Keratin and DAPI positive, but CD45 negative cells were considered as CTCs. Individual CTCs were picked by micromanipulation (micro injector CellTramVario and micromanipulator TransferManNKII, Eppendorf Instruments, Hamburg, Germany). Each individual cell was transferred in 1µl of PBS into the cap of a 200µl PCR tube and stored at -80°C for further processing.

### **Archival tumor tissue**

From the two enrolled subjects, formalin-fixed, paraffin-embedded tissue of the primary tumors was obtained. Five µm thick sections were cut and transferred onto membrane slides NF 1.0 PEN (415190-9081-000, Zeiss). The tissue sections were dried for 1 hour at 75°C, deparaffinized in xylene for 2x 10min, rehydrated in ethanol (2x 100%, 2x 96%, 2x 80%, 2x 70%) 30s each, and finally rinsed with water for 3min as described before [23]. DNA cross-links were removed by incubating the slides in 1 M NaSCN overnight at 37°C. Subsequently, the slides underwent heat-induced antigen retrieval by boiling with citrate buffer pH6 (S1699, Dako) in a pressure cooker at 125°C for 5min. After antigen retrieval the slides were washed 3x 3min with TBST (50mM Tris, 150mM NaCl, 0.05% Tween 20, pH 7.6) and incubated with primary anti-human ER antibody (ab16660, Abcam, diluted 1:50 with antibody diluent, S3022, Dako) at 4°C overnight. Next, the slides were washed with TBST 3x 3min and the staining was completed with Peroxidase/DAB+ based Dako REAL™ detection system (K5001, Dako) according to the manufacturer's recommendations. Briefly, the slides were incubated first with Dako REAL™ biotinylated secondary antibodies for 10 min (solution A, K5001), washed 3x 3min in TBST, incubated for 5 min with Dako REAL™ peroxidase-blocking solution (Dako, S2023), washed 3x 3min in TBST, incubated with Dako REAL™ streptavidin peroxidase (HRP) for 10 min (solution B, K5001), washed 3x 3min in TBST, and exposed to DAB+ substrate for 10 min (Dako REAL™ DAB+

Chromogen, solution C, diluted 1:50 with Dako REAL™ HRP Substrate Buffer, solution D, both from K5001). Subsequently slides were washed 3x 3min with water, stained with haematoxylin for 30 s, rinsed with water and completely air-dried (>2 hours).

Isolation of ER-positive and ER-negative cells from the tissue sections was performed by laser microdissection using a PALM MicroBeam system (Carl Zeiss Group, Goettingen, Germany). Cells were collected into adhesive cap 500 µl PCR tubes (415190-9211-000, Zeiss) with 5µL of freshly prepared lysis buffer (20 mM TrisHCl pH8.0, 0.1 mM EDTA, 0.5% Nonidet P40 (M3165.0250, Genaxxon), 1% proteinase K (19131, Qiagen)). The tubes were pulse-vortexed for 15 s in upside-down position. The tubes were centrifuged at 15000 rcf for 10min and incubated in thermocycler with preheated lid (110°C) at 56°C for 16h with final heating step at 90°C for 10min to inactivate proteinase K.

### **Whole genome amplification (WGA) and quality control**

WGA was performed according to the manufacturers' recommendations using the PicoPlex WGA Kit for single cells (New England Biolabs, E2620L). The WGA products were cleaned up with NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel, 740609). DNA concentration of WGA products was measured with a Nanodrop1000 (PepLab, Erlangen, Germany). The quality of the WGA products was assessed by a multiplex PCR of the *GAPDH* gene as described elsewhere [Babayan et al., 2016]. Samples were considered of sufficient quality for further analyses if at least one of 200-400 bp bands was detectable.

### **Next generation sequencing (NGS)**

Amplified DNA of each single cell was sequenced by using shallow whole genome sequencing with Illumina's HiSeq2000 NGS platform.

### **Data analysis**

Raw data obtained in fastq format underwent adapter clipping for removal of WGA adapters: the first/last 14 bases were trimmed as suggested by the manufacturer. Further data analysis included alignment of reads on human reference genome (hg 19) with BWA-MEM [24]. The resulting SAM file was filtered for ambiguously-mapped and low-quality reads (SAMtools [25]), sorted in coordinate order, indexed, filtered for duplicates (Picard [26]), realigned around indels, and base quality recalibrated (GATK

[27]). Copy number alterations (CNAs) were evaluated using Control-FREEC with a window size of 500kb [28, 29] and visualized and further analyzed using custom scripts (MATLAB R2015a, The MathWorks Inc.). Exact details on the complete preprocessing pipeline and a list of the employed reference datasets can be downloaded online (Supplementary Data).

### **Statistical analysis**

We used unsupervised phylogenetic cluster analysis to investigate the clonal organization of the tumors. This analysis allows for the organization of the samples into clusters according to their similarity with each other and the reference (a genome without CNAs). Increasing distance between a cluster and the reference line implies higher differences between them. Biologically, larger differences can be interpreted as further evolutionary progression.

Unsupervised phylogenetic cluster analysis was performed on the CNA data with 500kb bins along the whole genome. The analysis of samples was done for each patient separately. First, tissue samples only were clustered in order to determine the presence of tumor subclones.

Support vector machine (SVM) analysis was performed to allocate CTCs within the cluster structure, obtained for the tumor tissue. However, SVM analysis is not able to define new clusters, but calculates possibility of a sample belonging to predefined classes. As consequence, analyzing samples can be classified within the fixed structure only, in contrast to phylogenetic cluster analysis, which creates as many clusters as necessary to reflect the differences between the samples. In order to reveal possible clusters formed by CTCs beyond the tissue clusters we checked whether CTCs and tissue samples form mixed or individual clusters. CNA data of the CTCs was added to the CNA data of tissue samples and unsupervised phylogenetic clustering was repeated.

SVM analysis was performed on a multiclass model on the k-nearest neighbor classifier template, created based on euclidean distances between the neighbors. The model was trained on the tissue data with the use of phylogenetic analysis defined clusters, subsequently, the CTCs were classified.

SVM analysis was performed on an introduced variability score (VS). VS was calculated as the sum of the fragment lengths (FL) multiplied by the square of the copy number (CN) value's difference from 2 according to the formula  $[VS = \sum_{i=1}^n (FL_i * (CN_i - 2)^2)]$

$|\text{CN}_i - 2|^2$ ], of each CTC along the whole genome. VS quantifies frequency and amplitude of the CNAs, which we used to reflect evolutionary progress.

## RESULTS

CTCs and FFPE primary and metastatic tissue samples from two enrolled metastatic breast cancer patients underwent WGA and NGS. The data were used for the investigation of the clonal organization of the breast cancer, progression and metastatic pathways. The complete data is present in Table 1.

### **Patient UKE243**

Patient UKE243 (1945-2012) was diagnosed with primary breast cancer of the right breast in 1992 and with collateral ER-positive and ERBB2-negative breast cancer of the left breast in 1999, and received endocrine treatment (aromatase inhibitor) in 2000-2005. The first metastasis (ER-positive, ERBB2-negative) was detected in 2009, at which the endocrine treatment with aromatase inhibitor (aromasin) was started. Due to further metastatic progress (2010, ER-positive) the treatment was switched to endocrine therapy with selective ER-modulator (fulvestrant), and in 2011 switched to chemotherapy (docetaxel) due to further metastatic progress. Blood for CTC analysis was collected during the course of chemotherapy in November 2011 (Figure 1A).

The blood sample analysis revealed the presence of 270 CTCs in 1 ml of blood with heterogeneous ER expression (64% ER-positive and 36% ER-negative CTCs). In total, 42 CTCs were picked by micromanipulation for downstream analysis. The FFPE material of the second primary tumor, diagnosed in 1999, was used for obtaining 50 tissue sections containing each 10-20 cells using laser microdissection: 40% ER-positive, 40% ER-negative, and 20% with unknown ER status (Table 1).

Unsupervised phylogenetic cluster analysis was performed on the CNA data from the primary tumor tissue; as a result, 5 clearly distinguishable clusters were formed (Figure 1B). Next, support vector machine (SVM) analysis was performed to allocate the CTCs to the identified tissue clusters. Subsequently, results obtained by the SVM analysis were proved by phylogenetic clustering of CNA data of the CTCs and tissue fragments together. The resulting phylogenetic tree contained mixed CTC-tissue clusters. Most of the CTCs were tackled by phylogenetic cluster analysis on combined data to the same tissue clusters as by SVM analysis.

Among 42 analyzed CTCs, 12 CTCs resided to the first cluster, 11 to the second, 18 to the third, and 1 to the fourth, no CTCs were allocated to the fifth cluster. ER expression was heterogeneous among tissue samples and CTCs within each cluster (Figure 1C). No association was found between ER expression and the phylogenetic tree structure.

Chromosome 1q and 16p gain and chromosome 9p loss were present in all identified tissue clusters and respective CTCs. This observation suggests these CNA changes being the very early genomic rearrangements in the patient's carcinogenesis.

Based on distances between the clusters of the phylogenetic tree, we combined the clusters into 2 groups: the first group included clusters 1-3 and the second group contained clusters 4-5. Fisher's exact test of the 2 groups revealed significantly different CNAs: chromosome 4q and 8p losses were significantly more frequent in clusters 1-3, whilst chromosome 8p gains were more frequent in clusters 4-5 (Figure 2). Because all CTCs except one resided to the tissue clusters 1-3, we compared the aberration frequencies between the two groups: tissue clusters 1-3 vs. CTCs (Figure 3). Significant differences were chromosomes 8q gain (tissue) and 1q and 7 gains and 16q and 22 losses (CTCs). Losses of chromosome 22 and 16q were found exclusively in CTCs.

Introduction of the VS as a measurement of the CNAs' intensity was used to demonstrate that evolution of the identified clones was associated with increase of the CNAs' frequency and amplitude, reflecting incensement of the cancer genome instability with further progression (Figure 4A).

Noteworthy, the earliest detected clone, represented by the first cluster, is an agglomeration of close situated measurement points, united into one cluster on the basis of small distance from each other (Figure 1B). We identified a subgroup different from the other members of the cluster 1 by plotting of VS against copy number level. VS and copy number level increased along the evolution vector from cluster 1 to cluster 5 (Figure 4B). However it can be seen that the subgroup members, representing 3 tissue sections and 3 CTCs, demonstrated copy number level <2. Loss of genetic material was dominant mechanism of CNAs in these samples in contrast to the rest of the samples. We explain this observation with following proposal. The earliest cancer cell population evolved stepwise as described before in agreement with parallel progression model towards lineages of cluster 1-3 and 4-5. This process included accumulation of both losses and gains of genomic material, despite gains and

amplifications were dominant, while a small subgroup experienced leap progression as per loss of genetic material.

Taken together, the evolutionary pathway of the disease could be schematically present as on Figure 1D: initial or very early chromosomal aberrations included chromosome 1q and 16p gains, and 9p loss. These events probably caused chromosomal instability, required for further clonal evolution and progress of cancer. Chromosomal instability could lead to the development of at least 2 cell lineages. One lineage evolved towards luminal subtype and gave rise to clones 1-3, depicted by clusters 1-3. These clones experienced further evolutionary progress encompassing gain of 8q after a number of cells had spread into the systemic circulation. These cells might have given rise to metastases after a certain dormancy period. CTCs, released from these metastatic lesions, reflect inherent CNAs from primary tumor clones, as well as CNAs of further evolution within metastatic lesion, like losses of chromosomes 16q and 22. Another lineage experienced further chromosomal aberrations, resulted in development of clones identified as clusters 4 and 5, characterized by high chromosomal instability.

Our results indicate that the metastases of the patient UKE243 arise from cells, disseminated from almost all subclones of the primary tumor, from the most earliest to very progressed ones. These findings are in line with parallel progression model of carcinogenesis and metastasis, suggesting that tumor cells acquire metastatic potential in the early stages of tumor progression.

### **Patient UKE008**

Patient UKE008 (born 1978) was diagnosed with primary metastatic breast cancer in 2013 with multiple metastases in the spine and pelvis. Palliative therapy included irradiation of the primary tumor and systemic chemotherapy (paclitaxel, April – August 2013) in combination with anti-ERBB2 therapy (Trastuzumab and Pertuzumab, April 2013 – December 2015). The blood samples were collected before any systemic treatment was applied (1st sample) and 3 months after completion of the chemotherapy (2nd sample) (Figure 5A).

We detected 2 ER-negative CTCs in 7.5 ml blood of the first blood sample, collected before therapy (0.27 CTCs/ml) and 20 ER-negative CTCs per 1 ml in the second blood sample. In total, 1 CTC from the 1<sup>st</sup> and 11 CTCs from the 2nd blood sample were collected for downstream analysis (Table 1).

The primary tumor as well as one of the metastases in the L4 spine segment were biopsied and formalin-fixed and paraffin-embedded. The tumor as well as metastasis were ER- and ERBB2-positive. Microdissected fragments of the primary tumor (n=6) and spine metastasis (n=5) were ER-positive in 50% and 40% of cases, respectively.

Unsupervised phylogenetic cluster analysis of the tissue data only was performed. Because the patient was diagnosed with primary metastatic breast cancer, cluster analysis was performed on the combined data obtained from the primary tumor and metastasis. The obtained phylogenetic tree demonstrated the presence of 3 clearly distinguishable clusters. Subsequent SVM analysis tackled all CTCs (n=12) to the third tissue cluster, whereas phylogenetic cluster analysis of the combined CTC and tissue data demonstrated 1 distinct CTC cluster in addition to the 3 previously identified tissue clusters (Figure 5B). This discrepancy is explainable by the difference between the phylogenetic cluster analysis and the SVM. The cluster analysis identifies as many clusters as necessary according to the differences between the samples, whereas SVM analysis is not able to define new clusters. Taking this explanation into consideration, phylogenetic tree built on combined CTC and tissue data was considered as reflecting clonal organization the best: 3 distinct tissue clusters and 1 CTC cluster were identified. CTCs demonstrated highest similarity with the most progressed clone identified in the primary tumor and metastasis, but presumably did not arise directly from descendants of the clone.

The tumor subclone represented in the first cluster contained data obtained from 2 fragments of the primary tumor. The second cluster (represented by data of the metastasis only) might be considered an intermediate evolutionary step towards cluster 3. The third cluster, representing the most progressed evolutionary step, was made up of data from both primary tumor's and metastasis' tissue fragments. These results indicate that metastatic outgrowth could be initiated by collective dissemination of tumor cells from the 2 cooperating clones within a CTC cluster (Figure 5C). However, it cannot be excluded that cells from primary tumor clones disseminated not in a CTC-cluster, but as individual CTCs, arrived at the same distant location and cooperated there. Investigation of further metastatic lesions is needed to clarify mechanisms of metastasis-initiating dissemination in the patient.

Evolutionary history of the UKE008 patient's cancer might have been as follows (Figure 5D): chromosome 17p loss and chromosome 17q and 19q gain might be initial

or very early events in the carcinogenesis because these CNAs we identified in frequency plots of all the identified clusters (Figure 6). Later during carcinogenesis this early cancerous cell population branched in its evolution. One subclone experienced chromosome 4q loss and chromosome 6 gain and developed the clone, depicted by cluster 1. Possibly lineage, represented by clusters 2-3, originated from another branch. Further evolution of the lineage led via chromosome 1q, 8q and 11p gain and chromosome 11q loss towards the second clone (cluster 2), and additional gain of chromosome 7q resulted in cell clone, depicted by cluster 3. Cells from these cooperating clones disseminated either as CTC-cluster or as individual cells and built up distant metastasis we investigated, which is therefore reflecting the clonal structure of the primary tumor.

This scenario does not answer the question where the CTCs came from: the primary tumor or the metastasis. However, based on the phylogenetic tree, CTCs did not reside to any of tissue clusters, but formed separate clusters, not presented in the structure of the primary tumor and metastatic tissue (Figure 4B). Moreover, scatter plots of variability scores versus the number of genome-wide break points and copy number level (Figure 7A, 7B) demonstrated increase of genomic aberrations with the evolution from cluster 1 to cluster 4 with the maximum score among CTCs. The fact that the patient demonstrated multiple metastases suggests that CTCs of the patient UKE008 arise from the metastasis we did not investigate. In this case, the uninvestigated metastases embody further steps in evolutionary progression of the cancer in line with the linear progression model.

## DISCUSSION

In the study presented here we shed light on clonal evolution of the human breast cancer. The two investigated patients demonstrated different ways of clonal evolution of cancer towards tumor cell dissemination.

Clonality and evolution of the cancer can be investigated on single cell level with the use of primary tumors, metastases, and/or CTCs. Primary tumors are removed or biopsied in the majority of cases, delivering material for investigation. Administration of systemic therapy is usually based on characteristics of the primary tumor. However, the metastases may not resemble the primary tumor anymore due to genetic progression or selection of treatment-resistant clones. The differences between primary tumor and metastases might be the reason for treatment failure. Therefore

CTCs as “liquid biopsy” provide a unique, easy accessible source of tumor material [30].

CTCs that can be detected in the blood circulation many years after removal of the primary tumor are most likely coming from the metastases, because the half-life time of CTCs in circulation is <2.4 hours [31], despite dormancy cannot be ruled out completely. It has been shown that dormant tumor cells in bone marrow may sometimes divide into micrometastases, which release CTCs, and thus cause the presence of CTCs in blood of metastases-free breast cancer patients many years after mastectomy, but in small concentrations ( $\leq 1$ CTC/ml) [31], which is in contrast with 270 and 20 CTC/ml found by us in blood of the UKE243 and UKE008 patients, respectively. These findings suggest that CTCs detected in blood of the enrolled in our study patients arise from metastases present in the body at the time point of blood sampling.

