Identification of metastasis founder cells in breast cancer by cell lineage tracing

DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES DER NATURWISSENSCHAFTEN (DR. RER. NAT.)
DER FAKULTÄT FÜR BIOLOGIE UND VORKLINISCHE MEDIZIN
DER UNIVERSITÄT REGENSBURG

vorgelegt von
Manjusha S Ghosh

aus
Assam, India

im Jahr
2020

Das Promotionsgesuch wurde eingereicht am:
10.07.2020

Die Arbeit wurde angeleitet von:
Prof. Dr. Christoph A. Klein

Unterschrift:

Table of Contents

Abstract

Background: Metastatic dissemination often occurs at early stages of breast cancer progression implying that disseminated cancer cells (DCCs) evolve outside the primary tumor in a process of selection and adaptation. To identify metastasis founder cells in humans we strived to construct cell lineage trees by longitudinally tracking genetic changes in cancer cells isolated from patients at various disease stages.

Methods: We used sample triplets comprising primary tumor, bone marrow DCCs isolated at presumably curative surgery and CTCs/DCCs at a further time point before and after progressing into metastasis. To prepare samples for single cell lineage tree analysis we developed robust methods for isolation of single cells of high DNA quality from all tissues, including - (i) Isolation of single cells from flash frozen tumor tissue; (ii) Laser capture microdissection to isolate DCCs from diagnostic cytospins; (iii) Isolation of blood CTCs from CellSearch cartridges. In addition, we isolated CD3+ T cells, CD68+ macrophages and oral epithelial cells to serve as outgroups for cell lineage tree analysis. Cell lineage tree reconstructions are based on short tandem repeats (STRs) mutations reflecting cell divisions which determine cellular descent by tracing random mutational events. Around 12000 STR loci are sequenced after target enrichment using a patient-generic panel of duplex molecular inversion probes.

Results: Detection, isolation and whole genome amplification from all sample sources could be successfully established. Results of STR based lineage tree analysis showed a significantly separate clustering of advanced primary tumor cells from metastatic cells. Interestingly these metastatic cells have detectable early DCC ancestors.

Conclusions: The data allows reconstructing the genomic make-up of metastases founder cells from the available phylogenetic trees. Longitudinal analysis of systemic breast cancer evolution may provide insights for diagnostic monitoring and inform the development of novel adjuvant therapies.

1. Introduction

1.1. Cellular organization of the mammary gland

The human adult mammary gland consists of branching ducts and lobules which are ectodermal in origin and stromal tissue which is mesodermal in origin, consisting mostly of adipocytes (Javed and Lteif 2013). Mammary glands mature at puberty by a process called branching morphogenesis (Lyons 1958), when the ductal epithelium grows into the adipose tissue in response to hormonal cues (Inman et al. 2015). Small morphogenetic changes at menstrual cycles to major remodelling events during pregnancy are orchestrated by cell differentiation, proliferation and apoptosis (Inman et al. 2015). A mature mammary duct is bi-layered (Figure 1) and comprises myoepithelial cells, which are in contact with the basement membrane, and luminal epithelial cells, which line the ducts (Inman et al. 2015). The mammary epithelium also consists of stem and progenitor cells which can give rise to both basal and luminal cell types and thereby maintain the epithelium (Visvader and Stingl 2014). During pregnancy the mammary epithelium rapidly proliferates to give rise to secretory alveoli. During lactation, the luminal cells secrete milk while the basal cells contract to help transport the milk through the ductal structures to the nipples. At weaning, the mammary epithelium undergoes massive apoptosis in a process called involution to return to its 'pre-pregnant' state (Inman et al. 2015). These complex tissue remodelling processes are tightly regulated in the healthy state, understanding of which can provide insights into diseases like breast cancer which occur due to disregulation of normal mammary gland development pathways (Visvader and Stingl 2014; Rosen 2012).

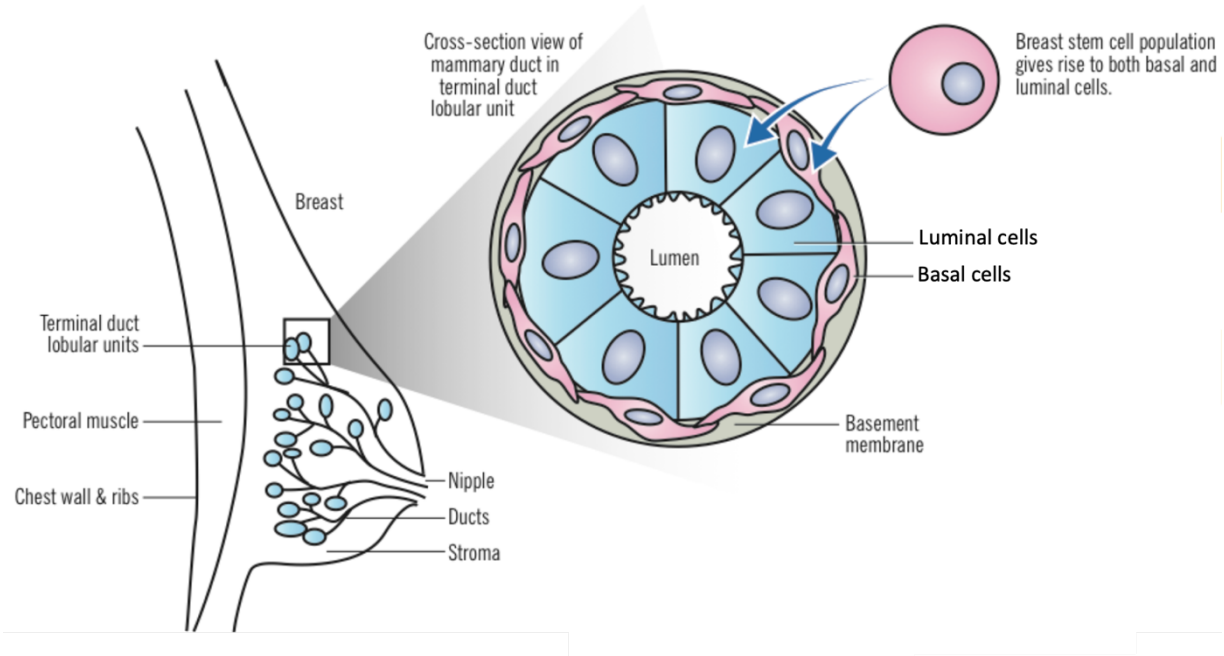


Figure 1. Breast anatomy. Mammary epithelium consists of a bilayer of two types of cells. Outer layer is made up of basal cells also called myoepithelial cells and inner layer consists of luminal epithelial cells. The basal layer also consists of stem and progenitor cells which gives rise to both cell types in the bilayer. The basement membrane separates the epithelium from the stromal tissue. Figure adapted from (McMaster Pathophysiology Review 2012-2018)

1.2. Breast Cancer

Breast cancer is one of the most common malignancies in women around the world. In 2018, more than 2 million diagnoses of breast cancer (11.6% of all diagnosed cancers worldwide) and 626,679 deaths (6.6% of global cancer related deaths) were estimated (Bray et al. 2018). In females, breast cancer accounts for 24.2% of all diagnosed cancers and contributes to 15% of all cancer related deaths (Torre et al. 2016; Seely and Alhassan 2018; Bray et al. 2018). It has been estimated from the current incidences that 1 in 8 to 10 women will develop breast cancer during their lifespan (Harbeck and Gnant 2017). The incidence of breast cancer in males is very low and accounts for <1% of all breast cancers and <1% of all diagnosed cancers in males (Abdelwahab Yousef 2017). Only female breast cancer has been addressed in this dissertation.

1.2.1. Breast cancer classification

Breast cancer can be classified broadly as either carcinoma in situ or invasive carcinoma (Malhotra et al. 2010) on the basis of histology. Carcinoma in situ is a non-obligate precursor of invasive carcinoma, refers to a stage where the cancer cells have not crossed the basement membrane to invade the surrounding tissues and includes ductal carcinoma in situ (DCIS) and lobular carcinoma in situ (LCIS) originating in the ducts and lobules respectively (Wen and Brogi 2018; Jaffer and Bleiweiss 2002). DCIS accounts for around 20% of breast tumors detected by mammography (Narod et al. 2015), however, LCIS cases are rare (0.5-2.5%) and are usually identified incidentally in breast biopsies (Wen and Brogi 2018). Similarly, invasive carcinoma includes both ductal and lobular types. Invasive ductal carcinoma (IDC) also called ‘no special type’ (NST) accounts for around 75% of breast carcinoma diagnoses (Li et al. 2003; Li, Uribe, and Daling 2005). Invasive lobular carcinoma (ILC) accounts for around 15% of all breast cancers (Arpino et al. 2004) and is characterized by small, round cells infiltrating the stroma in an irregular sheet-like pattern, instead of forming a tumor mass (Li, Uribe, and Daling 2005). ILCs are hard to detect by standard diagnostic tools including mammography and ultrasound (Hogan et al. 2015; Li, Uribe, and Daling 2005). Other classes of breast carcinoma include six rare histological subtypes namely inflammatory, medullary, papillary, mucinous, tubular and comedo carcinomas, each of which show differences in clinical characteristics, for example inflammatory carcinomas show poor survival rates (Li, Uribe, and Daling 2005).

Another important classifier reported by pathologists is the tumor grade, which is determined histologically according to growth patterns and degree of differentiation in the tumor tissue, where grade 1 refers to a well-differentiated tumor closely resembling normal breast tissue and grade 3 refers to a poorly differentiated tissue (Rakha et al. 2010). The degree of differentiation is calculated on the basis of three morphological features which include degree of tubule formation, cellular pleomorphism, and mitotic count (Rakha et al. 2010).

Assessment of anatomical extent of the disease is performed according to the TNM classification system which basically characterizes breast cancer on the basis of three major parameters, i) T denotes the features of the primary tumor, mostly size and local invasion and varies from T1-T4 with ascending order of severity; ii) N denotes the presence of disseminated cancer cells (DCCs) or micro-metastases in the lymph nodes, whereby N0 implies no involvement of lymph nodes, while N1-N3 indicate spread in increasing order; iii) M indicates

the absence or presence of distant metastasis, which could be either M0 or M1, respectively (Rosen and Sapra 2020). The combination of the above parameters defines the stage of the disease (Rosen and Sapra 2020). Stage 0 is assigned to non-invasive cancers like DCIS, stage I-III are considered early breast cancer without distant metastasis and stage IV refers to metastatic breast cancer. More recently, tumor grade and biomarkers including receptor status has also been incorporated in the staging system (O'Sullivan et al. 2017; Giuliano, Edge, and Hortobagyi 2018; Cserni et al. 2018).