The results obtained from patient UKE243 suggests a parallel progression of the breast tumor. CTCs were detected in the blood of patient UKE243 12 years after the primary tumor was removed. Based on bioinformatics analysis all CTC resided to 4 out of 5 phylogenetic clusters identified in data of the primary tumor (Figure 1C). These results suggest that metastases might have been founded by tumor cells that disseminated from multiple subclones of the primary tumor. In consideration of the time gap between primary tumor removal and detection of the first metastasis (10 years), it is likely that disseminated tumor cells underwent dormancy for a certain period before giving rise to distant metastases.

One question which may arise is whether metastases and CTCs of the patient UKE243 originate from the first primary tumor, diagnosed in 1992, or from the second contralateral primary tumor, diagnosed in 1999. The later tumor only was available for our analysis. According to the histology of both primaries, the metastases corresponded to the second primary tumor, which can be confirmed by our cluster analysis. Nevertheless, we cannot exclude the possibility that metastases and subsequent CTCs originate from the first primary tumor. However, in this case both primary tumors had similar clonal structure.

Additionally, primary tumor as well as CTCs of patient UKE243 demonstrated heterogeneous ER expression (Figure 1C). Outgrowth of further ER-positive metastases and presence of ER-positive CTCs after the completion of endocrine therapy suggests endocrine therapy failure in this patient. Since we did not find mutations in ER-coding gene (*ESR1*) in the CTCs of the patient [22], endocrine therapy

failure might have been caused by other mechanisms, e.g. epigenetic mechanisms or a dysfunctional ER-pathway.

Based on the observed CNA frequencies in identified clusters and CTCs, we conclude the existence of at least two lineages of tumor cells in the primary tumor of the patient UKE243. One of the lineages, presented by clones 1-3 (cluster 1-3), is characterized by chromosome 1q and 16p gain. A genomic signature 1q+/16p+/16q- was revealed in CTCs of the patient. CNA profile 1q+/16p+/16q- is associated with ER-positivity, luminal gene expression pattern, moderate to high differentiated tumors, and better outcome [32]. However, chromosome 16q loss was not observed in clones of primary tumor (Figure 3, 4) and therefore appeared at later stages of the metastatic process. Loss of chromosome 16q, if not appeared as early genomic effect as in luminal A tumors, might be produced due to genomic instability. This mechanism of 16q loss has been observed in luminal B tumors [33]. Another evidence for genomic instability of the patient's disease is increasing frequency of CNAs along the evolution vector (Figure 4).

The second lineage, depicted by clusters 4 and 5, exhibited CNA patterns typical for basal-like breast cancer (high frequency of narrow low-amplitude gains in losses) [32]. Basal-like subtype, typically ER-negative, is characterized by higher chromosomal instability than luminal subtypes [34]. Moreover, our results suggest that the clonal split happened at a very early stage of carcinogenesis. One of the lineages experienced further luminal-like differentiation, whereas the second lineage retained basal-like characteristics.

Presence of both basal- and luminal-like cell lineages in breast tumors has been demonstrated by others [35-37]. One of the possible explanations of the coexistence of basal- and luminal-like cells within a tumor can be given through the hypothesis that ER-positive cells, e.g. cells of luminal B subtype, and basal-like cells may arise from the same bipotent progenitor cell [14, 33]. Moreover, recently Cleary *et al.* demonstrated cooperation between basal- and luminal-like subclones playing a role in tumor maintenance [38].

Li *et al.* demonstrated in a mouse model that activation of Wnt signaling pathway transforms mammary progenitor cells, promoting heterogeneity of outgrowing cell lineages. The authors conclude that basal- and luminal-like lineages within the same tumor supposedly derive from a bipotent malignant progenitor cell [37]. Mammary

progenitor cells are typically ER-negative, but originating lineages might undergo luminal-like differentiation and become ER-positive [14, 33, 39].

In contrast to patient UKE243, our data obtained from the cancer from patient UKE008, suggest linear progression to metastases. According to the linear progression model, distant metastases originated from cells, disseminated from the primary tumor at late evolutionary stage(s).

Loss of chromosome 17p and gain of 17q were observed in frequency plots of all clusters identified on data of the patient UKE243, including the CTC cluster (Figure 6). Gain of chromosome 17q is typical for ERBB2-positive luminal B breast cancers (rev. in [32]), which is in agreement with the pathology report (ER-positive, ERBB2-positive tumor). Loss of chromosome 17p is a common aberration in many cancers, including breast cancer [40], due to the location of tumor suppressor gene *TP53*.

We demonstrated that at least one distant metastasis carries the genomic signatures observed in the two clones of the primary tumor (Figure 5C). This observation might have two explanations. First possible scenario suggests collective dissemination of tumor cells from these two clones within a mixed CTC-cluster. Another explanation implies individual dissemination of the tumor cells from the two clones and subsequent cooperation at distant site. Whichever dissemination way took place, our results indicate interaction of the two tumor clones. The two genetically similar clones of the primary tumor might have interacted to obtain a selective growth advantage and/or metastatic propensity.

Evidence for cooperating clones can be found in mouse and fruit fly models. It has been shown that two cell populations can interact to promote tumorigenesis and obtain the ability to metastasize [41-43]. Moreover, interclonal cooperation contributes to tumor growth and progression [44]. Tumor cells from cooperating clones might disseminate collectively by formation of CTC-clusters. CTC clusters demonstrate an increased metastatic capacity in comparison to single CTCs [45].

A 74-fold increase of the amount of CTCs in 1 ml blood of patient UKE008 was found in comparison to baseline before therapy, 3 months after completion of chemotherapy but still under anti-ERBB2 therapy. All CTCs were found to be ER-negative, whereas the primary tumor and the metastasis were ER-positive. It has been shown that ER activity provides a way for ER-positive ERBB2-positive cells to escape ERBB2-targeted therapy [46]. Co-expression of ERBB2 and ER has been found in

breast cancer patients patients [47-49]. The data suggests that ER-negative ERBB2-negative cells could escape therapy.

Although all CTCs identified in blood samples of patient UKE008 were ER-negative, it cannot be excluded that ER-positive CTCs were still present in the body, but could not be detected. The cells might have been escaped to bone marrow and underwent dormancy. Alternatively, EMT-associated downregulation of epithelial markers on the cell surface might have hampered detection of these cells.

Intra-tumor heterogeneity of breast cancer is a results of clonal expansion. In order to reconstruct cancer evolution and clonal organization on single cell level [50] one should assume that the tumor at any moment of the evolution contains all previous clones, or at least the most crucial ones. However, this assumption contradicts the Darwinian theory applied to carcinogenesis [51, 52]. According to the theory inter-clonal competition should lead to outcompeting of particular, not necessarily less aggressive clones [44], resulting in secondary mono- or oligoclonal structure of the primarily polyclonal tumor. As consequence the reconstructed clonal structure of the primary tumor does not necessarily reproduce cancer evolution.

Alternative look from an ecological perspective suggests that subclones within a tumor can be seen as individual units interacting with each other and their environment. This theory implies that not only competition, but other types of interaction, e.g. cooperation, are possible (rev. in [15]) and finds confirmation in cancer model systems [41-44]. Consequently, different cancer clones are not necessarily overgrown by one dominant clone, and poly- or oligoclonal structure of cancer can be revealed.

The results obtained in our study of the breast cancer clonality demonstrated oligoclonal structure of the investigated breast tumors, indicating that both mechanisms, competition and cooperation of tumor clones, are likely being involved in cancer evolution. Additionally, the results indicate that breast cancer might utilize both linear and parallel progression ways.

Technical obstacles can also hamper clonal analysis. Since comprehensive investigation of every single cell of the complete tumor is hardly possible, underrepresentation of certain clones as well as overrepresentation of other clones in a study cohort may lead to false reconstruction of tumor's clonal structure. Additionally, metastases are a difficult subject for clonal investigation. Distant metastases can be detected first when they reach a certain size, and are infrequently biopsied.

Nevertheless, our results demonstrate the feasibility of archival material accompanied with CTCs for investigation of clonal evolution of human breast cancer. Further research is needed to obtain information about the genetic heterogeneity of the metastases and possible identification of therapy sensitive and resistant clones.

## CONCLUSION

The investigated breast cancer cases represent parallel and linear metastases progression model (the UKE243 and UKE008 patient, respectively). Our results demonstrate that therapy resistant breast cancer metastases detected years after primary tumor removal may originate from tumor clones present at early and late stages of primary tumor carcinogenesis (the patient UKE243). Alternatively, metastasis in primary metastatic breast cancer originates according to the linear progression model from late interacting clones of the primary tumor, and CTCs most probably resemble further metastases, not resembling the primary tumor anymore (the patient UKE008). These results underline the importance of “liquid biopsy” for companion diagnostics in metastatic breast cancer.

## REFERENCES

1. Wang, D. and S. Bodovitz, *Single cell analysis: the new frontier in 'omics'*. Trends Biotechnol, 2010. 28(6): p. 281-90.
2. Chambers, A.F., A.C. Groom, and I.C. MacDonald, *Dissemination and growth of cancer cells in metastatic sites*. Nat Rev Cancer, 2002. 2(8): p. 563-72.
3. Fidler, I.J., *The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited*. Nat Rev Cancer, 2003. 3(6): p. 453-8.
4. van 't Veer, L.J., et al., *Gene expression profiling predicts clinical outcome of breast cancer*. Nature, 2002. 415(6871): p. 530-6.
5. Ramaswamy, S., et al., *A molecular signature of metastasis in primary solid tumors*. Nat Genet, 2003. 33(1): p. 49-54.
6. Klein, C.A., *Parallel progression of primary tumours and metastases*. Nat Rev Cancer, 2009. 9(4): p. 302-12.
7. Husemann, Y., et al., *Systemic spread is an early step in breast cancer*. Cancer Cell, 2008. 13(1): p. 58-68.
8. Zhang, L., et al., *Meta-analysis of the prognostic value of circulating tumor cells in breast cancer*. Clin Cancer Res, 2012. 18(20): p. 5701-10.
9. Joosse, S.A. and K. Pantel, *Genetic traits for hematogeneous tumor cell dissemination in cancer patients*. Cancer Metastasis Rev, 2016.
10. Rossi, S., et al., *Hormone Receptor Status and HER2 Expression in Primary Breast Cancer Compared With Synchronous Axillary Metastases or Recurrent Metastatic Disease*. Clin Breast Cancer, 2015. 15(5): p. 307-12.
11. Yang, Y.F., et al., *Discordances in ER, PR and HER2 receptors between primary and recurrent/metastatic lesions and their impact on survival in breast cancer patients*. Med Oncol, 2014. 31(10): p. 214.
12. Kimbung, S., N. Loman, and I. Hedenfalk, *Clinical and molecular complexity of breast cancer metastases*. Semin Cancer Biol, 2015. 35: p. 85-95.
13. Koren, S. and M. Bentières-Alj, *Breast Tumor Heterogeneity: Source of Fitness, Hurdle for Therapy*. Mol Cell, 2015. 60(4): p. 537-46.
14. Skibinski, A. and C. Kuperwasser, *The origin of breast tumor heterogeneity*. Oncogene, 2015. 34(42): p. 5309-16.
15. Marusyk, A. and K. Polyak, *Tumor heterogeneity: causes and consequences*. Biochim Biophys Acta, 2010. 1805(1): p. 105-17.

16. Vanharanta, S. and J. Massague, *Origins of metastatic traits*. *Cancer Cell*, 2013. 24(4): p. 410-21.
17. Joosse, S.A., T.M. Gorges, and K. Pantel, *Biology, detection, and clinical implications of circulating tumor cells*. *EMBO Mol Med*, 2015. 7(1): p. 1-11.
18. Burgio, E. and L. Migliore, *Towards a systemic paradigm in carcinogenesis: linking epigenetics and genetics*. *Mol Biol Rep*, 2015. 42(4): p. 777-90.
19. Vineis, P., A. Schatzkin, and J.D. Potter, *Models of carcinogenesis: an overview*. *Carcinogenesis*, 2010. 31(10): p. 1703-9.
20. Feinberg, A.P., *The epigenetics of cancer etiology*. *Seminars in Cancer Biology*, 2004. 14(6): p. 427-432.
21. Ciriello, G., et al., *Emerging landscape of oncogenic signatures across human cancers*. *Nat Genet*, 2013. 45(10): p. 1127-33.
22. Babayan, A., et al., *Heterogeneity of estrogen receptor expression in circulating tumor cells from metastatic breast cancer patients*. *PLoS One*, 2013. 8(9): p. e75038.
23. Joosse, S.A., et al., *Prediction of BRCA2-association in hereditary breast carcinomas using array-CGH*. *Breast Cancer Res Treat*, 2012. 132(2): p. 379-89.
24. Li, H. *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. 2013.
25. Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. *Bioinformatics*, 2009. 25(16): p. 2078-9.
26. *Picard tools*. Available from: <http://picard.sourceforge.net>.
27. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data*. *Genome Res*, 2010. 20(9): p. 1297-303.
28. Boeva, V., et al., *Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data*. *Bioinformatics*, 2012. 28(3): p. 423-5.
29. Boeva, V., et al., *Control-free calling of copy number alterations in deep-sequencing data using GC-content normalization*. *Bioinformatics*, 2011. 27(2): p. 268-9.
30. Alix-Panabieres, C. and K. Pantel, *Circulating tumor cells: liquid biopsy of cancer*. *Clin.Chem.*, 2013. 59(1): p. 110-118.

31. Meng, S., et al., *Circulating tumor cells in patients with breast cancer dormancy*. Clin Cancer Res, 2004. 10(24): p. 8152-62.
32. Kwei, K.A., et al., *Genomic instability in breast cancer: pathogenesis and clinical implications*. Mol Oncol, 2010. 4(3): p. 255-66.
33. Melchor, L. and J. Benitez, *An integrative hypothesis about the origin and development of sporadic and familial breast cancer subtypes*. Carcinogenesis, 2008. 29(8): p. 1475-82.
34. Pikor, L., et al., *The detection and implication of genome instability in cancer*. Cancer Metastasis Rev, 2013. 32(3-4): p. 341-52.
35. Knox, A.J., et al., *Modeling luminal breast cancer heterogeneity: combination therapy to suppress a hormone receptor-negative, cytokeratin 5-positive subpopulation in luminal disease*. Breast Cancer Res, 2014. 16(4): p. 418.
36. Haughian, J.M., et al., *Maintenance of hormone responsiveness in luminal breast cancers by suppression of Notch*. Proc Natl Acad Sci U S A, 2012. 109(8): p. 2742-7.
37. Li, Y., et al., *Evidence that transgenes encoding components of the Wnt signaling pathway preferentially induce mammary cancers from progenitor cells*. Proc Natl Acad Sci U S A, 2003. 100(26): p. 15853-8.
38. Cleary, A.S., et al., *Tumour cell heterogeneity maintained by cooperating subclones in Wnt-driven mammary cancers*. Nature, 2014. 508(7494): p. 113-7.
39. Bhagirath, D., et al., *Cell type of origin as well as genetic alterations contribute to breast cancer phenotypes*. Oncotarget, 2015. 6(11): p. 9018-30.
40. Gao, Y., et al., *Genetic changes at specific stages of breast cancer progression detected by comparative genomic hybridization*. J Mol Med (Berl), 2009. 87(2): p. 145-52.
41. Calbo, J., et al., *A functional role for tumor cell heterogeneity in a mouse model of small cell lung cancer*. Cancer Cell, 2011. 19(2): p. 244-56.
42. Ohsawa, S., et al., *Mitochondrial defect drives non-autonomous tumour progression through Hippo signalling in Drosophila*. Nature, 2012. 490(7421): p. 547-51.
43. Wu, M., J.C. Pastor-Pareja, and T. Xu, *Interaction between Ras(V12) and scribbled clones induces tumour growth and invasion*. Nature, 2010. 463(7280): p. 545-8.

44. Marusyk, A., et al., *Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity*. Nature, 2014. 514(7520): p. 54-8.
45. Aceto, N., et al., *Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis*. Cell, 2014. 158(5): p. 1110-22.
46. Wang, Y.C., et al., *Different mechanisms for resistance to trastuzumab versus lapatinib in HER2-positive breast cancers--role of estrogen receptor and HER2 reactivation*. Breast Cancer Res, 2011. 13(6): p. R121.
47. Pinhel, I., et al., *ER and HER2 expression are positively correlated in HER2 non-overexpressing breast cancer*. Breast Cancer Res, 2012. 14(2): p. R46.
48. Konecny, G., et al., *Quantitative association between HER-2/neu and steroid hormone receptors in hormone receptor-positive primary breast cancer*. J Natl Cancer Inst, 2003. 95(2): p. 142-53.
49. Harigopal, M., et al., *Multiplexed assessment of the Southwest Oncology Group-directed Intergroup Breast Cancer Trial S9313 by AQUA shows that both high and low levels of HER2 are associated with poor outcome*. Am J Pathol, 2010. 176(4): p. 1639-47.
50. Navin, N., et al., *Tumour evolution inferred by single-cell sequencing*. Nature, 2011. 472(7341): p. 90-4.
51. Nowell, P.C., *The clonal evolution of tumor cell populations*. Science, 1976. 194(4260): p. 23-8.
52. Greaves, M. and C.C. Maley, *Clonal evolution in cancer*. Nature, 2012. 481(7381): p. 306-13.

|  |   | UKE243              | UKE008               |                           |
|--|---|---------------------|----------------------|---------------------------|
| Characteristics of blood sample and CTCs | Time point of blood examination           | During chemotherapy | Before any therapy   | During anti-ErbB2 therapy |
|  | Number of CTCs per 1 ml of blood          | 270                 | 0.27                 | 20                        |
|  | ER-positive CTCs                          | 64%                 | 0                    | 0                         |
|  | ER-negative CTCs                          | 36%                 | 100%                 | 100%                      |
|  | Number of CTCs available for the analysis | 42                  | 1                    | 11                        |
|  | ER-positive CTCs                          | 16                  | 0                    | 0                         |
|  | ER-negative CTCs                          | 19                  | 1                    | 11                        |
|  | ER status unknown                         | 7                   | 0                    | 0                         |
| Characteristics of tumor tissue          | Tissue available for the analysis         | Primary tumor       | Primary tumor        | Metastasis                |
|  |   | 2 FFPE blocks       | Biopsy FFPE material | Biopsy FFPE material      |
|  | ER status                                 | positive            | positive             | positive                  |
|  | ErbB2 status                              | negative            | positive             |                           |
|  | Number of tissue fragments dissected      | 50                  | 6                    | 5                         |
|  | ER-positive fragments                     | 20                  | 3                    | 2                         |
|  | ER-negative fragments                     | 20                  | 3                    | 3                         |
|  | ER status unknown                         | 10                  | 0                    | 0                         |

Table 1. Characteristics of patient material and data available.