Breast tumors are clinically categorized depending on the expression of estrogen (ER), progesterone (PR) and human epidermal growth factor 2 (HER2) receptors. 1) ER, PR positive and HER2 negative cancers account for around 70% of all breast cancers; 2) HER2 enriched cancers, account for 15-20% of tumors; 3) Around 15% tumors do not express any of these cell surface receptors and are thus called triple-negative breast cancers (Waks and Winer 2019; Howlader et al. 2014). Hormone receptor positive status is identified by the expression of ER or PR in at-least 1% of the tumor cells (Hammond et al. 2010), found by immunohistochemistry (IHC). HER2 receptor status is also determined by IHC; for a positive status >10% tumor area must stain for HER2. In uncertain cases, this method is complemented by in-situ hybridization to identify amplification of the HER2 gene (Wolff et al. 2013).

Gene expression studies led to molecular subtype classification (Perou et al. 2000; Sorlie et al. 2001), which largely identified four clinically relevant subtypes of breast cancer namely luminal A, luminal B, HER2-enriched, and basal-like each of which has distinct prognosis, treatment strategies and response to therapies (Harbeck and Gnant 2017). The hormone receptor (HR) positive cancers, also known as luminal-like cancers are further classified as Luminal A and Luminal B, both showing expression patterns similar to the luminal cells in the breast epithelium (Sorlie et al. 2001). Luminal B cancers have higher expression of cell proliferation markers and lower expression of hormone receptors (Sorlie et al. 2001; Brenton et al. 2005; Prat et al. 2013). Ki67 is a cell proliferation marker and is used to clinically identify luminal B cancers by immunohistochemistry (Ades et al. 2014). Next, basal-like tumors have a gene expression pattern similar to basal/myo-epithelial cells of the breast epithelium (Sotiriou and Pusztai 2009). Around, 80% of triple-negative breast cancers (TNBCs) are basal like (Weigelt, Baehner, and Reis-Filho 2010; Prat et al. 2015). Another major subtype belonging to the TNBCs group include claudin-low tumors, which display enhanced mesenchymal and stem-like properties (Foulkes, Smith, and Reis-Filho 2010).

Furthermore, on the basis of distinct copy number alterations (CNAs) and expression profiles, up to ten different subtypes have been reported (Curtis et al. 2012). Thus, breast cancer is highly heterogenous and demands a personalized approach to diagnostics and therapy.

1.2.2. Breast cancer treatment

Primary breast cancer is treated loco-regionally by surgical resection of tumor and removal of adjoining lymph nodes, which is sometimes followed up with post-operative radiation therapy (Waks and Winer 2019). Systemic therapy includes cytotoxic chemotherapy and targeted therapies which are determined on the basis of the diagnosed molecular subtype. For example, HR positive breast cancers are treated by downregulation of ER receptor signaling (endocrine therapy) with tamoxifen or aromatase inhibitors, which is sometimes combined with

chemotherapy if the cancer is luminal B or has a higher nodal status (Waks and Winer 2019), trastuzumab, an anti-HER2 antibody is prescribed for HER2-positive breast cancer with chemotherapy, sometimes in combination with endocrine therapy if HR positivity exists (Slamon et al. 2011; Romond et al. 2005). TNBCs however cannot benefit from targetted therapies due to the lack of ER, PR or HER2 receptors and only rely on chemotherapy for tumors larger than 5mm (Waks and Winer 2019).

Unfortunately,metastatic breast cancer is considered incurable with current treatment regimes and treatment is mostly palliative (Harbeck and Gnant 2017). Systemic therapy is usually prescribed to manage metastatic disease and very rarely locoregional intervention takes place (Waks and Winer 2019). Metastic relapse is treated by sequential addition of new therapeutic agents, for example metastatic HR positive/HER2 negative disease, is treated at first with endocrine therapy along with cyclin-dependent kinase (CDK) 4/6 inhibitors. After development of resistance, chemotherapeutic agents are introduced sequentially to minimize toxicity (Waks and Winer 2019). Metastatic HER2 positive cancers are treated with chemotherapy with continuation of HER2 targeted therapy (von Minckwitz et al. 2009). Metastatic TNBCs (without BRCA1/2 mutations) are also treated with sequentially varying chemotherapeutic agents (Waks and Winer 2019). Some therapeutic decisions are also based on the genomics of the cancer cells, for example BRCA1/2 are tumor supressor genes and approximately 5% of breast cancer patients have mutations in these genes, and BRCA1/2 mutated cells can be targeted by FDA approved poly(adenosine diphosphate-ribose) polymerase (PARP) inhibitors which leads to cell death (Robson et al. 2017). Immunotherapy using anti-PDL-1 antibodies has also been used for treatment of metastatic breast cancers having microsatellite instability (Waks and Winer 2019).

1.2.3. Breast cancer prognosis and survival

Multiple factors influence the prognosis and survival of breast cancer patients, mainly histological type, grade, stage and molecular subtypes. Breast cancer specific mortality (percentage of deaths after diagnosis) for DCIS (Stage 0) and small invasive breast cancer (node negative) patients after 15 years of diagnosis was 2% and 4.1%, respectively (Giannakeas, Sopik, and Narod 2018). Histological grade is also an important parameter to predict survival as the ten year survival rate (percentage of patients alive) of breast cancer patients among different grades with identical TNM stages shows significantly different prognosis where higher grade corresponded to poor prognosis (Schwartz et al. 2014). Moreover, the higher the stage at diagnosis the poorer the survival (Polyak and Metzger Filho 2012). Ten years after diagnosis the survival rates for early (Stage 0-I), locally advanced (stage II-III) and stage IV breast cancer were 94%, 75% and 9%, respectively (Vondeling et al. 2018). Additionally, molecular subtypes display different survival, which is highest for luminal A, followed by luminal B and HER2-enriched cancers (Fallahpour et al. 2017). Triple-negative breast cancers have the poorest prognosis of all subtypes (Fallahpour et al. 2017; Haffty et al. 2006) and there is a sharp increase in relapse within three to five years of diagnosis (Foulkes, Smith, and Reis-Filho 2010). Apart from the clinical parameters, various other factors like genetic predisposition, age at diagnosis, socio-economic condition, and lifestyle choices also contribute to differences in incidence and overall survival rates of breast cancer (Vondeling et al. 2018; Oh et al. 2015; Foulkes, Smith, and Reis-Filho 2010; Shariff-Marco et al. 2015).

Around 90% of cancer related deaths are caused by metastatic disease (Brabletz et al. 2013). Although, advances in diagnostics have led to improved early detection of primary tumors so that more than 90% of breast cancers are now diagnosed at an early stage (Waks and Winer 2019), approximately 30% of women diagnosed with early stage breast cancer would relapse months to decades after curative surgery (O'Shaughnessy 2005; Tjensvoll et al. 2019). Unfortunately, the five year survival rate of metastatic breast cancer to date is a dismal 25% (Rabbani and Mazar 2007; Valastyan and Weinberg 2011). Median overall survival of metastatic breast cancer varies from approximately 1 year in TNBCs to 5 years in HR positive and HER2 positive subtypes (Waks and Winer 2019).

1.3. Metastasis

Metastasis is the process of dissemination of cancer cells from the primary tumor site to sucessful colonization of distant organs (Figure 2) and is the primary cause of cancer lethality (Valastyan and Weinberg 2011). It is a dynamic and multistep process in which the tumor cells enter circulation (blood or lymph) by various mechanisms like collective migration, undergoing epithelial to mesenchymal transition (EMT) or by passive shedding (Bill and Christofori 2015). Next the circulating tumor cells (CTCs) need to survive in the vasculature amid shear forces, anoikis and immune surveillance (Al-Mahmood et al. 2018; Blomberg, Spagnuolo, and de Visser 2018). CTCs are found associated with platelets which guard them from the immune system (Gay and Felding-Habermann 2011). Finally, the CTCs extravasate into distant organs or eventually get trapped in small capillaries (Bill and Christofori 2015). These extravasated cancer cells are then found in the distant sites as single cells or a micrometastasis which can remain quiescent (Valastyan and Weinberg 2011) or proliferate after undergoing mesenchymal to epithelial transition leading to colonization (Bill and Christofori 2015).

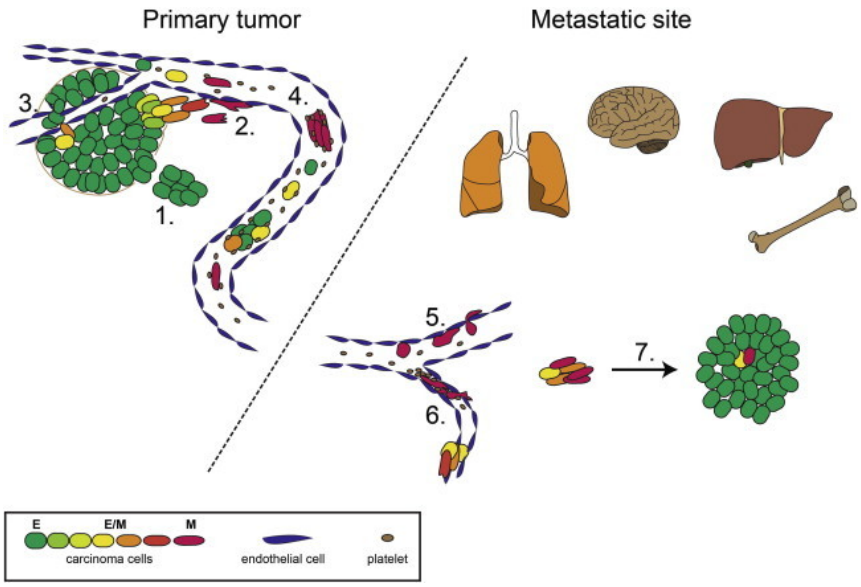


Figure 2: The process of metastasis. 1, Collective invasion of cells into surrounding tissue; 2, Single cells that underwent EMT enter circulation; 3, Passive shedding of metastatic cells; 4, Circulating tumor cells are found associated with thrombocytes in the circulation protecting them from shear forces and enhancing their sensitivity towards chemokine gradients (Al-Mahmood et al. 2018); 5, At distant sites circulating tumor cells extravasate through interactions with endothelial cells or; 6, get stuck in thin capillaries at distant sites; 7, The tumor cells start proliferating and colonizing by a process called mesenchymal to epithelial transition (MET). In the legend, E, epithelial characteristics; E/M having both epithelial and mesenchymal characteristics; M, mesenchymal characteristics. Figure taken from (Bill and Christofori 2015)

Breast cancer cells preferentially metastasize to bones, liver, brain and lung. Such preferential colonization can be seen in other cancers types as well, for example, prostate and colorectal cancer metastasize predominantly to bone and liver, respectively (Chambers, Groom, and MacDonald 2002). Various theories have been published about the reason for organotropism. The 'seed and soil' hypothesis (Paget 1989) proposed that a cancer cell (seed) is able to colonize a distant organ (soil), when the host microenvironment provides favourable growth conditions. Various microenvironmental factors, which supported this hypothesis, were subsequently found. For example, breast cancer cells carry CXCR4, which is the receptor for the chemokine CXCL12 (also called SDF1) secreted in bone marrow, thus making bone marrow a preferred niche for breast cancer metastasis (Luker and Luker 2006; Eckhardt et al. 2012). Paget's hypothesis was later challenged by James Ewing, who proposed that circulatory patterns from primary site to distant organs dictated the pattern of metastasis (Chambers, Groom, and MacDonald 2002). However, many autopsy studies subsequently found that both the theories hold ground and contribute to tumor metastasis (Chu and Allan 2012).

1.3.1. Fate of tumor cells in ectopic sites

Cancer cells residing at distant organs are termed disseminated cancer cells (DCCs). DCCs must acquire adaptive mechanisms within the new environment to become proliferative and give rise to metastases (Nguyen, Bos, and Massague 2009). In breast cancer, metastasis can occur months to decades after removal of the primary tumor (Pan et al. 2017; Fallahpour et al. 2017). This could be explained by DCCs remaining in a state of dormancy which alters their metabolic profiles and enables evasion of apoptosis (Senft and Ronai 2016; Havas et al. 2017; van der Toom, Verdone, and Pienta 2016). Alternatively, DCCs could exist in steady-state proliferation, meaning that the proliferation rate is balanced by the rate of cell death induced by immune surveillance or lack of angiogenesis (Endo and Inoue 2019). Interestingly, recent study in our lab showed that around 50% of DCCs in early breast cancer patients are in a state of proliferation, thus supporting this hypothesis (Irlbeck 2019).

Various factors contribute to the survival of DCCs and latency of breast cancer. Immune system (Suhail et al. 2019), anti-tumor therapy (Chatterjee and van Golen 2011) and the endothelial cells of the vasculature have been found to assist in keeping tumor cells dormant (Ghajar et al. 2013; Chatterjee and van Golen 2011; Sosa, Bragado, and Aguirre-Ghiso 2014). Bone marrow provides a permissive microenvironment for the DCCs to lay dormant until the arrival of an opportune moment to proliferate and colonize (Psaila and Lyden 2009; Psaila et al. 2006; Gomis and Gawrzak 2017; Linde, Fluegen, and Aguirre-Ghiso 2016). One such moment could be stochastic removal of immune surveillance that can bring dormant DCCs into proliferation (Malladi et al. 2016). However, it is not yet clear which combinations of genomic/epigenomic and microenvironmental properties lead to DCCs giving rise to lethal metastasis.

1.3.2. Models of metastasis

The linear progression model (Figure 3) of cancer evolution, holds that the metastasis founder cells (MFCs) evolve mostly within the primary tumor by sequentially acquiring somatic mutations (Lambert, Pattabiraman, and Weinberg 2017) and are among the most advanced clones which disseminate late during the progression of the primary tumor (Valastyan and Weinberg 2011). However, various clinical observations and data from animal models have hinted towards a parallel evolution of primary tumors and metastases.

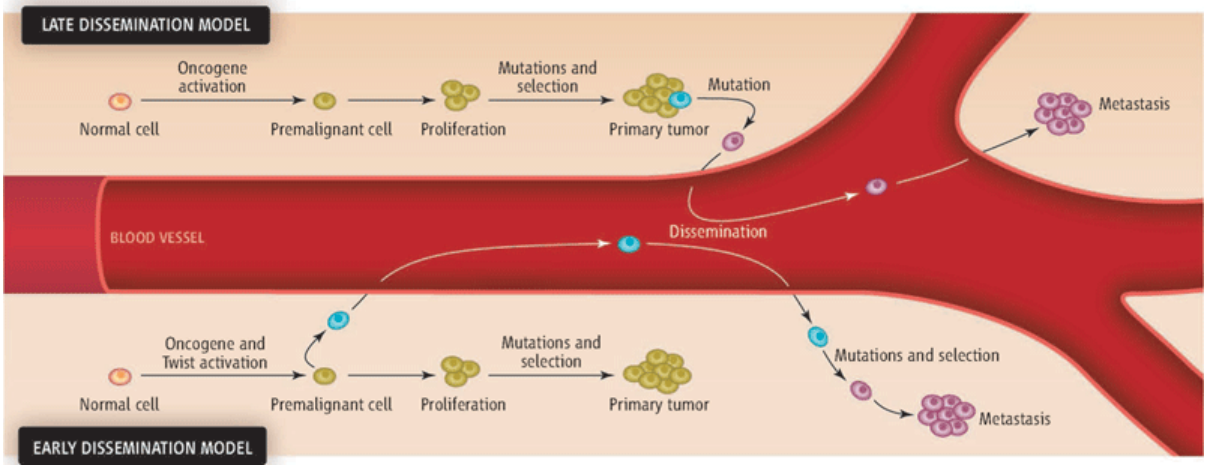


Figure 3: **Linear vs. parallel progression.** Top: The late dissemination model proposes that DCCs evolve largely within the primary tumor, gradually acquire somatic mutations there and resemble the most advanced clone in the primary tumor. In this scenario the DCC's genomic profile must be similar to that of the primary tumor and therapy decisions based on the primary tumor should curb relapse. Bottom: In the early dissemination model, DCCs leave the primary tumor at a genomically immature state and evolve outside the primary tumor by acquiring somatic mutations at the distant sites. In such a scenario the metastases could be irresponsive to the therapy strategies based on primary tumor characteristics. Figure adapted from (Klein 2008)

Many studies from as early as the 1950s pointed out that the late dissemination model of MFCs could not explain the growth rates of metastasis (Collins, Loeffler, and Tivey 1956; Klein 2009). By assessing the growth kinetics from mammography studies it was shown that manifestation of metastasis would take six to twelve years, if cancer cells disseminated from primary tumor at an advanced stage shortly before surgery, however median metastasis free survival of patients with tumors <2 cm is 35 months, which indicated that primary tumors and metastasis must develop in parallel (Klein 2009). Also, metastases are detected at diagnosis in 5% of patients with small primary (T1 and T2) tumors (Engel et al. 2003) and in cancer of unknown primary (CUP) where there is no clinically detectable primary tumor (van de Wouw et al. 2002). Furthermore, no significant difference was found in the tumor size and the presence of DCCs in bone marrow of breast cancer patients (Klein and Holzel 2006). Interestingly, DCCs were also observed in bone marrow of ductal carcinoma in-situ (DCIS) patients, which by definition is considered as a precursor of invasive breast cancer (Banys et al. 2012; Sanger et al. 2011; Husemann et al. 2008).

According to the SEER study involving 108,196 breast cancer patients, where decades long follow-up data is available, 54.1% of DCIS patients who died of breast cancer did not have an invasive in-breast recurrence prior to death (Narod et al. 2015), which indicates that DCIS cells have the intrinsic abilities to migrate and colonize. This phenomenon was also

demonstrated in mouse models of breast cancer where the majority of the metastases were generated from the pre-malignant tumors (Hosseini et al. 2016; Harper et al. 2016; Husemann et al. 2008). A similar demonstration of early dissemination was also seen in melanoma and pancreatic cancer models (Muzumdar et al. 2016; Eyles et al. 2010).

Finally, genomic analysis of the disseminated cancer cells in carcinoma patients from our laboratory and others confirmed that dissemination of cancer cells to distant sites occurs early and long before the diagnosis of a primary tumor (Schardt et al. 2005; Klein et al. 1999; Klein 2009) suggesting a parallel and non-linear evolution of primary tumor and metastases (Figure 3). Metastases might therefore be resistant to targeted therapies based on the biomarkers of the primary tumor (Klein 2008). Hence understanding the mechanisms of parallel evolution by analysis of the residual disease (DCCs) would improve therapy decisions.

1.4. Disseminated Cancer Cells

1.4.1. Detection of DCCs

Carcinoma cells can be detected in the bone marrow, lymph nodes and blood by immunocytochemical staining against epithelial markers, since these tissues are of mesenchymal origin and lack epithelial cells (Klein 2013; Schlimok et al. 1987). Many antibodies targeting various epithelial cell markers have been tested over the years, of which monoclonal antibodies against cytokeratins (CK) have been found to reliably detect DCCs (Pantel and von Knebel Doeberitz 2000). Using a consensus protocol where 2 million bone marrow cells were screened per patient, one to ten CK-positive cells can be found in approximately 30% of M0 patients (Fehm et al. 2006).

The epithelial cell adhesion molecule (EpCAM) is another common epithelial marker used for detection of DCCs (Schlimok et al. 1987). Although EpCAM is not as specific as CK, being a cell surface marker it allows isolation of live cells for transcriptomic analysis (Guzvic et al. 2014; Klein, Seidl, et al. 2002). EpCAM is also used for enrichment of tumor cells in bone marrow prior to CK staining (Woelfle et al. 2005). The FDA approved CellSearch system which enumerates CTCs in blood, is based on ferromagnetic enrichment of CTCs using anti-EpCAM antibody (Allard et al. 2004; Cristofanilli et al. 2005).

1.4.2. Prognostic significance of detected DCCs

Multiple studies have shown that the detection of DCCs in bone marrow leads to poor prognosis in carcinoma patients (Tjensvoll et al. 2010; Domschke et al. 2013; Hartkopf et al. 2014; Janni et al. 2011). In breast cancer, approximately 30% of patients were found to be DCC-positive, and among the patients with small tumors (<2 cm and lymph node negative), DCC-positive patients had more than three times increased breast cancer specific mortality rates within the first five years of diagnosis than the DCC negative counterparts (Braun et al. 2005). Recently, a multi-centre study involving 10,320 early breast cancer (T1–4, N0–2, M0) patients showed that detection of DCCs in bone marrow (27.4% of all patients) was an independent prognostic marker for reduced overall survival and distant disease-free survival (Hartkopf AD 2018 ; Hartkopf et al. 2014).

1.4.3. Genomic analysis of DCCs

CK positive cells in the bone marrow were found to harbor tumor specific chromosomal aberrations by fluorescence in-situ hybridization experiments, thus proving their malignant origin (Muller et al. 1996; Klein 2000). Single cell profiling by comparative genome hybridization (CGH) of DCCs from breast, prostate and oesophageal cancer patients without clinically detectable metastases revealed substantial heterogeneity (Klein, Blankenstein, et al. 2002; Schardt et al. 2005; Schmidt-Kittler et al. 2003; Stoecklein et al. 2008; Weckermann et al. 2009). M0 DCC genomes show early genomic separation from the predominant clone of the primary tumour (Klein 2013). Characteristics of most M0 DCC genomes include fewer chromosomal aberrations compared to M1 DCC genomes and absence of typical cancer type specific aberrations and point mutations (Weckermann et al. 2009; Schmidt-Kittler et al. 2003; Klein, Blankenstein, et al. 2002; Irlbeck 2019). Data from breast cancer patients can be seen in Figure 4. Advanced, PT specific gains and losses like in 8q, 8p are missing in majority of the M0 DCCs (Hosseini et al. 2016).