FFPE – formalin-fixed, paraffin-embedded.

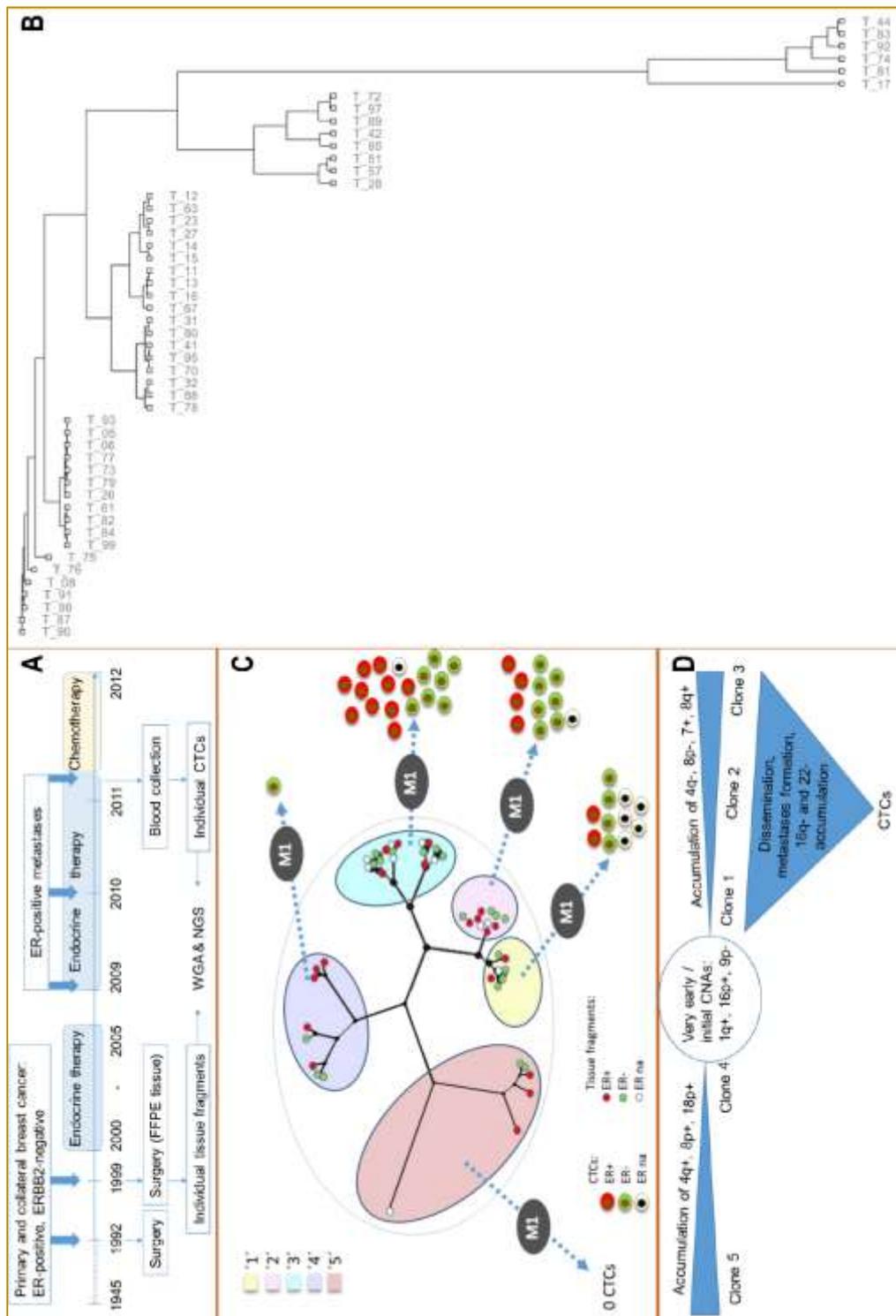


Figure 1. Evolutional history of the cancer disease of patient UKE243.

- Timeline of breast cancer disease and cancer therapy in patient UKE243
- Phylogenetic tree resulted from unsupervised phylogenetic cluster analysis of the tissue samples obtained from the primary tumor of the patient UKE243.
- Reconstruction of the primary tumor's structure and CTCs' origin.
- Schematic carcinogenesis of the primary tumor and metastases of UKE243 patient. Description in text.

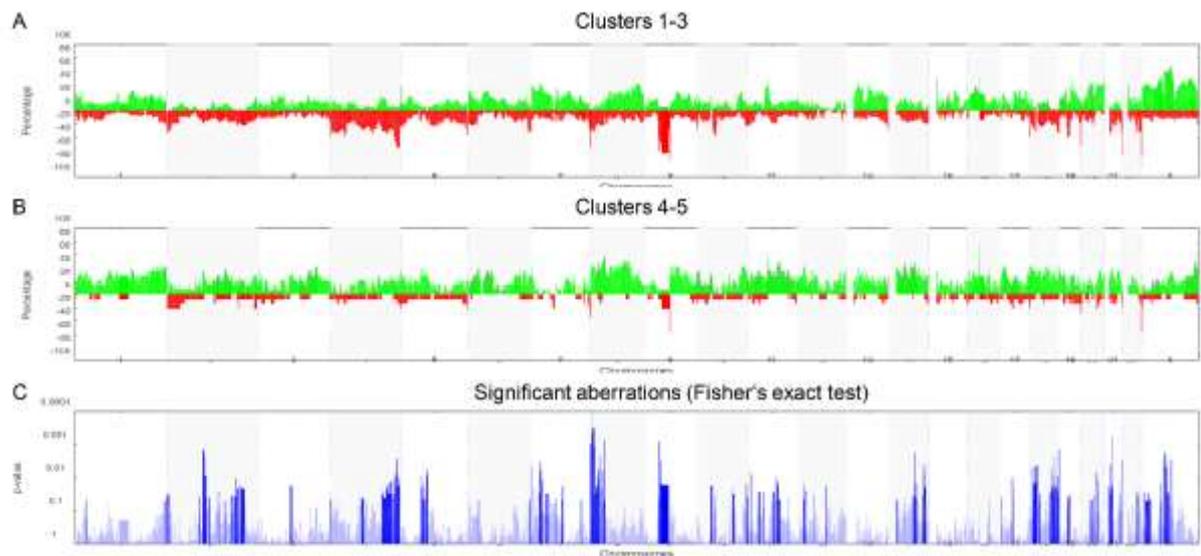


Figure 2. Frequency plots of tissue fragments from combined clusters 1-3 and 4-5 of patient UKE243.

Green regions demonstrate gain of genetic material, red – losses, along the chromosome from 1 to X (x axis).

- A. Tissue clusters 1-3.
- B. Tissue clusters 4-5.
- C. Plot of statistically significant differences between the compared groups by Fisher's exact test. Dark blue regions demonstrate significant differences with the threshold  $p < 0.05$ .

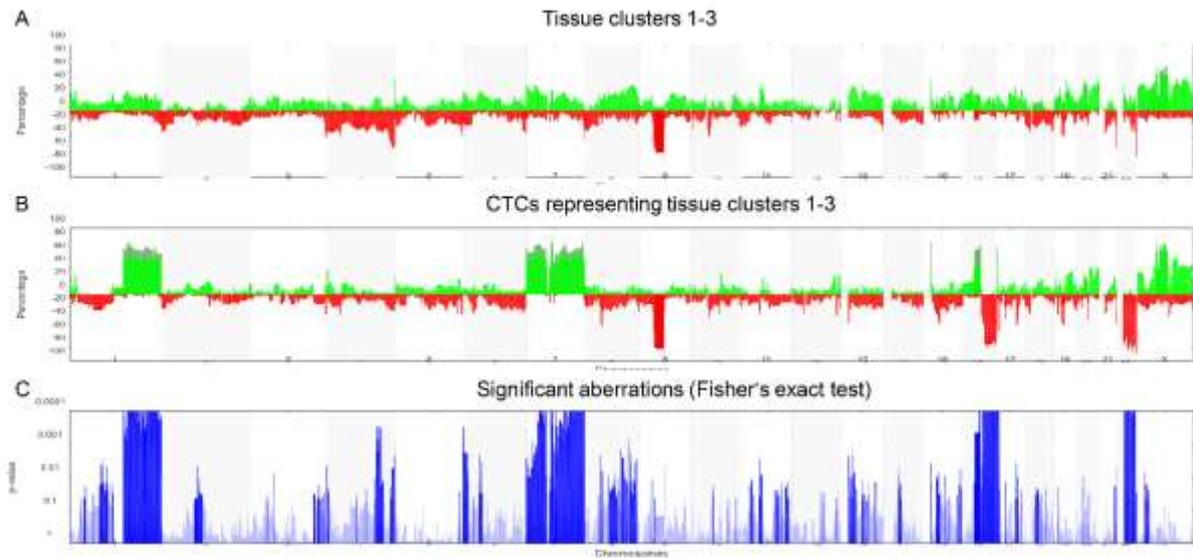


Figure 3. Frequency plots of tissue fragments and CTCs from clusters 1-3 of patient UKE243.

Green regions demonstrate gain of genetic material, red – losses, along the chromosome from 1 to X (x axis) in groups of samples.

- A. Tissue clusters 1-3.
- B. CTCs representing tissue clusters 1-3.
- C. Plot of statistically significant differences between the compared groups by Fisher's exact test. Dark blue regions demonstrate significant differences with the threshold  $p < 0.05$ .

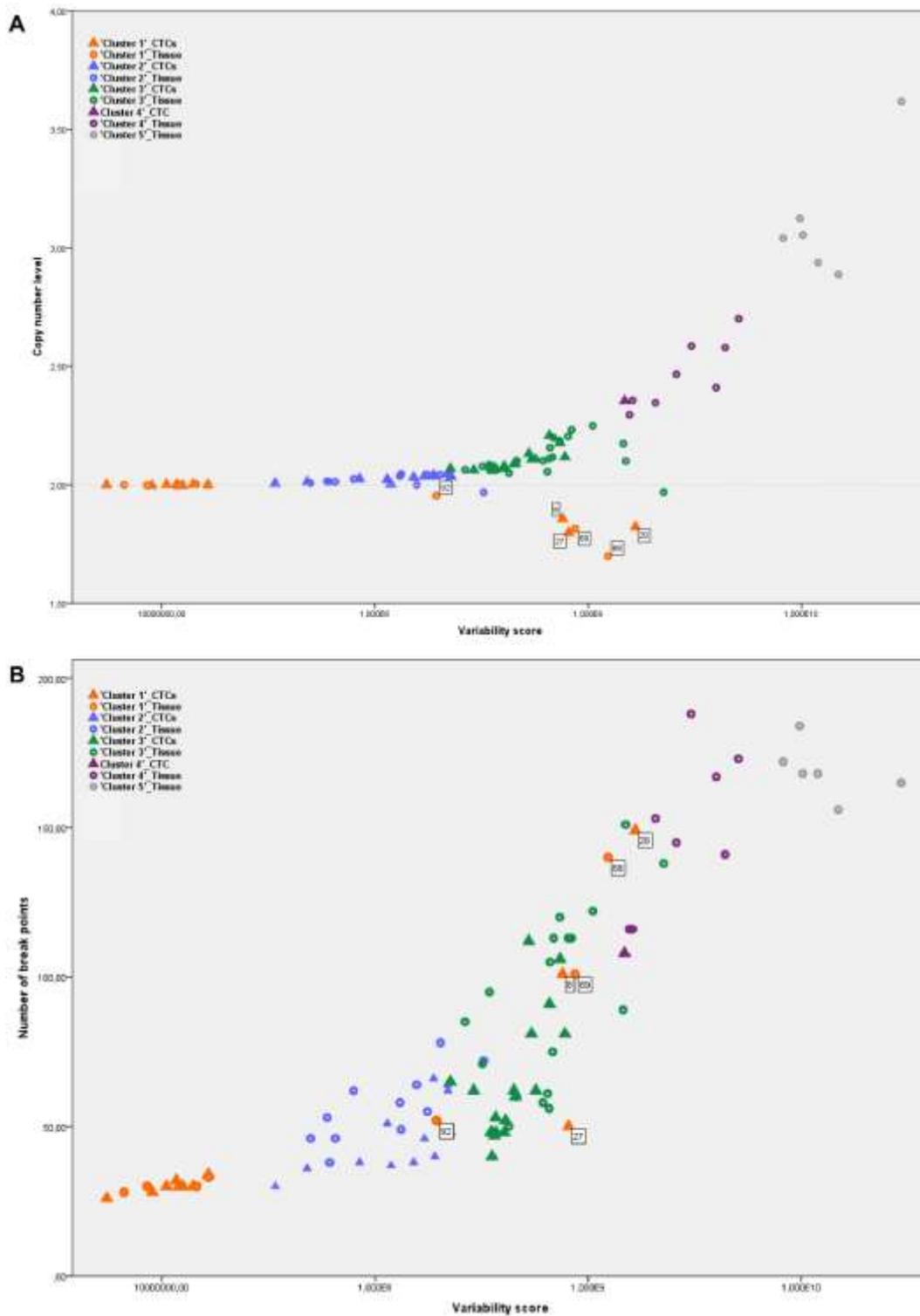


Figure 4.

- A. Plot of the variability score against number of break points in the CTCs and tissue samples of the patient UKE243.
- B. Numbers represent sample ID and are presented for the subgroup of the cluster 1.
- C. Plot of the variability score against average CNA value in the CTCs and tissue samples of the patient UKE243.
- D. Numbers represent sample ID and are presented for the subgroup of the cluster 1

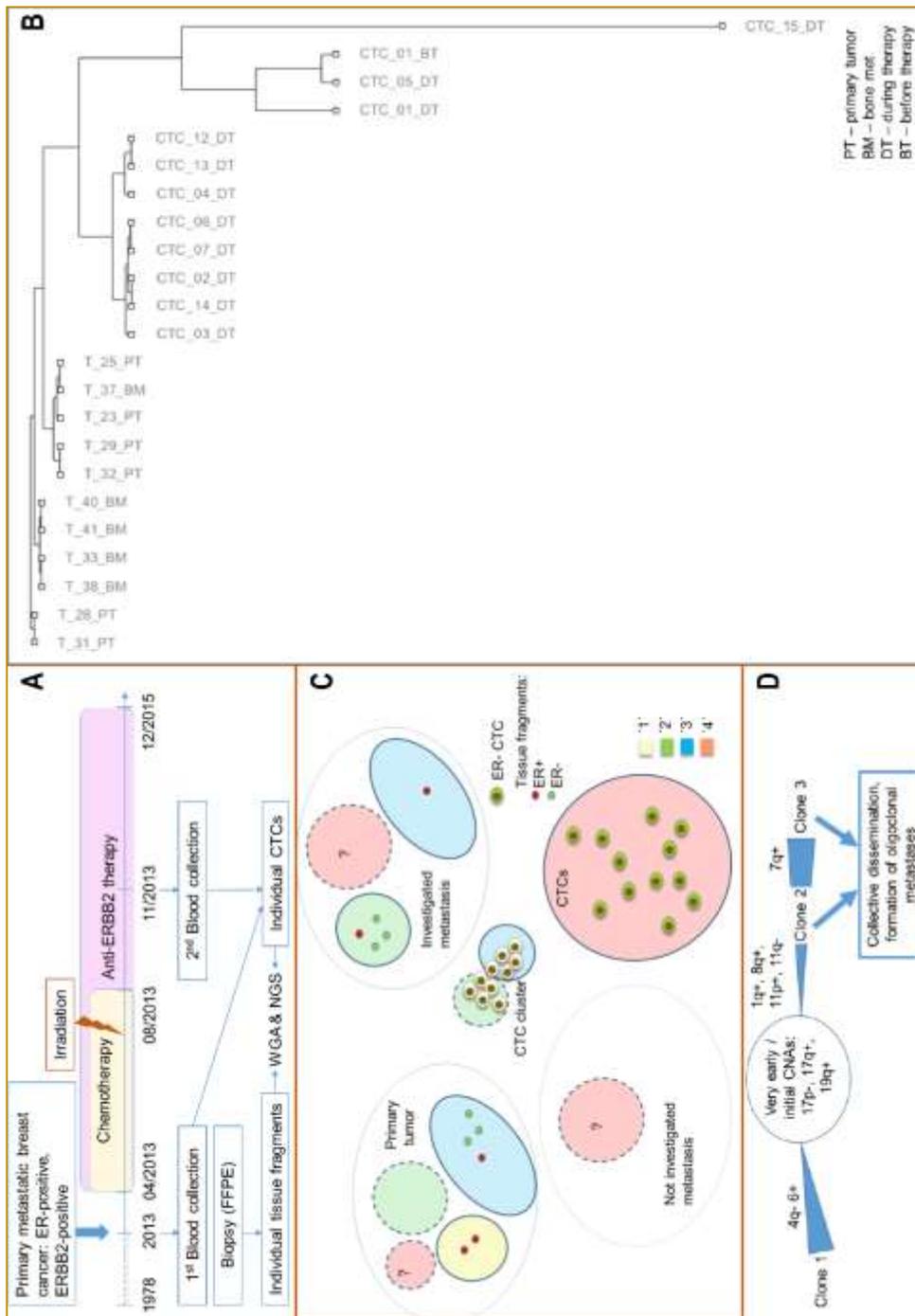


Figure 5. Evolutional history of the cancer disease of patient UKE008.

- Timeline of breast cancer disease and cancer therapy in patient UKE008.
- Phylogenetic tree resulted from unsupervised phylogenetic cluster analysis of the CTCs and tissue samples obtained from the primary and metastatic tumor of patient UKE008.
- Reconstruction of the primary and metastatic tumor's structure and CTCs' origin.
- Schematic carcinogenesis of the primary tumor and metastases of UKE008 patient. Description in text.

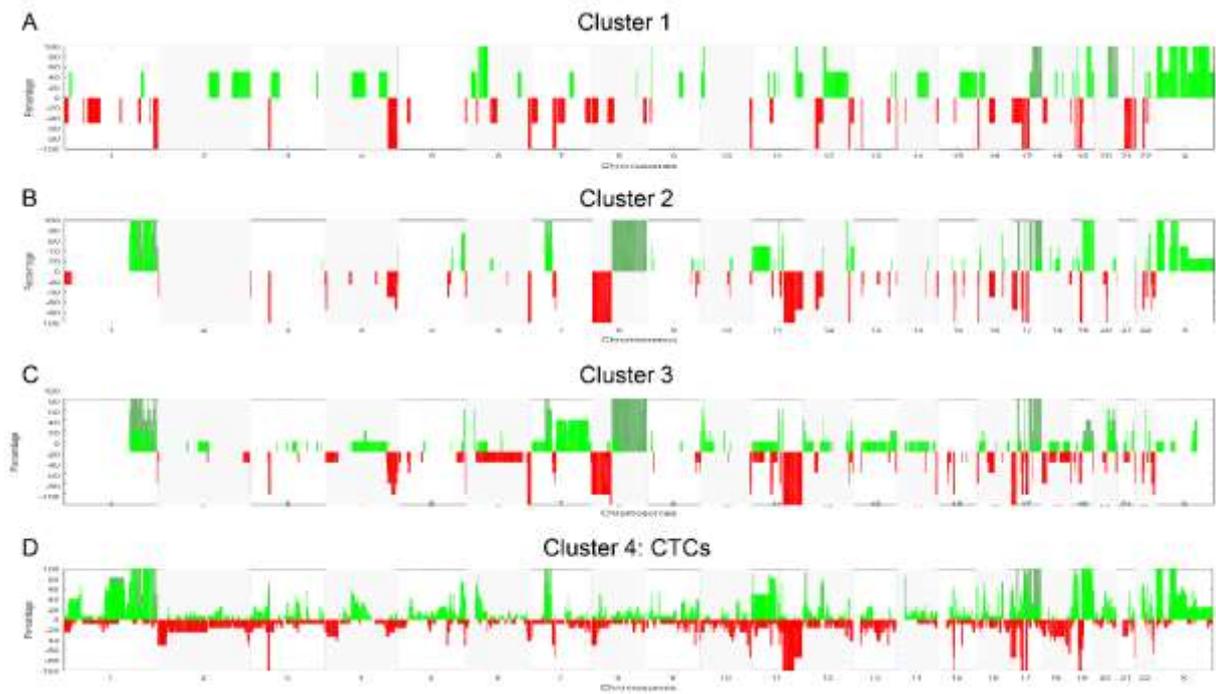


Figure 6. CNA profiles of clusters 1-4, combining tissue samples and CTCs of the patient UKE008.

Green regions demonstrate gain of genetic material, red – losses, along the chromosome from 1 to X (x axis) in groups of samples.

- A. Tissue cluster 1.
- B. Tissue cluster 2.
- C. Tissue cluster 3.
- D. CTC cluster.

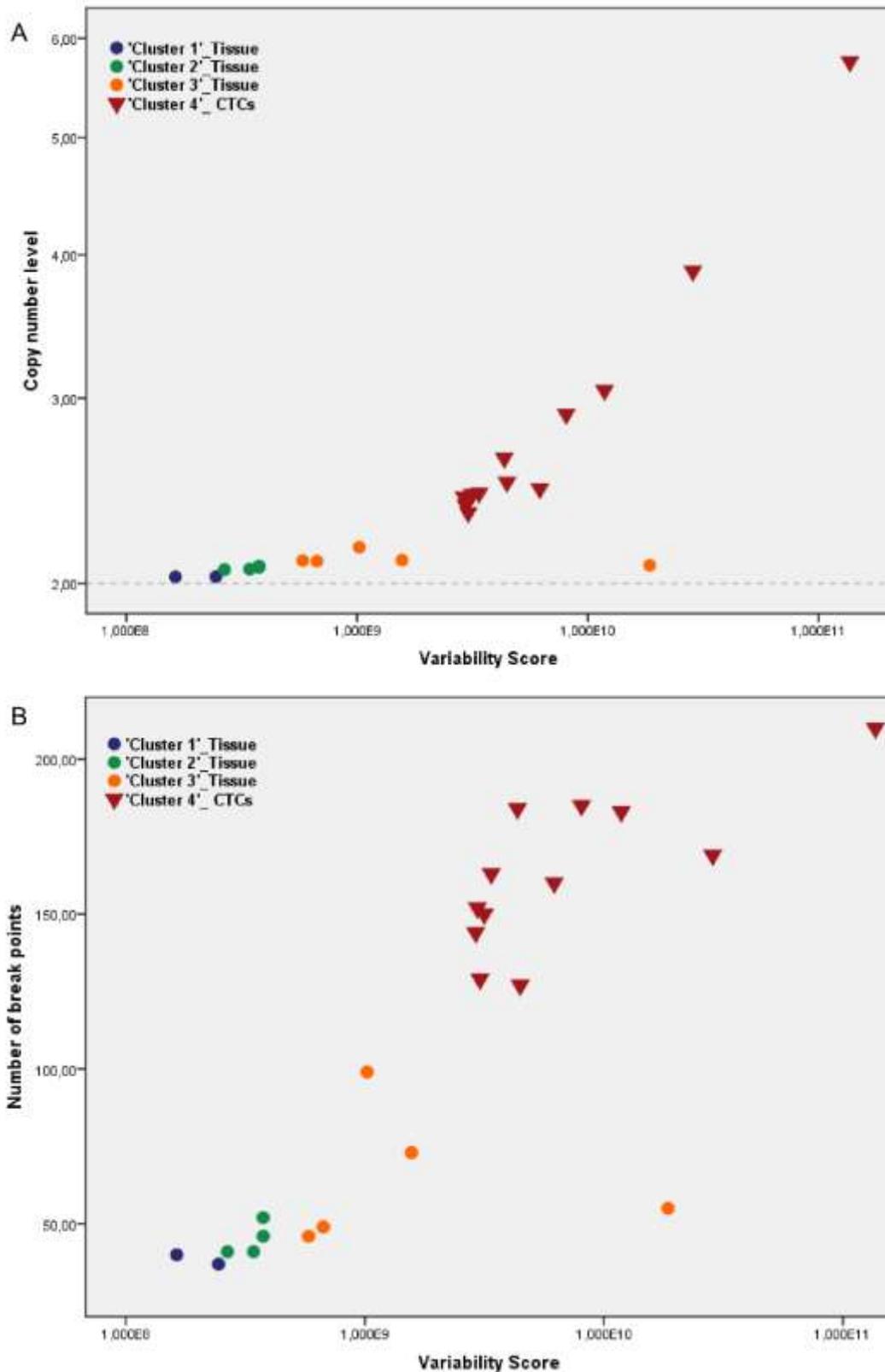


Figure 7.

- A. Plot of the VS against number of break points in the CTCs and tissue samples of the patient UKE008.
- B. Plot of the VS against average CNA value in the CTCs and tissue samples of the patient UKE008.

## 7. PUBLICATION 4

### RHAMM splice variants confer radiosensitivity in human breast cancer cell lines

Alexandra Schütze<sup>1</sup>, Christian Vogeley<sup>1</sup>, Tobias Gorges<sup>2</sup>, Sören Twarock<sup>1</sup>, Jonas Butschan<sup>1</sup>, Anna Babayan<sup>2</sup>, Diana Klein<sup>3</sup>, Shirley K. Knauer<sup>4</sup>, Eric Metzen<sup>5</sup>, Volkmar Müller<sup>6</sup>, Verena Jendrossek<sup>3</sup>, Klaus Pantel<sup>2</sup>, Karin Milde-Langosch<sup>6</sup>, Jens W. Fischer<sup>1\*</sup>, and Katharina Röck<sup>1\*</sup>

<sup>1</sup>Institut für Pharmakologie und Klinische Pharmakologie, Universitätsklinikum der Heinrich-Heine-Universität, Düsseldorf, Germany

<sup>2</sup>Department of Tumor Biology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

<sup>3</sup>Institute of Cell Biology (Cancer Research), University Hospital, University of Duisburg-Essen , Essen, Germany

<sup>4</sup>Institute for Molecular Biology II, Centre for Medical Biotechnology (ZMB), University of Duisburg-Essen, Essen, Germany

<sup>5</sup>Institute of Physiology, Faculty of Medicine, University Duisburg-Essen, Essen, Germany

<sup>6</sup>Department of Gynecology, University Hospital Hamburg-Eppendorf, Hamburg, Germany

\*contributed equally

Published: Oncotarget. 2016 Feb 8. doi: 10.18632/oncotarget.7258. [Epub ahead of print]

## RHAMM splice variants confer radiosensitivity in human breast cancer cell lines

Alexandra Schütze<sup>1</sup>, Christian Vogeley<sup>1</sup>, Tobias Gorges<sup>2</sup>, Sören Twarock<sup>1</sup>, Jonas Butschan<sup>1</sup>, Anna Babayan<sup>2</sup>, Diana Klein<sup>3</sup>, Shirley K. Knauer<sup>4</sup>, Eric Metzen<sup>5</sup>, Volkmar Müller<sup>6</sup>, Verena Jendrossek<sup>3</sup>, Klaus Pantel<sup>2</sup>, Karin Milde-Langosch<sup>6</sup>, Jens W. Fischer<sup>1,\*</sup>, Katharina Röck<sup>1,\*</sup>

<sup>1</sup>Institut für Pharmakologie und Klinische Pharmakologie, Universitätsklinikum der Heinrich-Heine-Universität, Düsseldorf, Germany

<sup>2</sup>Department of Tumor Biology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

<sup>3</sup>Institute of Cell Biology (Cancer Research), University Hospital, University of Duisburg-Essen, Essen, Germany

<sup>4</sup>Institute for Molecular Biology II, Centre for Medical Biotechnology (ZMB), University of Duisburg-Essen, Essen, Germany

<sup>5</sup>Institute of Physiology, Faculty of Medicine, University Duisburg-Essen, Essen, Germany

<sup>6</sup>Department of Gynecology, University Hospital Hamburg-Eppendorf, Hamburg, Germany

\*These authors have contributed equally to this work

Correspondence to: Jens W. Fischer, e-mail: jens.fischer@uni-duesseldorf.de

Keywords: breast cancer, ionizing radiation, RHAMM, cell death, extracellular matrix

Received: September 29, 2015

Accepted: January 20, 2016

Published: February 8, 2016

### ABSTRACT

**Biomarkers for prognosis in radiotherapy-treated breast cancer patients are urgently needed and important to stratify patients for adjuvant therapies. Recently, a role of the receptor of hyaluronan-mediated motility (RHAMM) has been suggested for tumor progression. Our aim was (i) to investigate the prognostic value of RHAMM in breast cancer and (ii) to unravel its potential function in the radiosusceptibility of breast cancer cells. We demonstrate that RHAMM mRNA expression in breast cancer biopsies is inversely correlated with tumor grade and overall survival. Radiosusceptibility *in vitro* was evaluated by sub-G1 analysis (apoptosis) and determination of the proliferation rate. The potential role of RHAMM was addressed by short interfering RNAs against RHAMM and its splice variants. High expression of RHAMMv1/v2 in p53 wild type cells (MCF-7) induced cellular apoptosis in response to ionizing radiation. In comparison, in p53 mutated cells (MDA-MB-231) RHAMMv1/v2 was expressed sparsely resulting in resistance towards irradiation induced apoptosis. Proliferation capacity was not altered by ionizing radiation in both cell lines. Importantly, pharmacological inhibition of the major ligand of RHAMM, hyaluronan, sensitized both cell lines towards radiation induced cell death. Based on the present data, we conclude that the detection of RHAMM splice variants in correlation with the p53 mutation status could help to predict the susceptibility of breast cancer cells to radiotherapy. Additionally, our studies raise the possibility that the response to radiotherapy in selected cohorts may be improved by pharmaceutical strategies against RHAMM and its ligand hyaluronan.**

### INTRODUCTION

Radiotherapy has become standard of care for most breast cancer cohorts [1]. Radiation has significantly reduced the risk of local recurrence and also improved overall survival [2]. However, cancer cells can acquire

radioresistance with its complementary risk of increased mortality rates [3]. Furthermore, radiotherapy has been shown to increase the risk of cardiovascular diseases [4]. Hence, discovery of targets predicting the response to radiotherapy as well as agents that sensitize cancer cells to ionizing radiation with low side effects, is of great interest.

lines (Fig. 3B). MCF-7 cells revealed a significant increase in the apoptotic rate as measured by sub-G1 analysis explaining the decrease of the total cell number (Fig. 3C-3D). In comparison, MDA-MB-231 cells were found to be radioresistant (Fig. 3A-3D). In order to investigate the underlying mechanism of increased cellular death in MCF-7 cells, a protein array was performed analyzing the phosphorylation pattern of proteins involved in apoptosis. p53 and p38 were significantly increased in MCF-7 cells 48h after initial radiation (Fig. 3E-3G). This effect was not seen in MDA-MB-231 cells. Of note, MCF-7 cells harbor wild type p53 whereas MDA-MB-231 cells harbor a mutated form of the protein leading to its accumulation in the nucleus [19]. The involvement of p53 in the induction of apoptosis in MCF-7 cells was further validated by treatment with siRNA against *p53* prior to radiation. Short interfering RNA specific for *p53* abolished the pro-apoptotic effect of radiation in MCF-7 cells (Fig. 3H). Inhibition of p38 by SB202190 did not abrogate irradiation induced cell death. These results indicate that MCF-7 cell death in response to ionizing radiation is induced by the *p53* pathway.

Since both analyzed BC cell lines cells exhibit a divergent estrogen receptor (ER) status (MCF7:positive; MDA-MB-231 negative [18]), we next tested whether the ER is involved in the radiosusceptibility. However, treatment of MCF-7 cells with the unspecific ER-antagonist ICI182780 did not abrogate the pro-apoptotic effect of ionizing radiation in this cell type (Suppl. Fig. S1).

### RHAMM is regulated by radiation in a p53 correlated manner

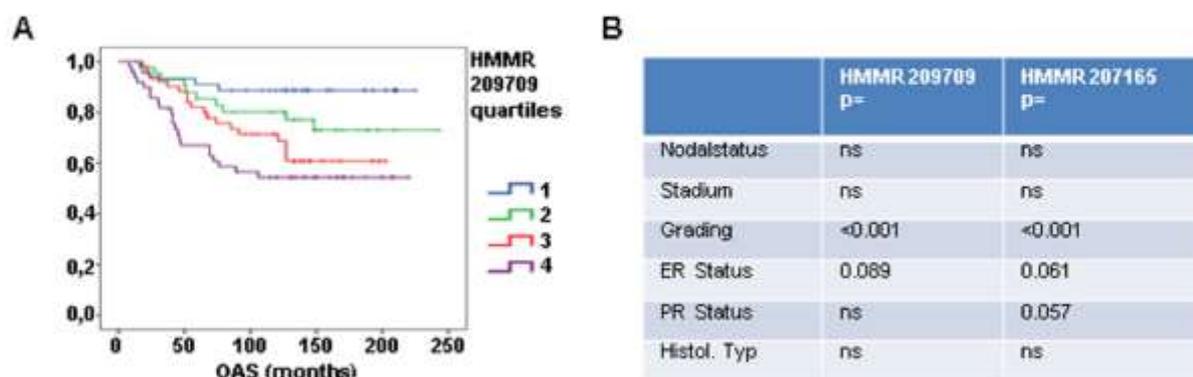
To investigate the role of RHAMM in response to radiation, both cell lines were irradiated with 2Gy and *RHAMM* mRNA level was measured by qRT-PCR (Fig. 4A). *RHAMM* mRNA was significantly reduced in MCF-

7 cells, likewise shown by immunocytochemical staining of RHAMM (Fig. 4B). No change was detected in MDA-MB-231 cells resembling the results on the apoptotic response (Fig. 3C). Immunocytochemistry staining of p53 showed increased signals in irradiated MCF-7 cells compared to sham-irradiated controls. Accumulation of p53 in MDA-MB-231 cells was found independent of ionizing radiation (Fig. 4B). Alternative splicing of RHAMM might be responsible for different cellular functions. The regulation four different protein isoforms regulation of four by ionizing radiation was analyzed by western blot. Of note, only splice variants v1/v2 which both run at 85kDa were decreased in MCF-7 cells while v3 and v4 were not reduced. In contrast, MDA-MB-231 cells displayed a significantly lower expression of all tested isoforms and were not further decreased by radiation (Fig. 4C-4D). Recently, *RHAMM* has been shown to be transcriptionally repressed by p53 [11]. Treatment of both cell lines with short interfering RNA against *p53* confirmed these results (Fig. 4E-4G). The endogenously increased level of p53 in the nucleus of MDA-MB-231 cells could therefore explain the reduced occurrence of RHAMM protein in this cell line. Of note, reduction of RHAMM by siRNA did not change the level of p53 in both cell lines (Suppl. Fig. S2).