In contrast, DCC genomes are highly similar in patients with clinical metastases, suggesting expansion of an aggressive clone and a preference for a shared state of genomic changes (Klein 2013). It was observed that this acquired state of genomic aberrations is relatively stable, and is maintained in patient DCCs even after several cycles of chemotherapy (Klein, Blankenstein, et al. 2002; Klein 2013). This preferred genomic state is also seen in peripheral blood derived CTCs from M1 patients (Magbanua et al. 2013; Magbanua et al. 2012; Polzer et al. 2014). Interestingly, the genomes of DCCs and CTCs from M1 patients resemble those found in corresponding primary tumours (Weckermann et al. 2009; Schmidt-Kittler et al. 2003; Klein, Blankenstein, et al. 2002; Heitzer et al. 2013). The genomic difference between DCCs isolated from M0-stage patients and M1-stage patients raises questions on the evolution of systemic cancer. Are there some cell intrinsic properties that could potentiate the genomically immature M0 DCCs to transition to M1 like genomes?

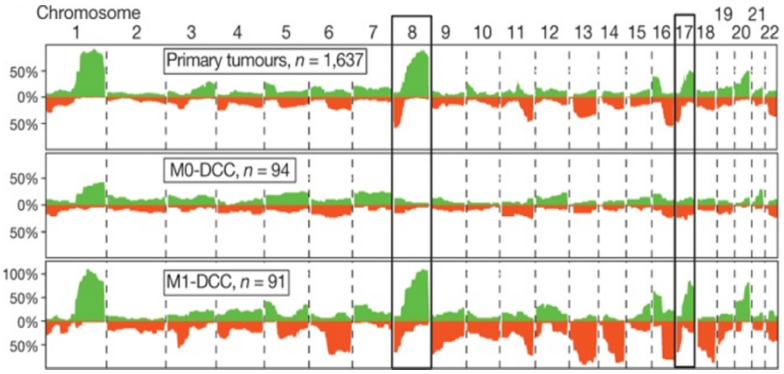


Figure 4. M0-DCCs are genomically immature. Copy number alterations found in primary tumors of breast cancer (n=1637 from Progenetix database) compared to that of M0 DCCs (n=94) and M1-DCCs (n=91). The ordinate depicts the percentage of cells with aberrations at the given chromosomal locus. Figure adapted from (Hosseini et al. 2016)

In the M0 stage, when the primary tumor has been surgically removed, M0 DCCs must overcome several obstacles to evolve to the M1-DCC-like karyotype to give rise to lethal metastasis. Matched primary tumors and single cell CTC analysis of metastatic breast cancer patients indicate that important driver mutations of breast cancer like *HER2 or PIK3CA* could be acquired outside the primary tumor (Polzer et al. 2014). *HER2* amplification was also found in some DCCs while their matched primary tumors did not have the amplification (Schardt et al. 2005). In mouse models of breast cancer (Hosseini et al. 2016) and human melanoma

(Werner-Klein et al. 2018) it was shown that typical driver alterations required for colonization, are acquired outside the primary tumor. Thus DCCs which are capable of colonization must be characterized, to identify early events in the evolution of metastasis which could be relevant for therapeutic intervention (Schardt et al. 2005).

It is noteworthy that the current data on genomic aberrations of M0-DCCs and M1-DCCs do not comprise longitudinally collected samples, i.e. no matched pairs of DCC from both M0 and M1 stage after detection of systemic metastasis of one individual patient have been obtained and analyzed so far. Phylogenetic analysis of longitudinally collected samples could help identify the metastasis founder cells (MFCs).

1.5. STR based cell lineage tracing

As multicellular organisms are products of multiple cell divisions of one single cell, the 'zygote', cell lineage tracing could allow identification of progenies of putative founder cells enabling us to understand various concepts in development and diseases like cancer (Shapiro 2018). While lineage tracing by direct observation is possible in the nematode *C.elegans* (Sulston et al. 1983; Sulston and Horvitz 1977) and transplantation and genetic engineering based invasive maneuvers could be performed in animal models (Kretzschmar and Watt 2012; Frieda et al. 2017; McKenna et al. 2016; van de Moosdijk et al. 2017), in human patients one could only probe retrospectively (Shapiro 2018). Human cell lineage reconstructions can be made by tracing somatic mutations which naturally occur in genomic loci during cell divisions. Such genomic regions include short tandem repeats (STRs), single nucleotide polymorphisms (SNPs), L1 retro transposable elements and copy number alterations (Behjati et al. 2014; Evrony et al. 2015; Lodato et al. 2015; Woodworth, Girskis, and Walsh 2017)

STRs - also called microsatellites - consist of repeats of one to six base pairs, which are frequently mutated during mitosis, due to polymerase slippage at replication (Frumkin et al. 2005). These mutations can be linked to cell divisions because each progeny acquires a distinct mutational signature (Frumkin et al. 2005). Among all naturally occurring somatic mutations, STRs possess one of the highest mutation rates of around 10^{-3} to 10^{-5} per locus per generation in humans (Willems et al. 2016). The higher the mutation rate, the lower is the number of genomic loci required to be sampled for obtaining reliable cell lineage trees (Frumkin et al. 2005). Owing to this fact, lineage relationships of cells with microsatellite instability are easier to resolve due to high STR mutation rates (Frumkin et al. 2005).

Most STRs are present in non-coding regions of the genome hence these mutations are silent and do not affect the phenotype and are not under selection pressure, unlike driver mutations in cancer (Frumkin et al. 2005). STR based cell lineage reconstructions have been applied to questions of development (Reizel et al. 2011; Reizel et al. 2012; Salipante et al. 2010) and tumor evolution (Frumkin et al. 2008) in mouse models. In human cancers like leukemia, STR based phylogenies have revealed diverse mechanisms for relapse initiation (Shlush et al. 2012). Thus, to identify MFCs in breast cancer patients, STR based phylogenetic analysis was applied.

1.6. Aims of the thesis

In order to identify the MFCs and obtain novel insights into the evolution of metastatic breast cancer, we strived to obtain cell lineage trees by longitudinal tracking of breast cancer patients progressing from M0-stage to M1-stage disease (Figure 5). Hence we had two aims; i) Establishment of a cohort of breast cancer patients with triplet samples of primary tumor, M0-DCCs at the time point of surgery and M1-CTCs at a time point after progression to metastatic disease; ii) STR based phylogenetic analysis of cells from longitudinally acquired samples of individual patients.

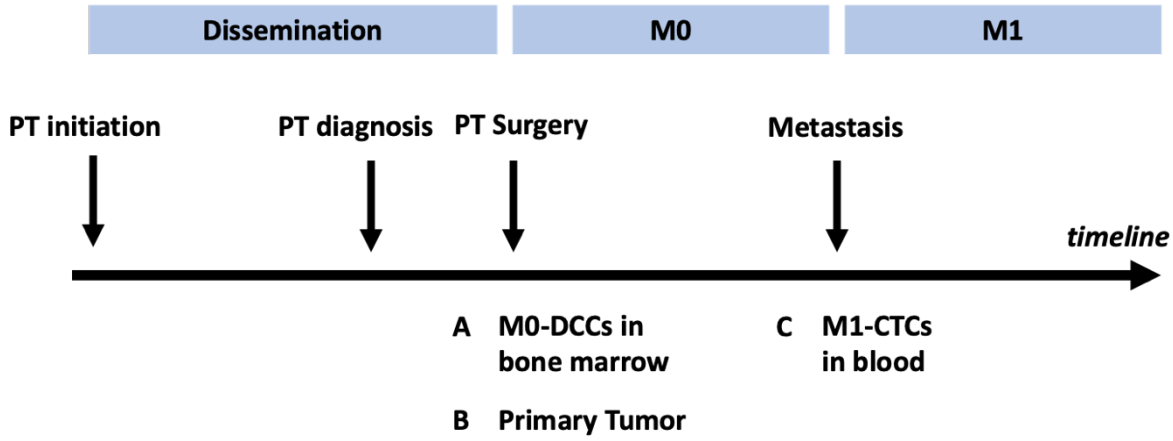


Figure 5: Longitudinal tracking of breast cancer patients. A. Bone marrow is sampled at the time of primary tumor surgery for isolation of M0 DCCs. B. Primary tumor is analyzed. C. After the patient is diagnosed with metastasis, blood is sampled for isolation of metastatic CTCs.

2. Methods

2.1. Patient samples and collaborations

2.1.1. Clinical cooperation

To generate the triplet cohort of patients we collaborated with the Department of Gynecology, University of Tübingen, owing to their huge archive of BM slides in collaboration with Dr Andreas Hartkopf. For establishing the protocol several unrelated patients bone marrow cytopsins were tested which, also included samples from Department of Gynecology and Obstetrics, LMU Munich in collaboration with Dr Nadia Harbeck.

2.1.2. Inclusion criteria

Inclusion criteria for the study were i) Female breast cancer patients, ii) M0 at primary tumor surgery, iii) Detection of CK positive cells in bone marrow aspirate at primary surgery, iv) bone marrow derived archived frozen cytospins available for isolation of DCCs, v) progression to M1 at the time of the study (2016-2019).

2.1.3. Ethics

Written informed consent was obtained from all patients involved in the generation of the triplet cohort for lineage tree analysis. The study was approved by the local ethics committee of University of Regensburg (ethics vote numbers 07-079, 18-948-1) and University of Tübingen (ethics vote number 535/2016BO2).

2.1.4 Bioinformatics Collaborations

Cell lineage trees were generated in collaboration with Dr Ehud Shapiro, Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Israel.

2.2. Single cell isolation techniques

2.2.1. Standard Micromanipulation

Single cells were isolated from adhesion slides or cell suspensions using a micromanipulator (PatchMan NP2, Eppendorf) coupled to an inverted fluorescence microscope (Olympus IX81). Cells were aspirated with the help of a 30 μm diameter glass capillary as previously described (Klein et al. 1999) and transferred to a separate field (called picking field) in an 8 chambered slide containing 200 µl PBS. Finally, the cell is aspirated in 1 µl PBS from the pick field into a 200 µl PCR tube (Axygen, MAXYMum Recovery) containing 2 µl cell lysis buffer for whole genome amplification (see chapter 2.9).

2.2.2. Laser Microdissection (LMD)

2.2.2.1. Sample preparation for LMD

The entire slide undergoing LMD was stained in addition to the immunohistochemical staining of cells of interest. For counterstaining, the slides were incubated with Mayer's hematoxylin for 2 minutes followed by washing 2 times with 1X PBS (pH 7.4) for 2 minutes. The cuvette containing the slides was left in 4 °C for at least 30 min for bluing. Next, the slides were dehydrated by dipping into pre-chilled (-20 °C) solutions of 70%, 90% ,100% ethanol and acetone for 5 seconds each (dehydration at RT should be avoided). Slides were used for LCM immediately after dehydration.

2.2.2.2. LMD

LMD was performed on Zeiss PALM Microbeam system consisting of a 355 nm pulsed frequency tripled solid state laser which can be controlled by the PALM Robo software. The beam diameter could be focused to <1 μm which allowed precise microdissection of single cells, the higher the magnification of the objective, the smaller would be the beam diameter. The sample to be micro-dissected was placed on the PALM RoboMover stage (which could be controlled using the PALM Robo software), after selection of the desired area of interest on the software the robotic stage automatically moved to the selected area. Default factory settings were loaded on the software followed by calibration of the laser marker (to calibrate a point seen on the screen as the point where laser should hit the sample) according to manufacturer's instructions. Calibration of laser marker was repeated before microdissection of each cell to ensure precision. Cells of interest were micro-dissected with the laser cutting function with energy levels varying from 35-55% focused through a 40X objective to ensure sharp and precise cutting along the edges of the cells. Next two methods were possible for cell isolation, i) Laser Pressure Catapulting (LPC) and ii) Laser Cutting Assisted Micromanipulation (LCAM)

i) LPC

After laser cutting, single cells were catapulted into a silicon filled adhesive cap of a 200 µl PCR tube (Carl Zeiss) containing 4.5 µl of lysis buffer for WGA (see chapter 2.9.2). Due to the silicon filling in the cap the flight distance of a catapulted cell was reduced and due to the adhesive properties of silicon the cells would stick to the cap, thus augmenting the success of cell isolation. LPC energy was set to the minimum energy required to detach and successfully catapult the cells into the cap of the tube, which varied from 55-65% depending on the type of sample preparation. Various sample preparation methods were tested by Eva Maria Hecht, which were described in detail in her masters dissertation (Hecht 2016).

ii) **LCAM**

In this method, after laser cutting, cells were isolated from the slide using standard micromanipulation instead of LPC. A large amount of surrounding area was ablated using higher laser cutting energy (50-65%). Higher energy allows bigger cutting widths which is required to clear a larger area around the cell. In order to create space for contamination free standard micromanipulation, laser ablation was carried out in a radius of 25-50 µm around the

cell. (Supplementary Figure 1B). For standard micromanipulation the slide was dipped shortly in 1X PBS for rehydration, followed by incubation with a detergent cocktail (10 µl 10X OPA buffer, 2.6 µl 10% Tween, 2.6 µl Igepal, 69.6 µl PCR water) for 5 minutes. After a short wash in 1X PBS the slide was ready for standard micromanipulation as described in chapter 2.2.1.

2.3. Isolation of DCCs from frozen archived cytospins

2.3.1. Cytospin preparation

Frozen cytospin slides collected over the last decades at the Department of Gynecology, University of Tübingen were sent to our lab on dry ice for this project. The slides were prepared according to a previously published protocol (Bauer et al. 2000). Briefly, 10-20 ml of bone marrow (BM) was aspirated from the anterior iliac crest during primary tumor surgery and processed within 24 hours. MNCs were isolated by density gradient separation using Ficoll (Biochrom, Berlin, Germany,1.077 g/ml). The isolated MNCs were counted and resuspended in 1X PBS to obtain a solution containing 1.5 million cells/ml. To prepare cytospins, 1 ml cell suspension was spun down at 1600 rpm for 2 min onto a glass slide using a cytocentrifuge (Hettich, Tuttlingen, Germany). The slides were allowed to dry at room temperature and thereafter stored at -20 °C.

2.3.2. Detection and isolation of DCCs

For detection of DCCs, frozen cytospins were stained for epithelial marker cytokeratin as described previously (Klein, Blankenstein, et al. 2002) with some modifications. Frozen cytospins were allowed to equilibrate to room temperature and incubated with blocking solution (10% AB serum/1X PBS) for 30 min. The blocking solution was discarded followed by incubation with A45-B/B3 antibody (Micromet, Munich, Germany) at a concentration of 2µg/ml. After 60 min, the primary anitbody was discarded and the slides were washed 3 times with 1X PBS. Next, slides were incubated for 30 minutes with secondary antibody anti-mouse AP-Polymer solution followed by washing 3 times with 1X PBS. For development, BCIP/NBT solution (AP Conjugate Substrate Kit; Bio-Rad Laboratories) was prepared according to manufacturer's instructions with 0.5% levamisole followed by incubating the slides for 12 min. After development the slides were washed again for 3 times with 1X PBS. Fixation with PFA was not performed for slides undergoing LCM. Screening of positive cells was performed using upright Olympus microscope (BX43). For cytopsins undergoing LCM, screening was performed under the PALM microbeam microscope (PALM Zeiss) where the location of positive cells could be saved as elements in the slide specific elements file on the PALM Robo Software. After screening, the slides were counterstained and dehydrated (chapter 2.2.2.1) followed by isolation by LCAM (chapter 2.2.2.2).

2.4. Isolation of vimentin positive cells

2.4.1. Preparation of adhesion slides from fresh blood

MNCs from fresh blood were isolated as previously described (Klein, Blankenstein, et al. 2002). Briefly, 5-10 ml of blood was washed with Hanks buffer solution and subjected to density gradient separation with 60% Percoll. The interphase cells were collected and

counted. Cells were diluted to a concentration of 500,000 cells/ml in 1X PBS and 1ml was dispensed per adhesion slide (Menzel, Braunschweig, Germany) at a density of $0.16*10^6$ cells per 14 mm diameter spot. The cells were allowed to sediment for 1 hour followed by discarding the supernatant PBS. The slides were allowed to air dry at RT and then stored at -20 °C.

2.4.2. Detection and isolation of vimentin positive cells

Adhesion slides from healthy donor PBL or frozen archived patient cytospins were stained for mesenchymal marker vimentin using monoclonal Mouse Anti-vimentin antibody (clone V9, Agilent DAKO). The rest of the staining protocol is as in chapter 2.3.2. Vimentin positive cells were isolated by standard micromanipulation or LPC as described in chapters 2.2.1 and 2.2.2 respectively.

2.5. Isolation of MCF10a spiked-in mouse BM-MNCs

2.5.1. Preparation of cytospins from mouse BM-MNCs with spiked in MCF10a cells

Balb/C BM-MNCs were isolated as previously published (Hosseini et al. 2016). MCF10a cells were cultured as described (Grujovic 2019), harvested at 70-80% confluence, resuspended in 1X PBS and counted. MCF10a cells were spiked into Balb/C BM-MNCs in a ratio of 1:100 to obtain a final suspension of $1.5*10^6$ cells/ml. Cytospins were made using the Hettich Centrifuge Rotofix 32A by centrifuging 100 µl of cell suspension at 1000 rpm for 5 minutes over a spot of 30 mm^2. After centrifugation the supernatant was carefully removed with a Pasteur pipette. The slides were allowed to dry at RT overnight. After drying, the slides were stored at -20 °C.

2.5.1. Detection and isolation of MCF10a cells from mouse BM-MNC cytospins

Frozen cytospins were allowed to equilibrate to room temperature and incubated with blocking solution (10% AB serum/1X PBS) for 30 min. The blocking solution was discarded followed by incubation with biotinylated A45-B/B3 antibody and biotinylated MOPC-21 antibody (isotype control) at a dilution of 1:150. Meanwhile ABC-AP complexes (VECTASTAIN® ABC-AP KIT) were prepared according to manufacturer's instructions. After 45 min, the biotinylated antibodies were discarded and the slides were washed 3 times with 1X PBS. Next, slides were incubated for 30 minutes with ABC-AP complexes followed by washing 3 times with 1X PBS. For development, BCIP/NBT solution (AP Conjugate Substrate Kit; Bio-Rad Laboratories) was prepared according to manufacturer's instructions with 0.5% levamisole followed by incubating the slides for 12 min. After development the slides were washed again for 3 times with 1X PBS. For isolation of MCF10a, slides were screened on the PALM Microbeam microscope and the coordinates of MCF10a cells were saved. Next the cytopsins were counterstained by Mayer's hematoxylin and dehydrated (chapter 2.2.2.1) followed by isolation by LCAM (chapter 2.2.2.2).

2.6. Isolation of CTCs from blood

2.6.1. Detection

Two samples of 7.5 ml blood were collected from patients in Cellsave tubes at University of Tübingen, sent to our laboratory at RT and analyzed within 72 hours of sample collection. Cellsearch was used for detection and enumeration of CTCs as described previously (Polzer et al. 2014). Briefly, CTCs were captured by ferrofluid coated with EpCAM-antibody, followed by staining for cytokeratin, CD45 and DAPI by an automated workflow. It is to be noted that for the first CellSave tube, the assay was conducted according to manufacturer's instruction but from the second tube, plasma was isolated first, followed by diluting the cellular fraction with dilution buffer to reach the original volume of 7.5 ml, which was then prepped according to manufacturer's instruction.

2.6.2. Cell extraction

When CTCs were detected in a Cellsearch run, cell suspension from Cellsearch cartridge was extracted into a 1,5ml Eppendorf LoBind tube using a 200 µl gel-loading pipette tip pre-coated in 2% BSA (to prevent cells from sticking to the tip, to avoid cell loss). Next, the cartridge was carefully rinsed with 325µl 1xPBS by pipetting on the inner walls repeatedly using a pre-coated gel-loading pipette tip for a total of two times. The entire fluid was collected in the 1,5ml Eppendorf LoBind tube. The sample was centrifuged with a swinging-bucket rotor centrifuge for 5 min at 1000 g, 20°C. The supernatant was discarded very carefully from above until approximately 100 µl was left in the tube. The cells are resuspended and immediately used for single cell picking.

2.6.3. Single cell isolation from suspension

The entire 100µl cell suspension containing CTCs was pipetted onto a chamber of an 8-chambered slide pre-coated with BSA. Next, 100µl PBS was used to rinse the tube containing cells and was added to the sample in the chamber to reach a final of 200 µl in one chamber. The cells were allowed to settle. Next, the sample was screened manually on the fluorescence microscope. Single cells positive for cytokeratin (Cy3), DAPI and negative for CD45 (APC) were isolated using standard micromanipulation as described in chapter 2.2.1. Additionally, two normal cells were picked per sample (APC positive, DAPI positive, Cy3 negative), a cell pool (blindly picking 1 µl of cell suspension) as positive control and 1 µl of PBS from the picking field was collected as reagent control for each sample.