Next it was investigated whether the increase of RHAMM v1/v2 in MDA-MB-231 cells after p53 knock down would restore radiosensitivity. Knock down of p53 and subsequent upregulation of RHAMM v1/v2 increased the rate of cellular death in MDA-MB-231 cells. Of note, the apoptotic effect was even further enhanced by subsequent ionizing irradiation (Fig. 4H).

### RHAMM splice variants increase the radiosensitivity of breast cancer cell lines

To establish the radiosensitizing ability of RHAMM - observed in terms of apoptosis - and to investigate the



**Figure 1: RHAMM is prognostic for patient overall survival.** **A.** Affymetrix analysis of *RHAMM* expression in 196 tissue samples from breast cancer patients is shown. Patients were stratified into subgroups according their *RHAMM* expression (low (1), medium (2), high (3), very high (4)) and the subgroups were correlated to overall survival. **B.** table showing results of statistic tests for clinical parameter in two affymetrix analysis.

involvement of the different *RHAMM* splice variants, cells were treated with siRNA against the respective mRNAs (Suppl. Fig. S3) and subsequently irradiated. In MCF-7 cells *siRHAMMpan* as well as siRNA against all individual *RHAMM* splice variants increased the rate of apoptosis (white bars Fig. 5A).

Again, irradiation induced the percentage of apoptotic cells. An additional apoptotic effect was observed in cells which were treated with *siRHAMM v1/v2* and irradiation. Knock down of *RHAMM v3* and *v4* did not alter this effect (Fig. 5A).

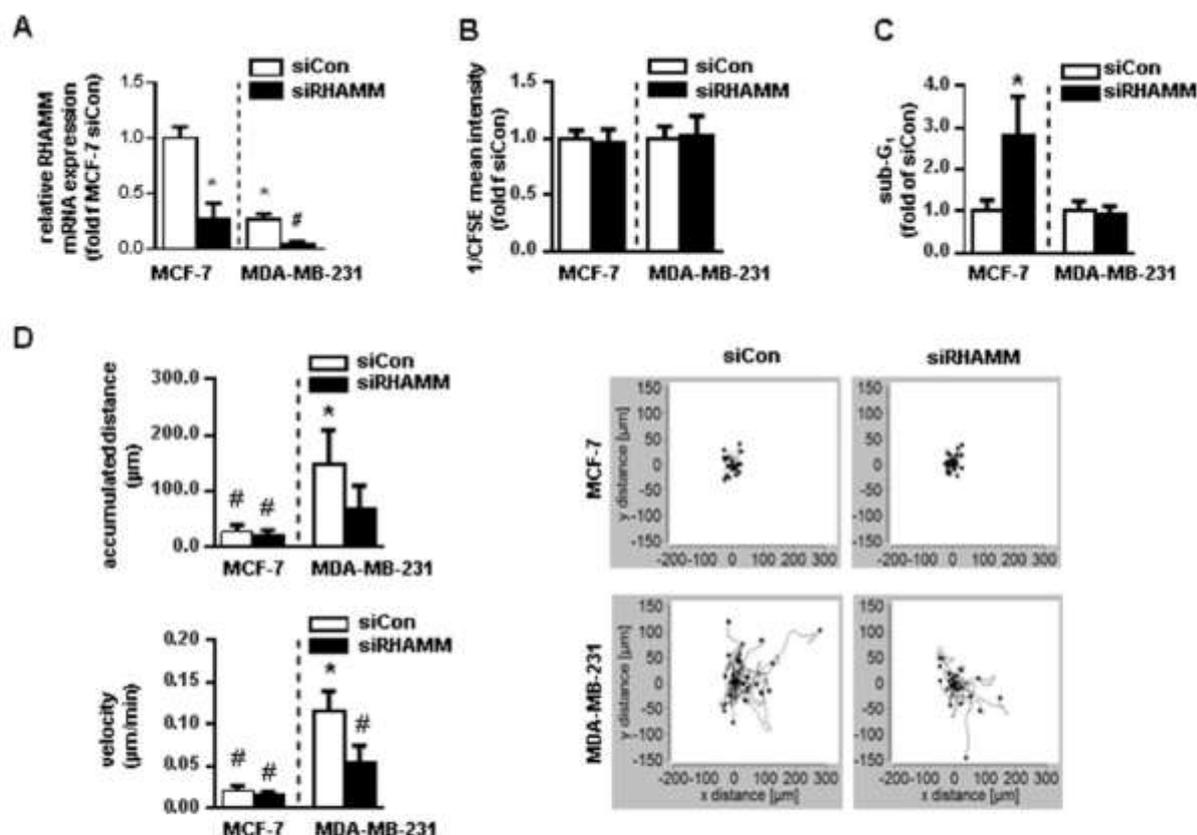
In MDA-MB-231 cells neither *siRHAMMpan* nor *siRHAMMv1/v2* revealed a significant induction of cell death (Fig. 5B). However, siRNA against *RHAMMv3* and *v4* induced cell death in MDA-MB-231 cells. Irradiation did not increase MDA-MB-231 cell death in any treatment group.

As expected from the previous experiments, the proliferation rate was not altered by either ionizing radiation or siRNA treatment (Fig. 5C-5D).

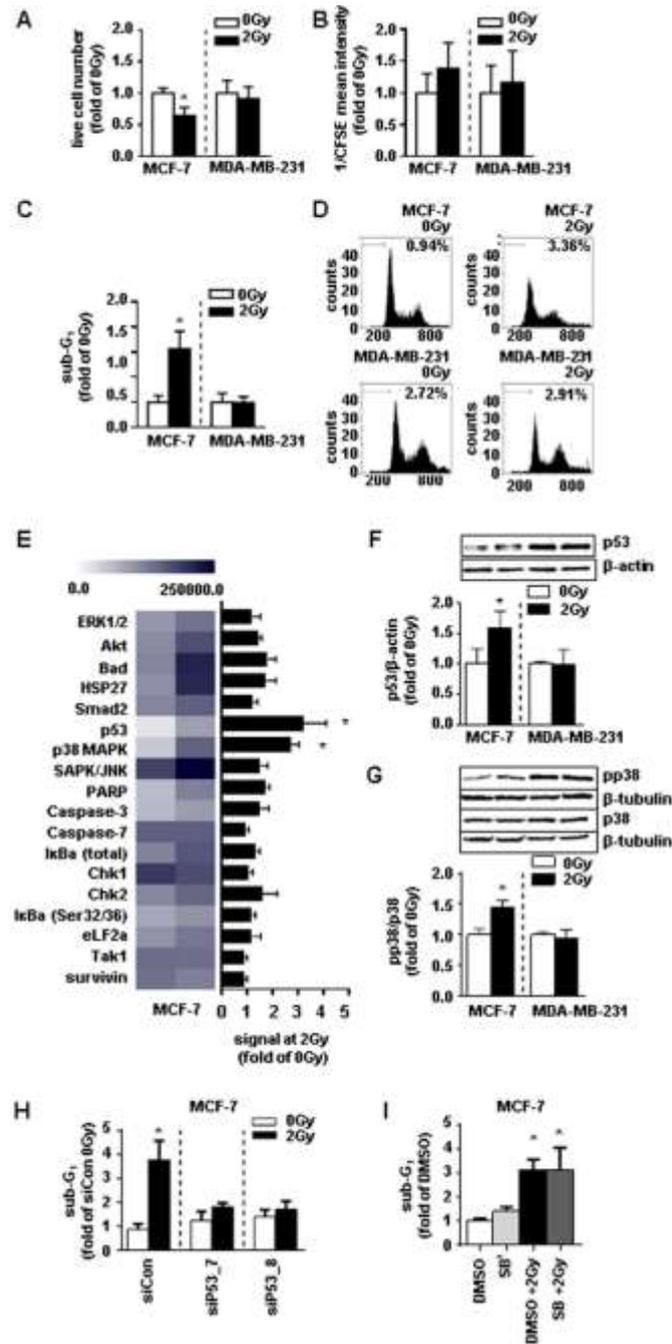
### Pharmacologic inhibition of hyaluronan synthesis increases the response to ionizing radiation

Finding pharmacological approaches which increase the susceptibility of cancer cells to radiation and thereby reduce the rate of recurrence is of great interest. HA is the main intra- and extracellular ligand of *RHAMM* thereby regulating both functional aspects [20]. Its pharmacological inhibition can be realized by treatment of cells with the HA-synthase inhibitor 4-methylumbelliferone (4-MU). 4-MU treatment suppressed HA synthesis in irradiated and non-irradiated breast cancer cells (Fig. 6A). Importantly, incubation of MCF-7 cells with 4-MU increased the radiosensitivity of the cells with respect to apoptosis fourfold. Whereas MDA-MB-231 cells did not respond to 4-MU treatment alone, the susceptibility of the cells after additional radiation was increased (Fig. 6B-6C).

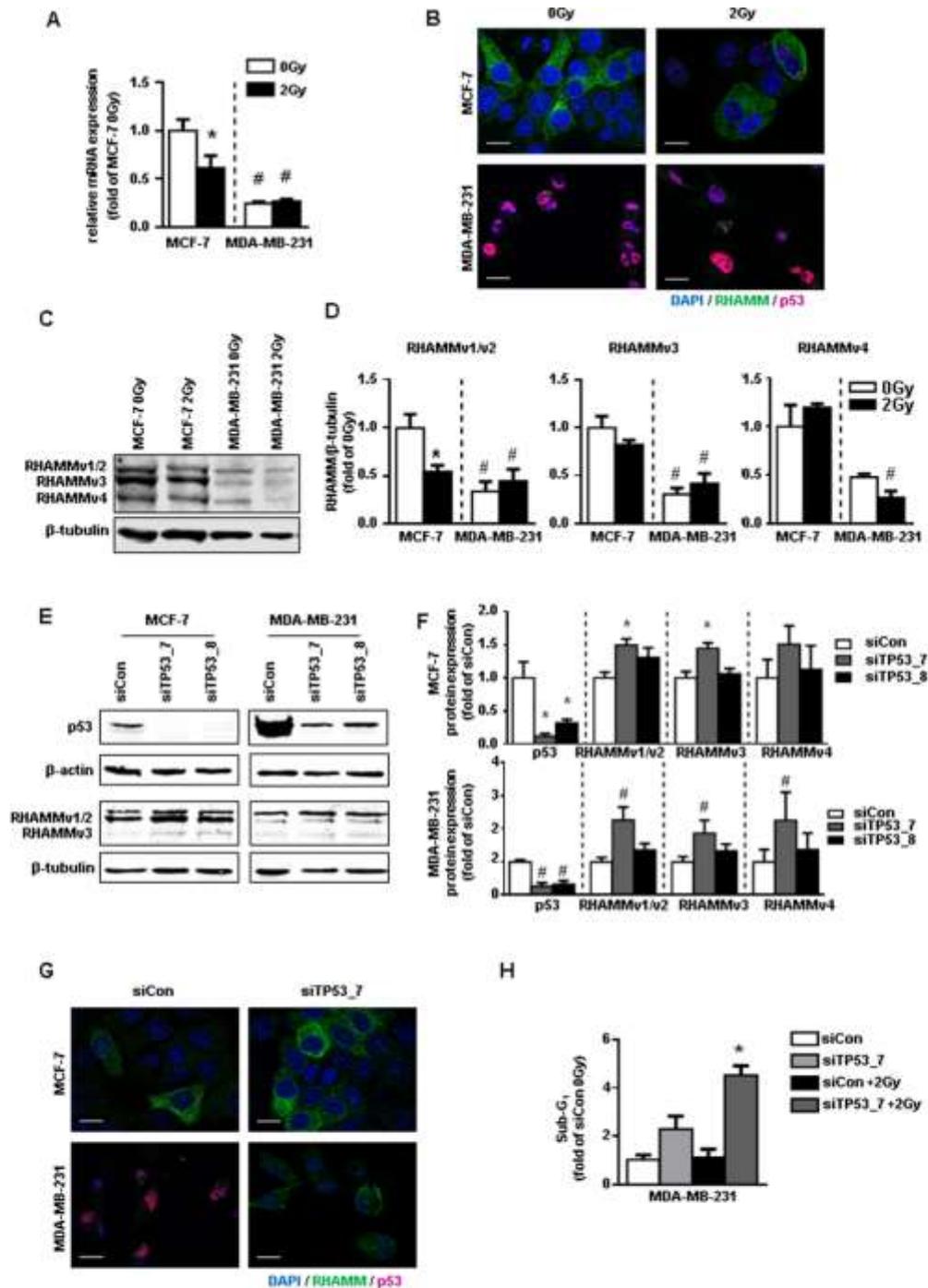
In conclusion, we provide evidence, that *RHAMM* is involved in the malignant phenotype of BC cells.



**Figure 2:** *RHAMM* has apoptotic and motility characteristics in different cancer cell lines *in vitro*. A. relative mRNA expression of *RHAMM* in *siRHAMM* transfected MCF-7 and MDA-MB-231 cells. B. proliferation rate measured with CFSE staining in MCF-7 and MDA-MB-231 cells 48h after siRNA knockdown of *RHAMM*. CFSE intensity is assigned reciprocally. C. sub-G<sub>1</sub> analysis of MCF-7 and MDA-MB-231 cells 48h after siRNA knockdown of *RHAMM*. D. accumulated distance and velocity of migration assay with exemplary pictures of one experiment. \*, p<0.05 in comparison to MCF-7 siCon; #, p<0.05 in comparison to MDA-MB-231 siCon.



**Figure 3: MDA-MB-231 cells are not radiosensitive at a dose of 2Gy whereas MCF-7 cells are radiosensitive via activation of p53 and p38.** A, live cell number, B, proliferation rate of via CFSE (assigned reciprocally), and C, sub-G1 analysis of MCF-7 and MDA-MB-231 cells 48h after irradiation with 2Gy measurement. D, representative histograms of sub-G1 analyses. E, absolute expression levels of proteins analyzed via the stress and apoptosis assay depicted in a heatmap (left) and the fold changes of these proteins in MCF-7 cells 48h after irradiation compared to the non-irradiated control (right). F, western blot analyses of MCF-7 cell lysates for p53 and G, pp38 and p38 protein expression. H, sub-G1 analysis in MCF-7 cells transfected with siRNA against p53 48h after irradiation with 2Gy. I, proliferation rate of MCF-7 cells treated with SB (SB202190, p38-inhibitor, 10  $\mu$ M) 4h before irradiation with 2Gy. \*,  $p < 0.05$  in comparison with untreated MCF-7 0Gy (A-G, J, K)/siCon (H, I).



**Figure 4: p53 and RHAMM variant status of MCF-7 and MDA-MB-231 in response to 2Gy irradiation.** **A.** relative mRNA expression of *RHAMM* pan. **B.** RHAMM (green) and p53 (pink) immunofluorescence staining 48h after irradiation with 2Gy. Scale bars: 20µm. **C.** western blot analysis of RHAMM protein expression and its quantification and **D.** quantification. **E.** exemplary blots of protein expression of p53 and RHAMM of p53 depleted MCF-7 and MDA-MB-231 cells and **F.** quantification. **G.** RHAMM (green) and p53 (pink) immunofluorescence staining 48h after siRNA knockdown of p53. Scale bars: 20µm. **H.** MDA-MB-231 were transfected with sip53. 48h after transfection cells were irradiated. Sub-G<sub>1</sub> was analyzed further 48h later. \*, p<0.05 in comparison to MCF-7 siCon, #, p<0.05 MDA-MB-231 in comparison to MCF-7 siCon.

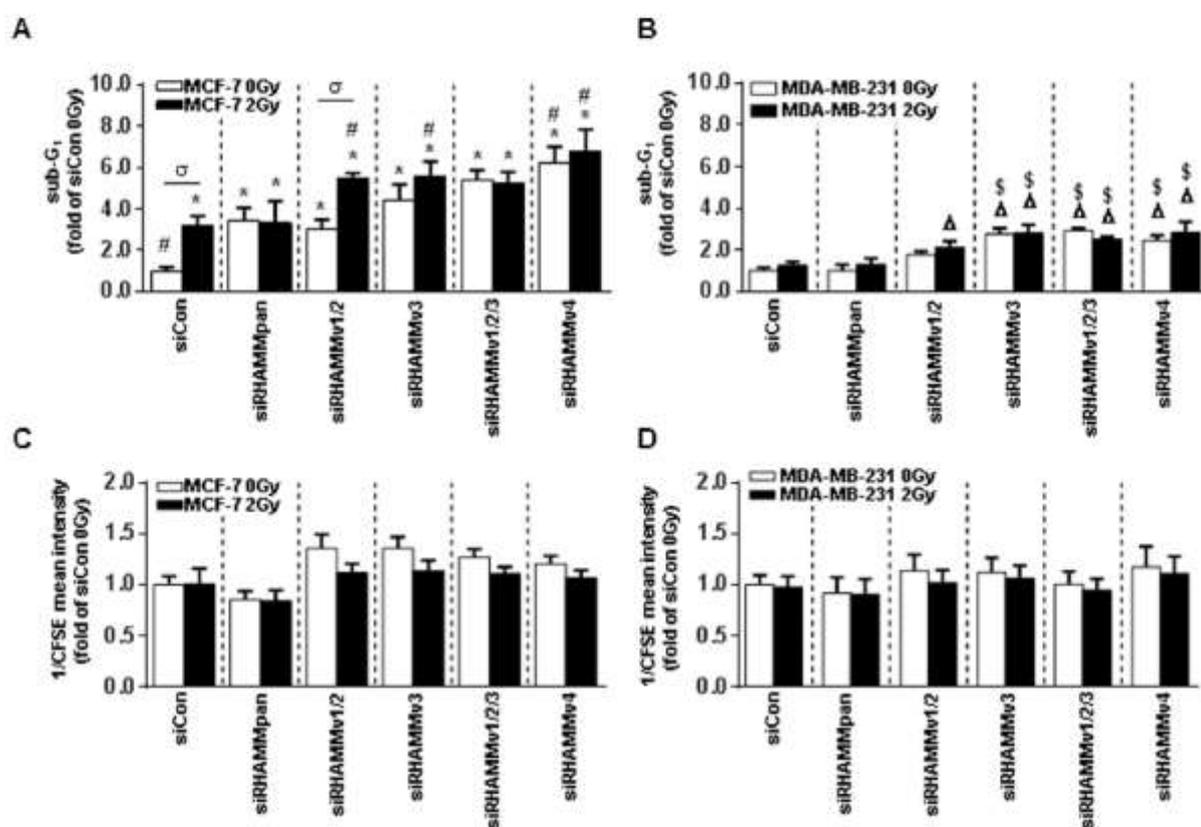
Detection of *RHAMM* isoform expression in correlation with the *p53* mutation status might allow to predict the responsiveness to radiation. Importantly, pharmacological inhibition of HA, the main binding partner of *RHAMM*, could help to increase the radiosensitivity of both *p53* wild type and mutated cancer types (Fig. 7).

## DISCUSSION

The receptor for hyaluronan-mediated motility (*RHAMM*) exhibits at least two distinct functions. As intracellular protein it is involved in maintaining the stability of the mitotic spindle [12]. In addition, by attachment to a GPI-anchor, it is associated with the cellular membrane and acts as cellular HA receptor promoting cell motility and invasion [15, 21]. Various studies report overexpression of *RHAMM* during tumor development and suggest a prognostic significance of *RHAMM* expression in e.g. leukemia, bladder cancer, and BC [13, 22, 23]. These findings have fostered the idea to

use *RHAMM* for therapy in acute myeloid leukemia and multiple myeloma, now being evaluated as vaccination against *RHAMM* in small clinical trials [24]. However, its value as prognostic marker and therapeutic target remains indeterminate.

In agreement with other reports, our data show that the level of *RHAMM* in tumor biopsies derived from BC patients is correlated with recurrence-free and OS. Interestingly *in vitro* its functional role also appears to depend on the expression level. Whereas in *RHAMM*<sup>high</sup> MCF-7 cells the cellular survival depends on the *RHAMM* expression, MDA-MB-231 cells (*RHAMM*<sup>low</sup>) survive independently of *RHAMM*. Of note, this circumstance is reversed with regard to the migratory capacity of both cell types. One possible explanation is that *RHAMM* is transported to the membrane in invasive cancer cell lines [21] and thereby increases migratory behavior. This suggests that (i) the amount of *RHAMM* mRNA is important for the prognosis of the disease and (ii) that the subcellular localization might be informative for the



**Figure 5: *RHAMM*pan and *RHAMM* variant knock-down in MCF-7 and MDA-MB-231 cells leads to increased radiosensitivity after irradiation with 2Gy. A, sub-G<sub>1</sub> cell cycle analysis of *siRHAMM* variants transfected MCF-7 and B, MDA-MB-231 cells 48h after irradiation with 2Gy. C, analysis of proliferation rate via CFSE measurement (assigned reciprocally) in *siRHAMM* variant transfected MCF-7 and D, MDA-MB-231 cells 48h after irradiation with 2Gy. \*, p<0.05 in comparison to MCF-7 siCon 0Gy, #, p<0.05 in comparison to MCF-7 siCon 2Gy, α, p<0.05 MCF-7 siRHAMMv1/2 2Gy in comparison to MCF-7 siRHAMMv1/2 0Gy, Δ, p<0.05 in comparison to MDA-MB-231 siCon 0Gy, \$, p<0.05 in comparison to MDA-MB-231 siCon 2Gy.**

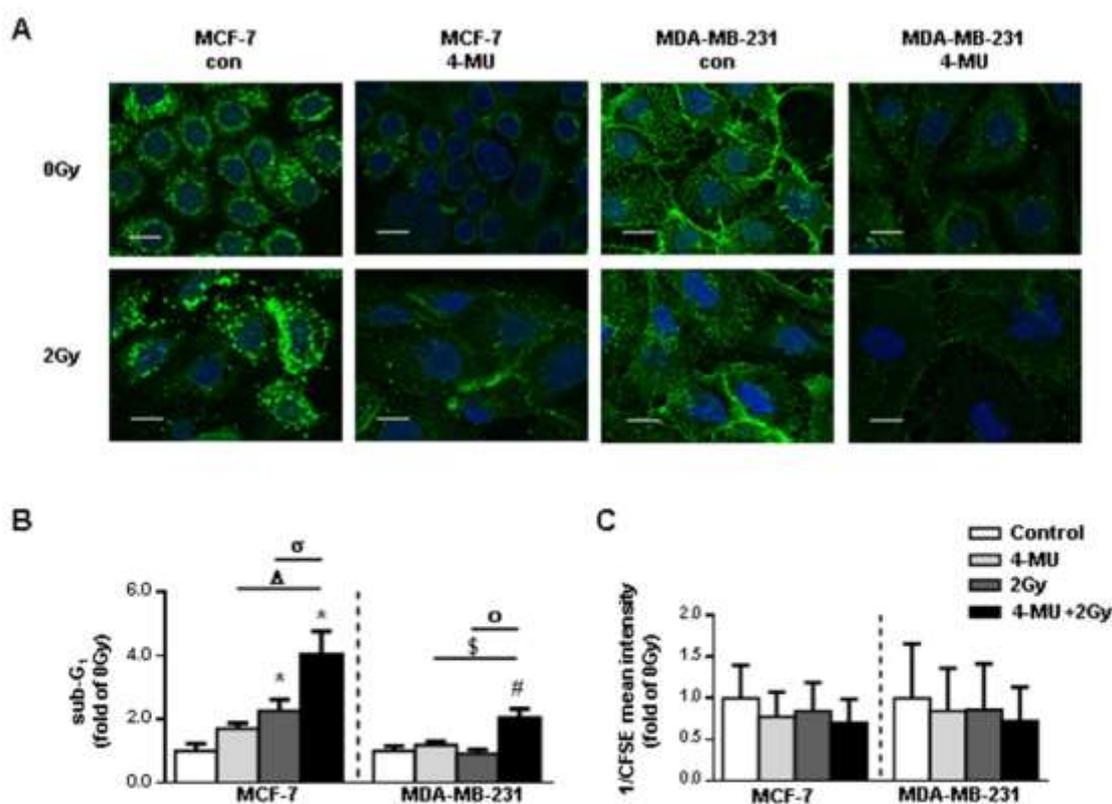
functional role. Future studies will have to be conducted to clarify this open issue.

Several factors may be associated with altered cellular radiosensitivity and the probability of cells to be killed by apoptotic mechanisms including *p53*, *FAS*-mediated pathways, and the *BCL-2* gene family [25]. The relationship between the level of endogenous apoptosis of tumors and their radiosensitivity has been investigated leading to conflicting results [26, 27]. However, it has been shown that cells from radiosensitive tumors are more susceptible to ionizing radiation compared to cells from unresponsive tumors [28]. Hence, the apoptotic response may serve as predictive assay for intrinsic radiosensitivity [29]. In the present study, a threefold increase in the apoptotic rate of *RHAMM*<sup>high</sup> MCF-7 was detected. In comparison, *RHAMM*<sup>low</sup> cells (MDA-MB-231) were found to be radioresistant. Apoptosis in caspase-3 deficient cells (MCF-7) has been reported to depend on alternative pathways like caspase-7 activation of DFF40 like DNase [30]. In this study analysis of

proteins involved in radiation induced cell death of MCF-7 cells revealed an involvement of *p53* in this process. Of interest MDA-MB-231 cells harbor mutations in the *p53* gene [31], whereas MCF-7 express wild-type *p53* [32]. A number of tumors acquire *p53* mutations with a high frequency [33-36]. In general, patients with these mutations respond poorly to treatment with ionizing radiation.

In genome wide mRNA screens *RHAMM* was observed to be decreased when *p53* is active [7, 37, 38]. Furthermore, it was recently described that *p53* can repress *RHAMM* expression via its promotor including the first exon and intron [11].

In the present study, we demonstrate that *RHAMM* expression in MDA-MB-231 was severely reduced in correlation with an accumulation of *p53* within the nucleus. Even though it is generally postulated, that mutant *p53* cells have lost the ability to bind consensus *p53* DNA binding regions, it has been demonstrated that mutated *p53* is still able to regulate gene expression directly or



**Figure 6: Pharmacological inhibition of HA system via 4-MU in MCF-7 and MDA-MB-231 leads to increased radiosensitivity.** A, affinitycytochemistry of HA 48h after irradiation with 2Gy. Scale bars: 20µm. B, sub-G1 cell cycle analysis of 4-MU treated MCF-7 and MDA-MB-231 cells 48h after irradiation with 2Gy. CFSE. C, analysis of proliferation rate via CFSE staining of 4-MU treated MCF-7 and MDA-MB-231 cells 48h after irradiation with 2Gy. CFSE intensity is assigned reciprocally. \*, p<0.05 in comparison to MCF-7 0Gy, Δ, p<0.05 in comparison to MCF-7 4-MU, σ, p<0.05 in comparison to MCF-7 2Gy, #, p<0.05 in comparison to MCF-7 4-MU, o, p<0.05 in comparison to MCF-7 2Gy.

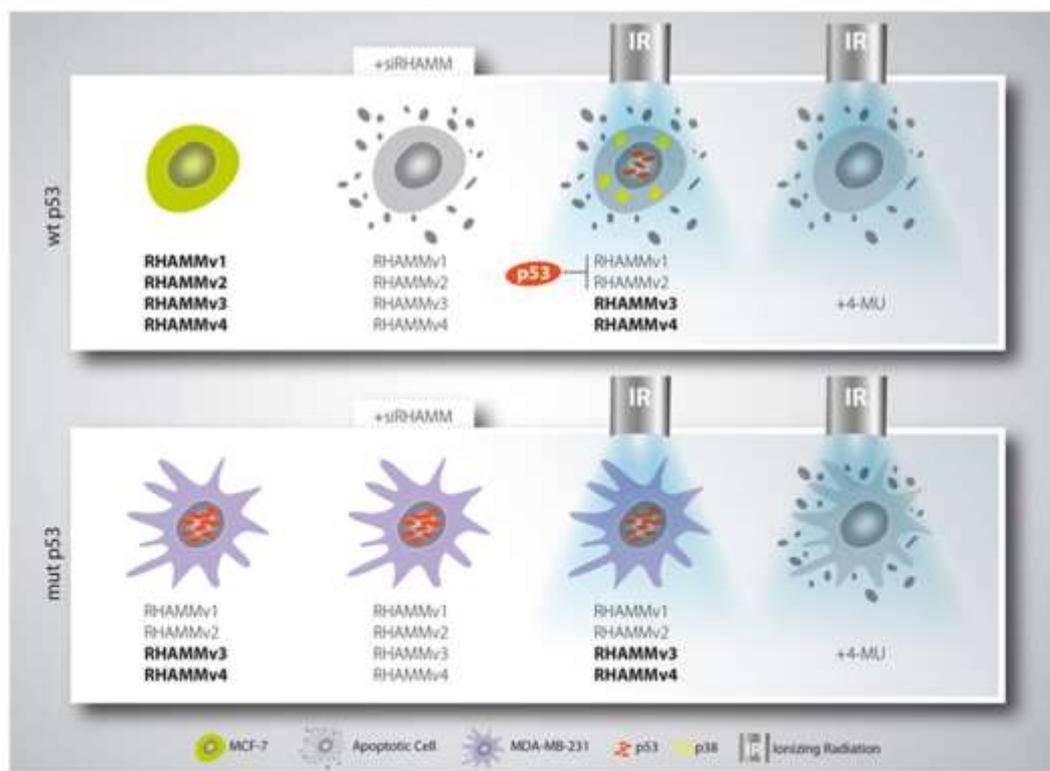
through mitochondrial or cytoplasmatic activities [39]. In MDA-MB-231 cells it was previously reported that the p53 mutant is required for cellular survival through yet unknown mechanism [40]. In the present study siRNA against p53 raised RHAMM protein levels indicating that mutated p53 in MDA-MB-231 cells is still able to repress RHAMM expression.

Of interest, not all RHAMM isoforms appear to be decreased by p53 as only RHAMMv1/v2 was reduced in MDA-MB-213 cells and elevated in response to siRNA against p53. Different functions of the RHAMM splice variants have been proposed previously. In multiple myeloma patients the ratio of RHAMMv3:RHAMMv1/v2 has been correlated with poor prognosis [16]. Furthermore RHAMMv3 appears to promote tumor growth and metastasis to lymph nodes and the liver in a mouse model [14]. As RHAMMv1/v2 encodes for longer proteins than RHAMMv3 and v4 and the transport mechanism of RHAMM to the membrane is still unknown, alternative splicing of RHAMM might present one reason for its presence in different subcellular compartments and thereby different cellular functions.

Regarding radiation responses the expression of RHAMM v1/v2 appears to be vital. Whereas

RHAMMv1/v2<sup>high</sup> MCF-7 cells were sensitive to ionizing radiation, RHAMMv1/v2<sup>low</sup> MDA-MB-231 cells were radioresistant. It appears plausible that RHAMMv1/2 is permanently reduced in p53 mutated cancer cells and other compensatory pathways exist, thereby conferring radioresistance.

CD44 is a widely expressed cell surface membrane receptor which participates in cell-cell and cell-matrix interactions. CD44 facilitates motogenic/invasive as well as proliferative cellular phenotypes [41]. Conflicting functions of CD44 have been proposed in experimental models of tumorigenesis and –progression in comparison to *in vivo* data. This may be due to the presence and absence of RHAMM and *vice versa* [42]. CD44 is known to co-operate with RHAMM and has been reported to compensate for loss of RHAMM. RHAMM and CD44 unify at least two distinct characteristics: i) both have been shown to be transcriptionally repressed by p53 [11, 43] and ii) they share the same binding partner, HA. HA is a glycosaminoglycan and an important component of the extracellular matrix which has been associated with BC progression [44]. *In vitro* HA induces BC cell motility [45] and survival [46]. HA has been predicted to bind all RHAMM isoforms near the carboxyl-terminus [47]. In



**Figure 7: Schematic model of mammary cancer radiosensitization.** Breast cancer cells with wildtype p53 status are radiosensitive and can be forced into apoptosis upon RHAMM downregulation. Mutant p53 breast cancer cells are only radiosensitive if treated with 4-MU which affects RHAMM ligand inhibition.

the present study treatment with the HA-inhibitor 4-MU augmented the radiation effect in both cell lines. However, 4-MU did only alter MDA-MB-231 cellular survival in combination with irradiation. This hints towards a HA dependent resistance mechanism possibly mediated by CD44 or RHAMM v3/v4. Future studies will have to address the role of CD44 and RHAMM for radiosensitivity of cancer cells, possibly also including further binding partners of both proteins, such as MET [48]. Apart from 4-MU the application of other pharmacological compounds interfering with HA-signaling pathways, as RHAMM/CD44 blocking antibodies or HA-binding peptides, should be considered.

Prognostic markers may help to improve the accuracy of risk stratification of cancer patients and therefore possibly provide important data to optimize therapeutic decisions. In the present study we report that the detection of *RHAMM* splice variants in correlation with the *p53* mutation status might help to pre-evaluate the susceptibility of breast cancer cells towards radiotherapy. Additionally, the data raise the possibility that the response to radiotherapy in selected tumors may be improved by targeting RHAMM and its ligand HA.

## MATERIALS AND METHODS

### Analysis of mRNA expression data in tumor tissue samples

In order to investigate the mRNA expression of *RHAMM*, microarray data (Affymetrix HG-U133A) from a cohort of 196 mammary carcinoma enrolled in the Department of Gynecology, Hamburg University medical Center, were analyzed. The clinical and histological characteristics of this cohort as well as technical details have been described elsewhere [49]. All microarray data have been submitted to Gene Expression Omnibus (GEO) under the following accession numbers: GSE26971 (samples GSM663775-GSM663852), GSE31519 (samples GSM782523-GSM782529), GSE31519 (samples GSM782554-GSM782568), GSE46184 (samples GSM1125783-GSM112856) [49]. Informed consent for the scientific use of tissue materials, which was approved by the local ethics committee (Ethik-Kommission der Ärztekammer Hamburg, #OB/V/03), was obtained from all patients. The study was performed in accordance to the principles of the declaration of Helsinki and REMARK criteria [50].

Expression data of two RHAMM probe sets present on the Affymetrix chips (209709\_s\_at and 207165\_at) were retrieved from the files. For statistical analysis, the cohort was divided into quartiles of similar size according to their expression values.  $\chi^2$ -tests were used to examine correlations between RHAMM expression and clinicopathological factors comparing the following groups: histological grading G1 vs. G2 vs. G3; stage

pT1 vs. pT2 vs. pT3-4; lymph node involvement N0 vs. N1; estrogen and progesterone receptor status (ER, PR), positive vs. negative; histological type ductal vs. others. Overall survival was analyzed by Kaplan-Meier analysis and Log-Rank-Tests. Probability values less than 0.05 were regarded as statistically significant.

### Treatment and transfection of cells

MCF-7 (CLS, Eppelheim) and MDA-MB-231 (CLS, Eppelheim) were cultured in DMEM-high glucose (Gibco, Carlsbad) with 2% FBS (Gibco) and 1% Penicillin/Streptomycin (Gibco). Ionizing radiation was applied with Gulmay RS225 (Xstrahl, Camberley) at 150kV and 15mA with 0.2mm copper filter 24h after treatment or seeding of 5000 cells per cm<sup>2</sup>. 10nM siRNA (AllStars Negative Control siRNA, *siHMMR* 9 (*siRHAMMpan*), *siTP53\_7* and *siTP53\_8* from Qiagen, Hilden and custom siRNA for *RHAMM* variants were purchased from Sigma) was used for transfection with 1:500 RNAiMax (Life Technologies, Carlsbad) in DMEM according to the manufacturer's manual. siRNA against sequences for knockdown of *RHAMM* variants are AAAGAGATTCGTGTTCTTCTACA (*siRHAMMv1/2*), AAGATTCGTGTTCTTCTACAGGA (*siRHAMMv3*), AAAGTTAAGTCTTCGGAATCAAA (*siRHAMMv1/2/3*), and AATGACCCCTTCGATTCGTGT (*siRHAMMv4*). Cells were irradiated 24h after reversed transfection with siRNA against *RHAMM* or 4-MU (Sigma-Aldrich Chemie, München) treatment. Cells receiving *p53* knockdown were irradiated 48h after reversed transfection with siRNA against *p53*.