2.7. Primary tumor material

2.7.1. Isolation of primary tumor single cells from cryo sections

Sections of 20-25 µm from cryopreserved PT tissues were placed on glass slides coated with poly L-lysine. The slides were immediately fixed in 0.5% PFA solution for 5 minutes followed by washing 3 times in 1X PBS for 3 minutes. The slides are then incubated with blocking solution (10% AB serum/1X PBS) for 30 minutes followed by staining for cytokeratin as described in chapter 2.3.2. After the development step, the section is incubated for 1h in a 1:1 mixture of collagenase A (33mg/ml) and hyaluronidase (10mg/ml) at 37°C in an oven for enzymatic digestion. To avoid drying of the tissue during incubation, the slide was placed in a pre-heated humid chamber in the oven (made by adding water to paper towels). After

incubation, the slide was shortly dipped into 1X PBS to wash away the enzymes. Next, 250-300 μl of 0.5M maltose solution was pipetted onto the slide and left for at-least 45 minutes to create a hypertonic environment for the cells. Next, the slide was scratched with a 200 μl pipette tip and the contents were collected in a microfuge tube already containing 250 μl of 0.5M maltose solution. The entire solution was pipetted vigorously several times to obtain a single cell suspension. For single cell picking, 50 μl of single cell suspension in maltose was added to 150 μl of pick PBS in one of the chambers of the 8 chambered slide pre-coated with BSA. Cell isolation was carried out as mentioned in section 2.6.3.

2.7.2. Isolation of primary tumor single cells from FFPE sections

2.7.2.1. Tissue deparaffinization and dissociation

For isolation of single cells from FFPE sections a protocol from Menarini Silicon Biosystems was used, now commercially available as DEPArray™ FFPE SamplePrep Kit. FFPE tissue section (40- 60 μm thick) was sealed inside a nylon biopsy bag. The sealed bag containing the tissue sample was placed in a 50ml falcon tube. The tissue was washed with zylene three times for 10 min each at RT for deparaffinization. To rehydrate, the tissue was washed with a decreasing ethanol series, 3 times with 100%, 3 times with 70% and 2 times with 50% ethanol, for 5 min per wash at RT. Next, the sample was washed once for 5 minutes with Milli-Q water at RT. For antigen retrieval, at first 1X HIAR buffer was added to the sample. After 5 min the 1X HIAR buffer at RT was changed with pre warmed (80 °C) 1X HIAR buffer and the sample was incubated at 80 °C in a water bath for 1 h. After incubation, the sample was allowed to cool to RT. Next 1X HIAR buffer was discarded and the sample was washed three times for 5 minutes with RPMI 1640 medium at RT under sterile conditions. To dissociate the tissue, RPMI 1640 medium was discarded and 10ml 1X Dissociation solution pre-warmed at 37 °C was added to the sample followed by incubation at 37 °C in a water bath for 45 min, with gentle agitation of sample after every 15 minutes. To stop dissociation, the sample was placed on ice, and was resuspended gently by pipetting with a 10ml serological pipettes. The dissociated sample was filtered through a 30μm mesh nylon filter into a 15ml conical tube and centrifuged at 1.000g for 5 minutes at 4 °C. The supernatant was discarded and the pellet was washed two times in 5ml ice-cold PBATw buffer by centrifugation at 1.000g for 5 minutes at 4 °C. Finally, the pellet was resuspended in 1ml ice-cold PBATw buffer and transferred to a 1,5ml Eppendorf protein LoBind tube. The cells were counted under fluorescence lamp in a Neubaur chamber by staining 10 μl of cell suspension with DAPI at a final concentration of 0.2 μg/ml.

2.7.2.2. Immunofluorescence staining

For immunofluorescence staining, 1 μl of anti-cytokeratin antibody (A45/Bb3) was added to 100 μl of ice-cold cell suspension (500.000 cells in PBATw buffer) and incubated for 1 h with gentle agitation every 15 minutes. Next 1ml of ice-cold PBATw was added to the cells and the solution was centrifuged at 1.000g for 5 minutes at 4 °C. This step was repeated for a total of two washes. Finally, the supernatant was discarded, and the cells were incubated with secondary antibody Alexa Fluor. 488 goat anti-mouse IgG1 (2mg/ml) diluted in ice cold PBATw buffer (1:500) for 30 minutes with gentle agitation every 15 minutes. Again, the cells were washed twice with 1ml ice cold PBATw buffer by centrifugation at 1.000g for 5 minutes at 4 °C. The supernatant was discarded and filled again with ice cold PBATw buffer, to which DAPI was added at a final concentration of 0.2 μg/ml and incubated for 15 minutes. The cell

suspension was again washed with ice cold PBATw buffer two times by centrifugation at 1.000g for 5 minutes at 4 °C. The cell pellet was finally diluted in 1 ml ice cold PBATw buffer and directly used for single cell isolation.

2.7.3. Isolation of primary tumor gDNA from FFPE tumor blocks

Tumor containing areas in PT blocks were determined by pathologist Dr Florian Weber and marked on hematoxylin and eosin sections (H and E sections). These sections were superimposed on the block to locate the tumor regions. Tumor regions were punched out using disposable biopsy punch (1.5mm). Next the punched-out regions were placed on UV sterilized glass slide, and the excess paraffin was cut out using a scalpel. Another UV sterilized glass slide was used to squeeze the tissue, followed by cutting the tissue into pieces in all directions with a scalpel. The tissue bits were transferred from the glass slide to a 1.5 ml tube using the scalpel. The tube was centrifuged shortly to collect all the tumor material at the bottom. Next, gDNA was isolated from the tumor material using QIAamp FFPE Tissue Kit according to the manufacturer's instructions. The isolated gDNA was quantified using NanoDrop ND-1000.

2.8. Isolation of outgroup cells

2.8.1. T cells and macrophages.

2.8.1.1. Preparation of adhesion slides from fixed blood

Blood samples from patients were collected in CellSave tubes (Menarini Silicon Biosystems). Approximately 7.5 ml of blood was washed with PBS in a 50 ml falcon tube and centrifuged at 200g for 10 min at 4 °C. The cell pellet was diluted in Hanks solution and centrifuged at 1000g, for 20 min at 4 °C on a 60% Percoll density gradient. The interphase was collected and washed in 1X PBS by centrifugation at 200g for 10 min at 4 °C and resuspended in 1X PBS, counted and transferred onto adhesion slides at a density of $0.16*10^6$ cells per 14 mm diameter spot. The cells were allowed to sediment on the slide for 1 hour at RT. The residual PBS was discarded and slides were allowed to air-dry overnight at RT and then stored at -20°C.

2.8.1.2. Detection and isolation of T cells and macrophages

Adhesion slides stored at -20 °C were thawed at RT and blocked with 1x PBS/10% AB-serum (Bio-Rad) for 30 min, followed by incubation with polyclonal rabbit anti-CD3 antibody (C7930, Sigma Aldrich) and monoclonal mouse anti-human CD68 antibody (clone KP1, M0814 Dako) in 1X PBS/10% AB-serum for T-cells and macrophages respectively, for 60 minutes each at room temperature. Next, slides were washed with 1x PBS and incubated with ready to use solutions of AP-polymer anti-rabbit (ZUC031-006, Zytomed Systems) and AP-Polymer anti-mouse (ZUC077-100, Zytomed Systems) respectively for 30 minutes at room temperature and developed using BCIP-NBT detection system (Bio-Rad Laboratories) with leavamisol (0.5%) for 12 mins. All steps of staining were carried out at RT.

2.8.2. OECs

2.8.2.1. Preparation of adhesion slides from buccal swabs

Buccal swabs collected from patients were immediately stirred into 1ml 1X PBS. The cell suspension thus obtained was pipetted onto adhesion slides. The cells were allowed to sediment for 1 hour at RT, and the residual PBS was discarded. The slides were allowed to air dry overnight at RT and then stored at -20 °C. These slides were prepared at the University of Tübingen and were shipped to us on dry ice.

2.8.2.1. Detection and isolation of OECs

Adhesion slides with buccal swabs were incubated with Mayer's hematoxylin for 2 min followed by washing in 1X PBS for 5 minutes. Next, the slide was dipped in fresh 1X PBS and was taken for picking by standard micromanipulation as described in chapter 2.2.1. Cells with a polygonal morphology with a distinct nucleus were isolated.

2.9. Whole Genome Amplification

2.9.1. Principle

WGA protocol used to amplify single cells or minute amounts of DNA was originally published as SCOMP (Stoecklein et al. 2002; Klein et al. 1999) and now commercially available as *Ampli*1™ WGA Kit. The method is based on a ligation mediated adaptor linker PCR (Figure 1) wherein the template DNA is deterministically fragmented using MseI restriction digestion, to obtain fragments that are around 150-1500 bp in length. Following restriction digestion, the DNA fragments are ligated to asymmetric double stranded adaptors (which are prepared separately in a step called pre-annealing). One of the oligonucleotide strands (ddMse11; 5'-TAACTGACAGddC-3') lacks phosphate group at the 5′-end which blocks its ligation to the Mse1 digested DNA fragments. This strand is later removed by denaturation. Since it also consists of a di-deoxy nucleotide at the 3′-end it is not able to prime amplification during PCR amplification of the fragments. The overhangs of the other strand (Lib1; 5'-AGTGGGATTCCTGCTGTCAGT-3'), is filled in by DNA polymerase during PCR amplification. The exponential PCR amplification continues with the help of excess Lib1 adaptors.

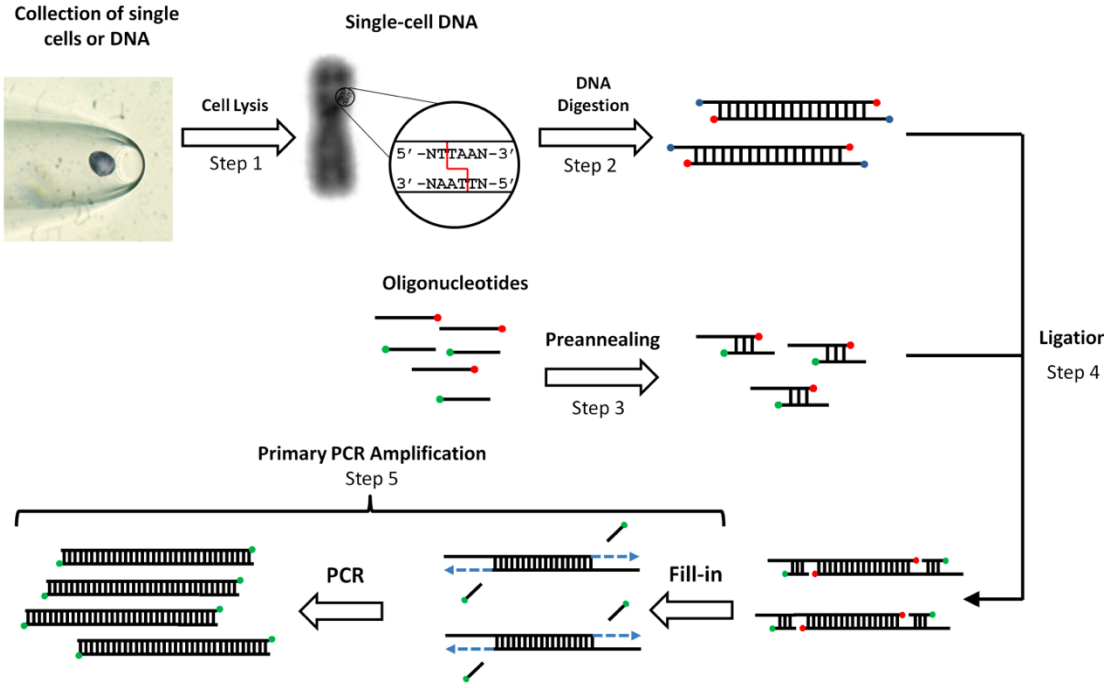


Figure 2: Principle of *Ampli*1™ WGA. The WGA protocol comprises of five steps including cell lysis, DNA digestion using MseI restriction enzyme, pre-annealing (where ddMse11 and Lib1 adaptors hybridize), ligation, where the double stranded adaptors ligate to digested DNA fragments, an PCR amplification, Chemical groups at the ends of the DNA fragments are indicated with coloured dots; blue, 5′ phosphate group; green, 3′ de-dioxy oligonucleotide; red, 3′-hydroxyl group. Figure adapted from (Czyz and Klein 2015).

2.9.2. Experimental Procedure

Composition of all master mixes are provided in Table 1. All steps were carried out in a PCR cycler. Single cells/cell pools were collected in lysis buffer containing proteinase-K (PK). The samples were incubated at 42 °C for 15 hours (for cell isolated by LCAM, LPC or FFPE) or for 10 hours (for cell picked from adhesion slides,cell suspensions). PK was inactivated at 80 °C for 10 min, and then allowed to cool to 4 °C. After PK digestion, the DNA was subjected to MseI restriction digestion by incubation at 37 °C for 3 h. Simultaneously, double stranded adaptors were prepared by pre-annealing at 65 °C and gradually decreasing temperature by 1 °C per minute till the sample reached 15 °C. Next ATP (10 mM) and T4 ligase (5U/µl) were added to the pre-annealing mixture. After 3 h of digestion, MseI was inactivated at 65 °C for 5 min and the sample was allowed to cool to 4 °C. The pre-annealing mixture containing ATP and T4 ligase was added to the MseI-digested samples, followed by overnight incubation at 15 °C for ligation. On the following day, the samples were incubated after addition of primary PCR master mix according to the amplification program in Table 2.

Master Mix	Constituents	*Volume per sample (µl)	**Volume per sample (µl)
Proteinase-K digestion	OPA 10x	0.2	0.5
	Tween 10 %	0.13	0.13
	Igepal 10 %	0.13	0.13
	Proteinase K (10mg/ml)	0.26	0.26
	PCR-water	1.28	3.48
MseI digestion	OPA 10x	0.2	--
	MseI 50,000 U/µl	0.2	0.25
	PCR-water	1.6	0.25
Pre-annealing	OPA 10x	0.5	0.5
	Lib1 100 µM	0.5	0.5
	ddMse11 100 µM	0.5	0.5
	PCR-water	1.5	1.5
Primary PCR	Buffer 1	3	3
	dNTPs 10 mM	2	2
	DNA Pol Mix	1	1
	PCR-water	34	34

Table 1: Composition of master-mix prepared for WGA.
***** Indicates amounts for standard WGA
** Indicates amounts for WGA of cells isolated by LPC

Step	Temperature	Duration	Cycles
Step 1	68.0 °C	3 min	
Step 2	94.0 °C	40 s	
Step 3	57.0 °C	30 s	
Step 4	68.0 °C	1:30 min	extend 1s/cycle
Step 5			GOTO 2, 14 times
Step 6	94.0 °C	40 s	
Step 7	57.0 °C	30 s	Increase 1.0°C/cycle
Step 8	68.0 °C	1:45 min	extend 1s/cycle
Step 9			GOTO 6, 8 times
Step 10	94.0 °C	40 s	
Step 11	65.0 °C	30 s	
Step 12	68.0 °C	1:53 min	extend 1s/cycle
Step 13			GOTO 10, 22 times
Step 14	68.0 °C	3:40 min	
Step 15	4.0 °C	infinite	

Table 2: WGA Primary PCR program

2.9.2. WGA of single cells isolated from FFPE sections

To improve the quality of DNA derived from FFPE sample, a DNA repair enzyme mix from NEB was used which constitute a mixture of various DNA repair enzymes (Taq DNA Ligase, Endonuclease IV, Bst DNA Polymerase, Fpg, Uracil-DNA Glycosylase, T4 Endonuclease V and Endonuclease VIII). The DNA repair step was introduced prior to MseI restriction digestion step in the WGA protocol. Single cells from FFPE tissue were collected in PK-lysis buffer, followed by incubation at 42 °C for 15 hours. Next, the DNA was treated with FFPE repair mix (NEB), by incubating at 37 °C for 20 min, followed by inactivation of the repair enzymes at 80 °C for 20 min. The samples were allowed to cool to 4 °C followed by MseI digestion (with half the quantity as standard WGA). Rest of the WGA protocol was similar as in chapter 2.9.1. The master mix composition for FFPE repair mix and MseI digestion mix is provided in Table 3.

Master Mix	Volume of constituents per sample
FFPE Repair	0.4 µl FFPE Repair buffer (10x) 0.1 µl FFPE Repair Mix
MseI digestion	0.5 µl PCR-water 0.1 µl OPA 10x 0.1 µl MseI 50,000 U/µl 0.8 µl PCR-water

Table 2: Composition of master-mixes for FFPE repair and modified MseI restriction digestion

2.9.3. Quality control (QC) PCR for WGA samples

Quality of WGA samples was assessed by a multiplex endpoint PCR consisting of primers (KRAS, CK 19, TP53 Exon2/3, D5S2117) to amplify four genomic regions corresponding to different MseI fragments, as previously described (Polzer et al. 2014) and commercially available as *Ampli*1™ QC kit. Primer sequences, amplicon lengths and corresponding MseI fragment length is provided in Table 3. Master mix for the multiplex PCR was prepared according to Table 4, 9 µl of the master mix was used for every 1 µl of template DNA (WGA product/reamplified WGA product). As negative control 1 µl of PCR water was used instead of template DNA, and a previously assessed good quality sample was used as a positive control. The multiplex PCR program is described in Table 5. The products of the PCR reaction were visualized by agarose gel electrophoresis (see chapter 2.9.5). The number of bands displayed on the gel corresponded to the genome integrity index (GII) of the sample. Samples with GII>2 were considered to be of good quality. Since KRAS corresponds to a small MseI fragment, it is usually amplified in the QC PCR if a cell was isolated, irrespective of its DNA quality. Absence of all the bands GII=0, thus means that the starting material might not have contained any cell.

Primer	Chromosomal Location	Sequence	MseI fragment length (bp)	Amplicon length (bp)
KRAS_91bp_5'	12p	ATAAGGCCTGCTGAAAATGAC	192	91
KRAS_91bp_3'		CTGAATTAGCTGTATCGTCAAGG		
*D5S2117_5'	5q	CCAGGTGAGAACCTAGTCAG	1376	140
D5S2117_3'		ACTGTGTCCTCCAACCATGG		
CK19 5'	6q	GAAGATCCGCGACTGGTAC	1146	621
CK19 3'		TTCATGCTCAGCTGTGACTG		
TP53 Exon2/3 5'	17p	GAAGCGTCTCATGCTGGATC	1374	301
TP53 Exon2/3 3'		CAG CCC AAC CCTTGTCCTTA		

Table 3: QC-Multiplex PCR Primers and amplicon description
***D5S2117** is a microsatellite marker, homozygous samples show one band, heterozygous samples show two bands but with a minor difference in length, 108-166 bp. These two bands would be counted as one band for calculating GII.

Reagent	Volume per sample
10x FastStart PCR Buffer (with 20mM MgCl2)	1 µl
Primer mix (4 µM per primer)	1 µl
dNTPs (from FastStart kit)	0.2 µl
BSA (20 mg/ml)	0.2 µl
FastStart Taq Polymerase (5 U/µl)	0.1 µl
PCR-water	6.5 µl

Table 4: Master mix composition for QC-Multiplex PCR

Step	Temperature	Duration	Cycles
Step 1	95.0 °C	4 min	
Step 2	95.0 °C	30 s	
Step 3	58.0 °C	30 s	
Step 4	68.0 °C	1:30 min	
Step 5			GOTO 2, 32 times
Step 6	72.0 °C	7 min	
Step 7	4.0 °C	forever	

Table 5: Multiplex QC PCR program

2.9.4. Quality control (QC) PCR for FFPE WGA samples

Since FFPE DNA is damaged and heavily fragmented, longer Mse1 fragments are scarcely present in WGA samples, owing to which FFPE samples have lower GII values. Therefore, different sets of primers targeting MseI fragments of various lengths, mostly including smaller and medium length MseI fragments were designed to judge the suitability of the FFPE DNA for downstream analysis. FFPE QC PCR consist of a total of 12 primer sets constituting three different multiplex PCRs. Primer sequences, amplicon lengths and corresponding MseI fragment length is provided in Table 6. Master mix for the multiplex PCR was prepared according to Table 7 and 9 µl of the master mix was used for every 1 µl of template DNA. As negative control 1 µl of PCR water was used instead of template DNA, and a previously assessed good quality sample was used as a positive control. The multiplex PCR program is as described previously in Table 5. The products of the PCR reaction were visualized by agarose gel electrophoresis (see chapter 2.9.5).