### Migration assay

Migration of the cells was investigated via time lapse microscopy starting 24h after ionizing radiation with 2Gy. Three videos per sample were recorded with Axio Observer.Z1 under standard culturing conditions and eight cells per video were tracked. Tracking was performed using ImageJ 1.47t (Wayne Rasband, National Institutes of Health, USA) manual tracking plugin (Fabrice Cordeli, Institut Curie, Orsay, France) and exemplary pictures are processed with Chemotaxis and Migration Tool (Ibidi, Gerhard Trapp, Martinsried, Germany).

### Western blotting

Western blots were performed using standard procedures and the following reagents: RHAMM (GTX62573, GeneTex, Irvine), p53 (OP43, Calbiochem, Billerica), p38 (#9212, Cell Signaling Technologies, Danvers), pp38 (#9211, Cell Signaling Technologies, Danvers),  $\beta$ -tubulin (T7816, Sigma-Aldrich Chemie, München) and  $\beta$ -actin (A5316, Sigma-Aldrich Chemie, München). Binding protein for affinitycytochemistry of hyaluronan was biotinylated HABP (Calbiochem,

Billerica) and as secondary antibody streptavidin-FITC (Dako, Glostrup) was used. Secondary antibodies for western blotting were IRDye<sup>®</sup> 800CW goat anti-rabbit IgG, IRDye<sup>®</sup> 800CW goat anti-mouse IgM, IRDye<sup>®</sup> 680RD goat anti-rabbit IgG and IRDye<sup>®</sup> 680LT goat anti-mouse IgM (LI-COR Biotechnology, Bad Homburg). Immunofluorescence secondary antibodies were AF488 goat anti-rabbit IgG (Life Technologies, Carlsbad) and AF568 (Fab'), fragment of goat anti-mouse IgG (H+L) (Life Technologies, Carlsbad).

### Immunofluorescence staining and affinitycytochemistry of hyaluronic acid

For RHAMM and p53 immunofluorescent stainings cells were fixed with 4% paraformaldehyde (PFA). For hyaluronic acid affinitycytochemical staining cells were fixed with 70% EtOH/4% PFA/0.5% acidic acid. Stainings were performed as described previously [51, 52].

### RNA isolation, cDNA transcription, and qRT-PCR

RNA isolation was performed according to the PeqLab peqGOLD TriFast (Erlangen) protocol. cDNA transcription was carried out with QuantiTect Reverse Transcriptase Kit (Qiagen, Hilden) according to the instruction manual. qRT-PCR was performed in duplicate on StepOnePlus RealTime PCR System using Platinum SYBR Green qPCR SuperMix-UDG (Life Technologies, Carlsbad) with ROX reference dye according to the manufacturer's protocol. Primer sequences (5'→3') for *GAPDH* f are GTGAAGGTCCGAGTCAACG, *GAPDH* r TGAGGTCAATGAAGGGGTC, *RHAMMpan* f GAATTTGAGAATTCTAAGCTTG and *RHAMMpan* r CCATCATACCCCTCATCTTTGTT were used. Data were analyzed by  $\Delta\Delta$ CT-method using *GAPDH* as reference gene.

### Intracellular signaling array

Cells were lysed in Path Scan<sup>®</sup> Sandwich ELISA Lysis Buffer (Cell Signaling Technology, Danvers) and analyzed using PathScan<sup>®</sup> Stress and Apoptosis Signaling Antibody Array Kit (Cell Signaling Technology, Danvers) according to the vendor's protocol. Chemiluminescent signals were detected via Odyssey Infrared Imaging System (application software version 3.0) from Li-Cor biosciences (Bad Homburg).

### Sub-G1 cell cycle analysis

Cells were detached via incubation with Trypsin/EDTA (Gibco), washed once with PBS and resuspended in 75  $\mu$ L Lysis buffer (0.1% Sodium citrate, 0.1% Triton X-100). Directly before FACS analysis 25  $\mu$ L GUAVA cell cycle reagent (Millipore, Billerica) were added and

doublet-discrimination was performed via signal height/area dot plot. 10000 cells per sample were analyzed employing a guava easyCyte 5 flow cytometer (Millipore).

### Proliferation and determination of proliferation rate

Cells were detached and diluted 1:2 with TrypanBlue (Life Technologies, Carlsbad) for live dead discrimination and determination of live cell number in Countess (Invitrogen, Carlsbad).

To evaluate proliferation rate cells were stained with CFSE (Life Technologies, Carlsbad) as shown previously [53]. Cells were harvested 48h after irradiation. Analysis was performed employing the easyCyte5 flow cytometer (Millipore, Billerica). 10000 cells were analyzed per sample. Mean of fluorescence signal was assigned reciprocal.

### Statistical analysis

Real-time data were analyzed by logarithmic data transformation. Data were analyzed with GraphPad Prism 6 (GraphPad Software, La Jolla, CA, USA) and are represented as mean  $\pm$  SEM. Comparison of two groups was analyzed via two-tailed t-test. Comparison of more than two independent variables was analyzed via one-way ANOVA and Sidak's multiple comparison *post hoc* test. Statistical significance was considered when  $p < 0.05$ .

### ACKNOWLEDGMENTS

We thank Prof. Reiner Jänicke, Laboratory for molecular Radiooncology, University Hospital Düsseldorf, to provide the Gulmay's RS225 irradiator, and Petra Rompel for her excellent technical assistance. This work was supported by grant from the DFG (GRK1739/10).

### CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

### REFERENCES

1. Favourable and unfavourable effects on long-term survival of radiotherapy for early breast cancer: an overview of the randomised trials. *Lancet* 2000; 355:1757–1770.
2. Mokbel K. Current management of ductal carcinoma in situ of the breast. *Int. J. Clin. Oncol.* 2003; 8:18–22.
3. Boyages J, Delaney G, Taylor R. Predictors of local recurrence after treatment of ductal carcinoma in situ: a meta-analysis. *Cancer* 1999; 85:616–28.
4. Hooning MJ, Botma A, Aleman BM, Baaijens MH, Bartelink H, Klijn JG, Taylor CW, van Leeuwen FF. Long-term risk of cardiovascular disease in 10-year survivors of breast cancer. *J. Natl. Cancer Inst.* 2007; 99:365–75.

5. Wilder RB, Curcio LD, Khanijou RK, Eisner ME, Kakkis JL, Chittenden L, Agustin J, Lizarde J, Mesa AV, Macedo JC, Ravera J, Tokita KM. Results with accelerated partial breast irradiation in terms of estrogen receptor, progesterone receptor, and human growth factor receptor 2 status. *Int. J. Radiat. Oncol. Biol. Phys.* 2010; 78:799–803.
6. Chang I, Graham PH, Hao J, Ni J, Bucci J, Cozzi PJ, Kearsley JH, Li Y. PI3K/Akt/mTOR pathway inhibitors enhance radiosensitivity in radioresistant prostate cancer cells through inducing apoptosis, reducing autophagy, suppressing NHEJ and HR repair pathways. *Cell Death Dis.* 2014; 5:e1437.
7. Kho PS, Wang Z, Zhuang L, Li Y, Chew JL, Ng HH, Liu ET, Yu Q. P53-Regulated Transcriptional Program Associated With Genotoxic Stress-Induced Apoptosis. *J. Biol. Chem.* 2004; 279:21183–21192.
8. Yoon KW, Byun S, Kwon E, Hwang SY, Chu K, Hiraki M, Jo SH, Weins A, Hakrrouch S, Cebulla A, Sykes DB, Greka A, Mundel P, et al. Control of signaling-mediated clearance of apoptotic cells by the tumor suppressor p53. *Science* (80- ). 2015; 349:1261669–1261669.
9. Meilwrath AJ, Vasey P a, Ross GM, Mellwrath AJ, Brown R. Cell Cycle Arrests and Radiosensitivity of Human Tumor Cell Lines : Dependence on Wild-Type p53 for Radiosensitivity Advances in Brief Cell Cycle Arrests and Radiosensitivity of Human Tumor Cell Lines : Dependence on Wild-Type p53 for Radiosensitivity'. 1994; 3718–3722.
10. Meyn MS, Strasfeld L, Allen C. Testing the Role of p53 in the Expression of Genetic Instability and Apoptosis in Ataxia-telangiectasia. *Int. J. Radiat. Biol.* 1994; 66:S141–S149.
11. Sohr S, Engeland K. RHAMM is differentially expressed in the cell cycle and downregulated by the tumor suppressor p53. *Cell Cycle* 2008; 7:3448–3460.
12. Chen H, Mohan P, Jiang J, Nemirovsky O, He D, Fleisch MC, Niederacher D, Pilarski LM, Lim CJ, Maxwell CA. Spatial regulation of Aurora A activity during mitotic spindle assembly requires RHAMM to correctly localize TPX2. *Cell Cycle* 2014; 13:2248–2261.
13. Niedworok C, Kretschmer I, Röck K, Vom Dorp F, Szarvas T, Heß J, Freudenberg T, Melchior-Becker A, Rübber H, Fischer JW. The impact of the receptor of hyaluronan-mediated motility (RHAMM) on human urothelial transitional cell cancer of the bladder. *PLoS One* 2013; 8:e75681.
14. Du Y-CN, Chou C-K, Klimstra DS, Varmus H. Receptor for hyaluronan-mediated motility isoform B promotes liver metastasis in a mouse model of multistep tumorigenesis and a tail vein assay for metastasis. *Proc. Natl. Acad. Sci.* 2011; 108:16753–16758.
15. Hamilton SR, Fard SF, Paiwand FF, Tolg C, Veiseh M, Wang C, McCarthy JB, Bissell MJ, Koropatnick J, Turley EA. The hyaluronan receptors CD44 and Rhamm (CD168) form complexes with ERK1,2 that sustain high basal motility in breast cancer cells. *J. Biol. Chem.* 2007; 282:16667–80.
16. Maxwell CA, Rasmussen E, Zhan F, Keats JJ, Adamia S, Strachan E, Crainie M, Walker R, Belch AR, Pilarski LM, Barlogie B, Shaughnessy J Jr, Reiman T. RHAMM expression and isoform balance predict aggressive disease and poor survival in multiple myeloma. *Blood* 2004; 104:1151–8.
17. Assmann V, Marshall JF, Fieber C, Hofmann M, Hart IR. The human hyaluronan receptor RHAMM is expressed as an intracellular protein in breast cancer cells. *J. Cell Sci.* 1998; 111:1685–1694.
18. Veiseh M, Kwon DH, Borowsky AD, Tolg C, Leong HS, Lewis JD, Turley EA, Bissell MJ. Cellular heterogeneity profiling by hyaluronan probes reveals an invasive but slow-growing breast tumor subset. *Proc. Natl. Acad. Sci.* 2014; 111: E1731–E1739.
19. Brosh R, Rotter V. When mutants gain new powers: news from the mutant p53 field. *Nat. Rev. Cancer* 2009; 9:701–13.
20. Ziebell MR, Zhao ZG, Luo B, Luo Y, Turley EA, Prestwich GD. Peptides that mimic glycosaminoglycans: High-affinity ligands for a hyaluronan binding domain. *Chem. Biol.* 2001; 8:1081–1094.
21. Hall CL, Yang B, Yang X, Zhang S, Turley M, Samuel S, Lange LA, Wang C, Curpen GD, Savani RC, Greenberg AH, Turley EA. Overexpression of the hyaluronan receptor RHAMM is transforming and is also required for H-ras transformation. *Cell* 1995; 82:19–26.
22. Crainie M, Belch a R, Mant MJ, Pilarski LM. Overexpression of the receptor for hyaluronan-mediated motility (RHAMM) characterizes the malignant clone in multiple myeloma: identification of three distinct RHAMM variants. *Blood* 1999; 93:1684–1696.
23. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM. ONCOMINE: A Cancer Microarray Database and Integrated Data-Mining Platform. *Neoplasia* 2004; 6:1–6.
24. Greiner J, Schmitt A, Giannopoulos K, Rojewski MT, Götz M, Funk I, Ringhoffer M, Bunjes D, Hofmann S, Ritter G, Döhner H, Schmitt M. High-dose RHAMM-R3 peptide vaccination for patients with acute myeloid leukemia, myelodysplastic syndrome and multiple myeloma. *Haematologica* 2010; 95:1191–1197.
25. Zhivotovsky B, Joseph B, Orrenius S. Tumor radiosensitivity and apoptosis. *Exp. Cell Res.* 1999; 248:10–7.
26. Sirzén F, Zhivotovsky B, Nilsson A, Bergh J, Lewensohn R. Spontaneous and radiation-induced apoptosis in lung carcinoma cells with different intrinsic radiosensitivities. *Anticancer Res.* 1997; 18:695–9.
27. Belcheva I, Stoychev T, Karadjov K, Danchev D. Neuropharmacological activity of newly-synthetized derivatives of 3,3-diethyl-2,4-pyridinedione. III. N-Acyl derivatives of 3,3-diethyl-2,4-pyridinedione. *Acta Physiol. Pharmacol. Bulg.* 1979; 5:75–81.
28. Olive PL, Durand RF. Apoptosis: an indicator of radiosensitivity *in vitro*? *Int. J. Radiat. Biol.* 1997; 71:695–707.

29. Decaudin D, Delic J, Dumont J, Tertian G, Blot E, Dubray B, Grandpeix C, Peffault de Latour R, Cosset JM. Clinical efficacy of irradiation in CLL patients: predictive value of *in vitro* radio-induced apoptosis. *Leuk. Lymphoma* 2002; 43:827–9.
30. Devarajan E, Sahin AA, Chen JS, Krishnamurthy RR, Aggarwal N, Brun AM, Sapino A, Zhang F, Sharma D, Yang XH, Tora AD, Mehta K. Down-regulation of caspase 3 in breast cancer: a possible mechanism for chemoresistance. *Oncogene* 2002; 21:8843–8851.
31. Olivier M, Feles R, Hollstein M, Khan MA, Harris CC, Hainaut P. The IARC TP53 database: new online mutation analysis and recommendations to users. *Hum. Mutat.* 2002; 19:607–14.
32. Lu X, Errington J, Curtin NJ, Lunec J, Newell DR. The impact of p53 status on cellular sensitivity to antifolate drugs. *Clin. Cancer Res.* 2001; 7:2114–2123.
33. Stretch JR, Gatter KC, Ralfkiaer E, Lane DP, Harris AL. Expression of mutant p53 in melanoma. *Cancer Res.* 1991; 51:5976–5979.
34. Chiba I, Takahashi T, Nau MM, D'Amico D, Curiel DT, Mitsudomi T, Buchhagen DL, Carbone D, Piantadosi S, Koga H, et al. Mutations in the p53 gene are frequent in primary, resected non-small cell lung cancer. *Lung Cancer Study Group. Oncogene* 1990; 5:1603–10.
35. Hinds PW, Finlay CA, Quartin RS, Baker SJ, Fearon ER, Vogelstein B, Levine AJ. Mutant p53 DNA clones from human colon carcinomas cooperate with ras in transforming primary rat cells: a comparison of the “hot spot” mutant phenotypes. *Cell Growth Differ.* 1990; 1:571–580.
36. Yamaguchi T, Mukai H, Yamashita S, Fujii S, Ushijima T. Comprehensive DNA Methylation and Extensive Mutation Analyses of HER2-Positive Breast Cancer. *Oncology* 2015; 88:377–84.
37. Spurgers KB, Gold DL, Coombes KR, Bohnenstiel NL, Mullins B, Meyn RE, Logothetis CJ, McDonnell TJ. Identification of cell cycle regulatory genes as principal targets of p53-mediated transcriptional repression. *J. Biol. Chem.* 2006; 281:25134–25142.
38. Rother K, John C, Spiesbach K, Haugwitz U, Tschöp K, Wasner M, Klein-Hitpass I, Möröy T, Mössner J, Engeland K. Identification of Tcf-4 as a transcriptional target of p53 signalling. *Oncogene* 2004; 23:3376–3384.
39. Kim E, Deppert W. Transcriptional activities of mutant p53: When mutations are more than a loss. *J. Cell. Biochem.* 2004; 93:878–886.
40. Hui L, Zheng Y, Yan Y, Bargonetti J, Foster D a. Mutant p53 in MDA-MB-231 breast cancer cells is stabilized by elevated phospholipase D activity and contributes to survival signals generated by phospholipase D. *Oncogene* 2006; 25:7305–7310.
41. Hiraga T, Ito S, Nakamura H. Cancer stem-like cell marker CD44 promotes bone metastases by enhancing tumorigenicity, cell motility, and hyaluronan production. *Cancer Res.* 2013; 73:4112–4122.
42. Maxwell CA, McCarthy J, Turley E. Cell-surface and mitotic-spindle RHAMM: moonlighting or dual oncogenic functions? *J. Cell Sci.* 2008; 121:925–932.
43. Godar S, Ince TA, Bell GW, Feldser D, Donaher JL, Bergh J, Liu A, Miu K, Watnick RS, Reinhardt F, McAllister SS, Jacks T, Weinberg RA. Growth-inhibitory and tumor-suppressive functions of p53 depend on its repression of CD44 expression. *Cell* 2008; 134:62–73.
44. Edward M, Gillan C, Micha D, Tammi RH. Tumour regulation of fibroblast hyaluronan expression: A mechanism to facilitate tumour growth and invasion. *Carcinogenesis* 2005; 26:1215–1223.
45. Tzircotis G, Thorne RF, Isacke CM. Chemotaxis towards hyaluronan is dependent on CD44 expression and modulated by cell type variation in CD44-hyaluronan binding. *J. Cell Sci.* 2005; 118:5119–5128.
46. Auvinen P, Tammi R, Parkkinen J, Tammi M, Agren U, Johansson R, Hirvikoski P, Eskelinen M, Kosma VM. Hyaluronan in peritumoral stroma and malignant cells associates with breast cancer spreading and predicts survival. *Am. J. Pathol.* 2000; 156:529–536.
47. Yang B, Yang BL, Savani RC, Turley EA. Identification of a common hyaluronan binding motif in the hyaluronan binding proteins RHAMM, CD44 and link protein. *EMBO J.* 1994; 13:286–96.
48. Ponta H, Sherman L, Herrlich PA. CD44: from adhesion molecules to signalling regulators. *Nat. Rev. Mol. Cell Biol.* 2003; 4:33–45.
49. Milde-Langosch K, Karn T, Schmidt M, zu Eulenburg C, Oliveira-Ferrer L, Wirtz RM, Schumacher U, Witzel I, Schütze D, Müller V. Prognostic relevance of glycosylation-associated genes in breast cancer. *Breast Cancer Res. Treat.* 2014; 145:295–305.
50. Meshane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. REporting recommendations for tumor MARKer prognostic studies (REMARK). 2005; 2:416–422.
51. Röck K, Tigges J, Sass S, Schütze A, Florea AM, Fender AC, Theis FJ, Krutmann J, Boege F, Fritsche E, Reifemberger G, Fischer JW. miR-23a-3p Causes Cellular Senescence by Targeting Hyaluronan Synthase 2: Possible Implication for Skin Aging. *J. Invest. Dermatol.* 2014; 135:369–377.
52. Twarock S, Tammi MI, Savani RC, Fischer JW. Hyaluronan Stabilizes Focal Adhesions, Filopodia, and the Proliferative Phenotype in Esophageal Squamous Carcinoma Cells. *J. Biol. Chem.* 2010; 285:23276–23284.
53. Quah BJC, Warren HS, Parish CR. Monitoring lymphocyte proliferation *in vitro* and *in vivo* with the intracellular fluorescent dye carboxyfluorescein diacetate succinimidyl ester. *Nat. Protoc.* 2007; 2:2049–56.

## 8. SUMMARY (IN ENGLISH) / ZUSAMMENFASSUNG (AUF DEUTSCH)

Cancer metastasis is the main cause of cancer related death. Administration of systemic therapy against cancer metastasis is usually based on characteristics of the primary tumor, however, metastases may not resemble the primary tumor anymore due to genetic progression or selection of treatment-resistant clones. The discordance in characteristics of primary tumor and metastases, and the fact that metastases are infrequently biopsied for clinical diagnostics, might lead to treatment failure. Circulating tumor cells (CTCs) extracted from peripheral blood could aid diagnostics and clinical management of cancer patients. Individual CTCs can be investigated on proteomic and genomic levels. Immunocytochemical (ICC) staining allows for CTC detection and characterization. Downstream genetic analysis is possible with a combination of whole genome amplification (WGA) and next generation sequencing (NGS). Such comprehensive analysis provides us with information about the origin and genetic heterogeneity of the metastases and might aid in possible identification of therapy sensitive and resistant clones.

In the studies presented here, a triple immunostaining protocol and a workflow for genetic analysis of single CTCs were established. Subsequently, these procedures were applied for the investigation of clonal evolution towards breast cancer metastasis, its intra-patient heterogeneity, and the role of heterogeneity in acquired resistance to hormone treatment and radiotherapy in breast cancer patients.

The obtained results suggest that genetic heterogeneity of breast cancer plays a key role in the resistance of therapy. Cross activation of proliferative signaling pathways results in resistance to endocrine and radiotherapy. We demonstrate that breast cancer might utilize both linear and parallel progression ways of metastasis. Accordingly, therapy resistant metastases in breast cancer patients might originate from tumor clones present at early stages of carcinogenesis, as well as from more progressed ones. The clonality of the investigated breast tumors indicates that both competition and cooperation of tumor clones are likely being involved in cancer evolution.

In conclusion, our results demonstrate the feasibility of genomic and protein expression analyses on single CTCs and underline the importance of “liquid biopsy” for companion diagnostics in metastatic breast cancer.

Metastasen sind die Hauptursache krebsbedingter Todesfälle. Die Verabreichung systemischer Therapien gegen Krebsmetastasen basiert in der Regel auf Eigenschaften des Primärtumors. Stetige Tumorprogression sowie Selektion behandlungsresistenter Klone führen jedoch dazu, dass Metastasen dem Primärtumor nicht mehr gleichen. Diese Diskordanz zwischen dem Primärtumor und den Metastasen, sowie die Tatsache, dass Metastasen selten einer Biopsie unterzogen werden, können zum Therapieversagen führen. In diesem Zusammenhang bieten zirkulierende Tumorzellen (sogenannte Circulating Tumor Cells, oder CTCs), die sich von dem Primärtumor oder den Metastasen abspalten und im Blut befinden, eine verhältnismäßig leicht zugängliche und dennoch kostbare Quelle für Tumor-, bzw. Metastasen-Material.

Die Charakterisierung der, aus dem peripheren Blut von Krebspatienten extrahierten, CTCs kann sowohl die Diagnostik als auch die klinische Behandlung der Patienten unterstützen. Einzelne CTCs können auf proteomischer und genomischer Ebene untersucht werden. Immunzytochemische Färbungen ermöglichen die CTC-Detektion und -Charakterisierung. Darauf folgende genetische Einzelzellanalysen sind nur auf Grund einer Kombination von Amplifikation des gesamten Genoms (Whole Genome Amplification, WGA) und Next Generation Sequencing (NGS) möglich. Solch umfassende Analysen liefern Informationen über die Herkunft und die genetische Heterogenität der Metastasen und können der Identifizierung von therapiesensiblen und resistenten Klonen dienen.

In den hier vorgestellten Studien wurden ein immunozytochemisches Dreifachfärbeprotokoll und ein Verfahren für die genetische Analyse einzelner CTCs etabliert. Anschließend wurden diese Verfahren zur Untersuchung der Klonalität und Heterogenität des Mammakarzinoms und der Metastasierungswege angewendet. Hierbei untersuchten wir die Rolle der genetischen Heterogenität bei der erworbenen Resistenz gegen Hormonbehandlung und Strahlentherapie bei Brustkrebs-Patientinnen. Die erhaltenen Ergebnisse deuten darauf hin, dass genetische Heterogenität in Brustkrebs eine wichtige Rolle bei der Therapieresistenz spielt. Die Kreuzaktivierung von proliferativen Signalwegen führt zu einer Kreuzresistenz gegenüber Endokrin- und Strahlentherapie. Anschließend zeigten wir, dass Brustkrebsmetastasierung sowohl über lineare als auch parallele Tumorprogression erfolgt. Dementsprechend können therapieresistente Metastasen bei Mammakarzinom-Patientinnen von Tumorklonen stammen, die bereits in frühen

Stadien der Karzinogenese entstanden sind, sowie von weiter fortgeschrittenen. Die Klonalität der untersuchten Brusttumore legt die Vermutung nahe, dass beide Mechanismen: Wettbewerb und Zusammenarbeit von Tumorklonen, in der Krebsentwicklung beteiligt sind.

Schließlich unterstreichen unsere Ergebnisse die Umsetzbarkeit und Bedeutung der Genom- und Proteinexpressionsanalyse auf einzelnen CTCs im Rahmen einer "liquid biopsy" für therapiebegleitende Diagnostik in Mammakarzinom-Patientinnen.

## 9. ERKLÄRUNG DES EIGENANTEILS AN DEN PUBLIKATIONEN

Meine Teilnahme an den folgenden Publikationen beinhaltet:

Publikation 1, "Heterogeneity of Estrogen Receptor Expression in Circulating Tumor Cells from Metastatic Breast Cancer Patients", PlosOne 2013 [107]:

- Planung und Durchführung aller Experimente, sowie Statistische- und Patientendaten-Analyse, Manuskriptverfassung, Erstellung von Abbildungen und weitere Bearbeitung zwecks Revision.

Publikation 2, "Comparative study of whole genome amplification and next generation sequencing performance of single cancer cells", submitted:

- Studiendesign, Planung und Durchführung aller Experimente, Analyse der NGS Daten, Statistische Datenaufarbeitung und Interpretation, Manuskriptverfassung, Erstellung von Abbildungen, Einreichen.

Publikation 3, „Clonal evolution of metastatic breast cancer: two cases, two progression models“,manuscript in preparation:

- Studiendesign, Planung und Durchführung aller Experimente, Analyse der NGS Daten, Analyse der Patientendaten, Datenaufarbeitung und Dateninterpretation, Manuskriptverfassung, Erstellung von Abbildungen.

Publikation 4, „RHAMM splice variants confer radiosensitivity in human breast cancer cell lines“, Oncotarget 2016 [108]:

- Teilnahme an Dateninterpretation und Datenaufarbeitung, Erstellung von Abbildungen, Korrekturlesen des Manuskripts.

## 10. DANKSAGUNG

Ich bedanke mich bei Herrn Prof. Dr. Pantel für seine Unterstützung, seinen Input, sein Lob und Vertrauen, und ganz besonders für die mir gegebene Möglichkeit die Doktorarbeit im Institut für Tumorbiologie zu vollenden trotz Wechsels des Wissenschaftlichen Betreuers meiner Doktorarbeit.

Ich danke herzlich meinem direkten Betreuer, Herrn Dr. Simon Joosse, für seine stetige wissenschaftliche Betreuung, motivierende und angenehme Arbeitsweise, Hilfsbereitschaft und Verständnis in allen Lebenslagen. Ich danke ihm ganz besonders für die Möglichkeit mich zu entfalten, meine Ideen, sogar die Verrücktesten, einzubringen und auszuprobieren. Die Zusammenarbeit war nicht nur unglaublich spannend und herausfordernd, sondern hat auch viel Spaß gemacht.

Ich bedanke mich bei Herrn Prof. Dr. Volkmar Müller für die gute und zuverlässige Zusammenarbeit und seinen Wissenschaftlichen Rat und Unterstützung.

Außerdem danke ich Julia Spötter für ihre hochqualifizierte Unterstützung und Mitarbeit, sowie ihre Freundschaft, die mir viel bedeutet.

Ich danke allen ehemaligen und aktuellen Mitarbeitern des Instituts für Tumorbiologie für das nette Arbeitsklima und die gute Zusammenarbeit. Besonders danke ich Dr. Sabine Riethdorf, Cornelia Coith, Antje Andreas, Sandra Schwentesius, Malgorzata Stoupiec, Susanne Hoppe und Andreas Bauche.

Außerdem danke ich all meinen Freunden, die ich im Institut kennengelernt habe, für ihre Präsenz in meinem Leben, sowohl in guten wie in schlechten Zeiten.

Ich danke Claudia Hille und Nele Schulte für die Durchsicht und Verbesserungsvorschläge meiner Doktorarbeit.

Des Weiteren möchte ich mich bei meiner Familie, meiner unerschöpfliche Kraftquelle, bedanken, sowie bei meinen Schwiegereltern in spe für die Unterstützung, Hilfsbereitschaft und den Glauben an mich.

Ich danke von ganzem Herzen meinem Verlobten Christoph Dette für die seelische Unterstützung und das aufgebrachte Verständnis für die langen Arbeitstage und ausfallenden Wochenenden.

Abschließend möchte ich diese Arbeit meinem im Januar 2016 verstorbenem Vater widmen. Seinen unbegrenzten Glauben an mich und seine unendliche Liebe zu mir möchte ich auf dieser Weise würdigen.

## 11. LEBENSLAUF

|              |   |
|--------------|---|
| Geburtsdatum | 25.02.1986                              |
| Geburtsort   | Moskau, Russische Föderation            |
| Wohnort      | Norderstedt, Bundesrepublik Deutschland |

### **Berufserfahrung**

|                     |   |
|---------------------|---|
| 11 / 2009 – aktuell | wissenschaftliche Mitarbeiterin<br>Universitätsklinikum Hamburg-Eppendorf,<br>Institut für Tumorbilogie, Leiter Herr Prof. Dr. Pantel   |
| 2012 – aktuell      | Arbeitsgruppe Dr. Joosse,<br>wissenschaftliche Mitarbeiterin,<br>seit 2013, in Verbindung mit der Doktorarbeit im Rahmen<br>des PhD-Programms für Nicht-Mediziner der medizinischen<br>Fakultät der Universität Hamburg |
| 2009 – 2012         | Arbeitsgruppe Dr. Hannemann,<br>wissenschaftliche Mitarbeiterin   |

### **Hochschulstudium**

|                       |   |
|-----------------------|---|
| 09 / 2003 – 06 / 2009 | Studium an der Russischen Staatlichen Medizinischen<br>Universität<br>Medizinisch-biologische Fakultät, Fachgebiet Biochemie.<br>Abschluss: Mit Auszeichnung<br>Die Qualifikation „Arzt-Biochemiker“ in der Fachrichtung<br>„Medizinische Biochemie“ wurde zuerkannt. |
| 08 / 2008 – 06 / 2009 | Diplomarbeit im Ärztlich-genetischen wissenschaftlichen<br>Zentrum, Laboratorium für Epigenetik, Moskau.<br>Thema: „Characterization of molecular-genetic alterations<br>as prognostic markers in superficial bladder cancer“   |

07 / 2006 – 06 / 2008      Wissenschaftliche Praxiserfahrung,  
Moskauer Medizinische Akademie I. M. Setchenow,  
Institut für Molekulare Medizin

### **Schulbildung**

09 / 1993 – 06 / 2003      Staatliche Mittelschule Nr.949 der Stadt Moskau  
inkl. Zusatzprogramm im Fach Psychologie (2001 – 2003)  
Abschluss: mit Auszeichnung

### **Weiterbildungen**

- Teilnahme an diversen Workshops im Rahmen des PhD-Programms: Projektmanagement, Presentation Skills, Advanced Writing Skills, Publication Process, Statistics
- Instrumentale Workshops: Advanced Imaging Methods, Scientific Programming
- Teilnahme an Führungs-Workshops

### **Sprachkenntnisse**

- Englisch - verhandlungssicher
- Deutsch - verhandlungssicher
- Russisch - Muttersprache

### **Publikationen**

- Babayan A., Alawi M., Müller V., Burandt E., Indenbirken D., Grundhoff A., Pantel K., Joosse S.A. Clonal evolution of metastatic breast cancer: two cases – two progression models. Manuscript in preparation.
- Babayan A., Alawi M., Müller V., Pantel K., Joosse S.A. Comparative study of the performance of whole genome amplification and next generation sequencing of single cancer cells. Submitted.
- Schütze A, Vogeley C., Gorges T.M., Twarock S., Butschan J., Babayan A., Klein D., Knauer S., Metzen A., Jendrossek V., Pantel K., Milde-Langosch K., Fischer J.W., Röck K. RHAMM splice variants confer radiosensitivity in human breast cancer cell lines. Oncotarget. 2016 Feb 8. doi: 10.18632/oncotarget.7258. [Epub ahead of print]

- Babayan A., Hannemann J., Spötter J., Müller V., Pantel K., Joosse S.A. Heterogeneity of Estrogen Receptor Expression in Circulating Tumor Cells from Metastatic Breast Cancer Patients. PLoS One. 2013 Sep 18;8(9):e75038. doi: 10.1371/journal.pone.0075038.

- Babayan A.Yu., Zaletaev D.V., Nemtsova M.V. Confirmation Of Field Cancerization In The Genesis Of Superficial Bladder Cancer. Molecular Medicine. 2013;1:24-28. ISSN 1728-2918.

- Babayan A.Yu., Karyakin O.B., Teplov A.A., Zaletaev D.V., Nemtsova M.V. Some molecular-genetic markers defining the pathogenesis of superficial and invasive bladder cancer. Molecular Biology. 2011;45(6):929-932.

- Nemtsova M.V., Paltseva E.M., Babayan A.Yu., Mihaylenko D.S., Babenko O.V., Samofalova O.Yu., Tsar'kov P.V., Zaletaev D.V. Molecular genetic analysis of the intratumoral clonal heterogeneity of colorectal adenocarcinomas. Molecular Biology. 2008;42(6):925-931.

- Babaian Alu, Andreeva Iulu, Zaletaev DV, Nemtsova MV. Clonal origin of multiple foci of urinary bladder cancer. Arkh Patol. 2012;74(5):44-50.

- Weitere 3 Publikationen als Erstautor in russischen Fachzeitschriften sowie mehrere Abstracts.

### **Konferenzteilnahmen**

- X. International Symposium on Minimal Residual Cancer, 2016, Hamburg, Deutschland – Poster

- VII. Forschungsretreat des Universitären Cancer Centers Hamburg, 2015, Jesteburg, Deutschland – Vortrag

- I. UKE Graduate Day 2015 for PhD Students and Postdocs, Hamburg, Germany – Poster

- V. Forschungsretreat des Universitären Cancer Centers Hamburg, 2013, Jesteburg, Deutschland – Vortrag
  
- III. Forschungsretreat des Universitären Cancer Centers Hamburg, 2011, Jesteburg, Deutschland – Vortrag
  
- II European Multidisciplinary Meeting on Urological Cancers 2009, Barcelona, Spanien – Vortrag. Molecular genetic alterations and their predictive values in superficial bladder cancer. Abstract book of the 2nd European Multidisciplinary Meeting on Urological Cancers. 2009. P. 42.
  
- European Human Genetics Conference, 23.-29. May 2009, Wien, Österreich – Poster
  
- The 24th Annual Congress of the European Association of Urology. 17.-21. März 2009, Stockholm, Sweden – Poster

## 12. EIDESSTATTLICHE ERKLÄRUNG

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe.

Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.

Ich erkläre mich einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Unterschrift: .....