Multiplex	Primer	Sequence	MseI fragment length (bp)	Amplicon length (bp)
1	KRAS_91bp_5'	ATAAGGCCTGCTGAAAATGAC	192	91
	KRAS_91bp_3'	CTGAATTAGCTGTATCGTCAAGG		
	BRAF_Exon 15_5'	CTCTTCATAATGCTTGCTCTG	264	171
	BRAF_Exon 15_3'	TCCAGACAACTGTTCAAACTG		
	PIK3CA_Exon20 (HS2)_5'	TCTAGCTATTCGACAGCATGC	296	221
	PIK3CA_Exon 20 (HS2)_3'	TTGTGTGGAAGATCCAATCCAT		
	EGFR_Exon 19_5'	TCCTCGATGTGAGTTTCTGC	425	350
	EGFR_Exon 19_3'	ATGCCTCCATTTCTTCATCC		
	TP53_Exon 7_5'	GAGGCTGAGGAAGGAGAATG	724	400
	TP53_Exon 7_3'	AGTATGGAAGAAATCGGTAAGAGG		
2	NRAS_Exon 2_5'	ACACCCCCAGGATTCTTACA	221	174
	NRAS_Exon 2_3'	TCCGCAAATGACTTGCTATT		
	TP53_Exon 8_5'	AGGTAGGACCTGATTTCCTTACTG	558	245
	TP53_Exon 8_3'	AGGCATAACTGCACCCTTG		
	EGFR_Exon 21_5'	CAGCGGGTTACATCTTCTTTC	427	418
	EGFR_Exon 21_3'	AAACAATACAGCTAGTGGGAAGG		
	EGFR_Exon 18_5'	TTGTCCTTCCAAATGAGCTG	743	496
	EGFR_Exon 18_3'	TGCCTTTGGTCTGTGAATTG		
3	TP53_Exon 5/6_5'	ACGCATGTTTGTTTCTTTGC	507	1034
	TP53_Exon 5/6_3'	ACCCCTCCTCCCAGAGAC		
	EGFR_Exon 20_5'	AAACGTCCCTGTGCTAGGTC	442	1324
	EGFR_Exon 20_3'	CATGGCAAACTCTTGCTATCC		
	CKND2A _Exon 3_5'	TGGCTCTGACCATTCTGTTC	380	1367
	CKND2A _Exon 3_3'	TGGAAGCTCTCAGGGTACAA		

Table 6: FFPE WGA Multiplex QC-PCR Primers and amplicon description

Multiplex	Reagent	Volume per sample
1	10x FastStart PCR Buffer (with 20mM MgCl2)	1 µl
	Primer mix 1(KRAS_Ex1, PIK3CA_Ex20, EGFR_Ex19, 8 µM per primer)	1 µl
	Primer mix 2 (BRAF_Ex15, TP53_Ex7, 8 µM per primer)	1 µl
	dNTPs (from FastStart kit)	0.2 µl
	BSA (20 mg/ml)	0.2 µl
	FastStart Taq Polymerase (5 U/µl)	0.1 µl
	PCR-water	5.5 µl
2	10x FastStart PCR Buffer (with 20mM MgCl2)	1 µl
	Primer mix (EGFR_Ex18, EGFR_Ex21, TP53_Ex8, NRAS_Ex2, 8 µM per primer)	1 µl
	dNTPs (from FastStart kit)	0.2 µl
	BSA (20 mg/ml)	0.2 µl
	FastStart Taq Polymerase (5 U/µl)	0.1 µl
	PCR-water	6.5 µl
3	10x FastStart PCR Buffer (with 20mM MgCl2)	1 µl
	Primer mix (CKND2A_Ex3, EGFR_Ex20, TP53_Ex 5/6, 8 µM per primer)	1 µl
	dNTPs (from FastStart kit)	0.2 µl
	BSA (20 mg/ml)	0.2 µl
	FastStart Taq Polymerase (5 U/µl)	0.1 µl
	PCR-water	6.5 µl

Table 7: Master mix composition for FFPE WGA Multiplex QC-PCR

2.9.5. Agarose gel electrophoresis

PCR products from WGA Multiplex QC PCR and FFPE WGA Multiplex QC PCR were visualized with gel electrophoresis, on 1.5% agarose and 2% agarose respectively. The agarose gel was prepared in 1X TBE buffer with ethidium-bromide (0,5 µg/ml). PCR product (10 µl) was mixed with loading dye (3 µl), loaded into the wells of the gel and were allowed to run for 45 min at 160V. Finally, the gel was imaged on a UV transilluminator.

2.10. Lineage tree analysis

2.10.1. Reamplification of *Ampli*1™ WGA products

WGA products were reamplified using a master mix shown in Table 8. For each sample 49 µl of the master mix, was used for 1 µl of primary WGA product to reach a final volume of 50 µl. The re-amplification was performed in a PCR cycler according to the program described in Table 9. Samples undergoing lineage tree analysis were reamplified using both the variations of the programs (Reamp 65 and Reamp 68), followed by mixing the products of both the reactions in 1:1 ratio. Samples undergoing *Ampli*1™ LowPass CNA analysis were reamplified using Reamp 68 only. A WGA-QC PCR was performed to check if the reamplification reaction was successful according to the protocol described in 2.9.3.

Master Mix	Volume of constituents per sample
Expand Long Template Buffer 1 (Buffer 1)	5 µl
Lib1 (10 µm)	5 µl
dNTPs (10 mM)	1.75 µl
BSA	1.25 µl
DNA Pol mix	0.5 µl
PCR water	35.5 µl

Table 8: Master mix composition for Reamp 65/68

Step	Temperature	Duration	Cycles
Step 1	95.0 °C	1 min	
Step 2	60.0 °C	30 s	
Step 3	**65.0** °C or **68.0** °C	2 min	
Step 4	94.0 °C	30 s	
Step 5	60.0 °C	30 s	
Step 6	**65.0** °C or **68.0** °C	2 min	extend 20 s/cycle GOTO 4, 10 times
Step 7	4.0 °C	forever	

Table 9: Reamp 65/68 program

2.10.2. Double strand synthesis of *Ampli*1™ WGA products

Double strand (ds) DNA synthesis of WGA products was performed as one of the approaches to optimize processing of *Ampli*1™ WGA products for lineage tree pipeline. Currently dsDNA synthesis is skipped during processing samples for lineage tree. Components of the master mix is provided in Table 10. For dsDNA synthesis of 10 µl WGA/ reamplified WGA sample, 2 µl of the master mix was used. The samples were incubated at 68 °C for 2 h, and then cooled till 12 °C in a PCR cycler.

Master Mix	Volume of constituents per 10 µl WGA sample
Expand Long Template Buffer 2 (Buffer 2)	0.2 µl
Lib1 (100 µm)	0.2 µl
dNTPs (10 mM)	0.2 µl
DNA Pol mix	0.1 µl
PCR water	1.3 µl

Table 10: Master mix composition for dsDNA synthesis

2.10.3. Purification of reamplified WGA products

Samples meant to undergo Duplex MIPs-based targeted enrichment pipeline (at Dr Ehud Shapiro's laboratory) for lineage tree analysis were purified using AMPure XP beads (Beckman Coulter) according to manufacturer's protocol to eliminate residual PCR reagents. Briefly, 27 µl (1.8X) of AMPure XP beads were mixed with 15 µl of sample (1:1 mixture of reamp 65 and reamp 68 products or primary WGA product) and incubated for 5 min at RT to allow the DNA to bind to the beads. Next, the samples were put on a magnetic rack where the beads are separated and the supernatant containing the impurities was discarded. While on the magnetic rack, the beads were washed two times for 30s with 80% ethanol, followed by air drying of the residual ethanol for 5-10 minutes making sure that beads were not overdried (cracks should not appear on beads). Next, the samples were removed from the magnetic rack and the DNA was eluted by adding 13 µl of nuclease free water on the beads followed by vortexing and incubating for 5 min. The samples were put back on the magnetic rack and the beads were allowed to separate for 5 min or until the solution was clear. The supernatant was collected in a fresh eppdendorf Lobind tube/96-well plate and sent to our collaborators on dry ice for downstream analysis.

2.10. Lineage tree analysis

All steps involving target enrichment of STR loci by molecular inversion probes (MIPs), sequencing and data analysis were performed at the Weizmann Institute of Science, Israel according to a published pipeline (Tao et al. 2018). FastTree2 algorithm was used for final lineage tree reconstruction with the mutation count distance matrix (Price, Dehal, and Arkin 2010).

2.12. Array-CGH

WGA libraries were PCR labelled or RP labelled (random primer based labelling with exo-Klenow fragment) and hybridized on oligonucleotide-based 4x180K Agilent Microarrays (design code: 022060) as previously published (Czyz et al. 2014; Polzer et al. 2014). Microarrays were scanned using Agilent Microarray Scanner Type C and the image was processed by Agilent Genomic Feature Extraction Software 10.7 and analyzed using Agilent Genomic Workbench Software 6.5 lite. Aberration calls were defined using ADM-2 algorithm with a threshold at 7.0. For centralization a threshold of 6.0 and a bin size of 10 was used. The minimum number of probes in the aberrant interval was set to 50 and a log2 ratio of 0.25 was set to minimize false positives.

2.13. Shallow sequencing of MCF10a cells spiked in Balb/C bone marrow MNCs

Ampli1™ WGA products were subjected to dsDNA synthesis according to the protocol in chapter 2.10.2. Lib1 adaptor was released by Tru1I digestion at 65 °C for 3 h followed by an AMPureXP-purification (1.8X) and the DNA was fragmented using Covaris ultrasonication for obtaining average insert size of 350 bp. Next, Illumina TruSeq DNA PCR-free library prep was used to make libraries according to manufacturer's instructions. Quality of resulting libraries was evaluated by Bioanalyzer 2100 using the HS DNA kit, and high-quality libraries were quantified by qPCR (KAPA library quantification kit) and pooled in equimolar amounts for sequencing on the Illumina MiSeq platform with 150bp paired-end reads. The sequenced data was analyzed using BioBloom Tools (Chu et al. 2014) to evaluate the amount of mouse DNA contamination in the sample.

2.14. Ampli1 LowPass CNA Analysis

Genome wide CNA sequencing was performed using *Ampli*1 LowPass Kit according to the manufacturer's instructions (Menarini Silicon Biosystems). Briefly, *Ampli*1™ WGA products were reamplified using Reamp68 protocol (chapter 2.10.1.), 5 µl of reamplified WGA product was diluted in 5 µl NGS water and purified with 1.8X SPRIselect or AMPure XP beads (Beckman Coulter). Next, 3 µl of purified sample was used to obtain barcoded libraries, which were quantified using the Bioanalyzer 2100 HS DNA kit (Agilent) and Qubit HS DNA kit (ThermoFisher Scientific) according to manufacturer's instructions followed by pooling the libraries in equimolar amounts (4nM). Sequencing was performed in single-end mode using Illumina Miseq or Novaseq 6000 according to the manufacturer's instructions. A maximum of 32 samples were pooled in one Miseq run and 96 samples were pooled in one Novaseq 6000 run. It should be noted that a custom primer for Read1 was used for sequencing which was provided in the *Ampli*1™ LowPass kit and PhiX DNA was not spiked-in. To obtain copy-number aberration (CNA) calls a modified version of the open source software Control-FREEC (Control-Free Copy number caller), kindly provided from Menarini Silicon Biosystems, has been used.

3. Materials

3.1. List of reagents

Reagent	Manufacturer	Catalog Number
1kb Plus DNA Ladder	New England Biolabs	N3200L
AB serum, human	Bio Rad	805135
Acetone	VWR Chemicals	MFCD00008765
Adenosine triphosphate (ATP) 100 mM	Roche Diagnostics	11140965001
Agarose LE	Anprotec	AC-GN-00009
AMPure XP purification beads	Beckman Coulter	A63882
Bovine serum albumin (BSA) (20 mg/ml) (for PCR)	Roche Diagnostics	10711454001
BSA Fraction V	Sigma Aldrich	05479
CellTracks AutoPrep Instrument Buffer	Menarini Silicon Biosystems	7901003
Collagenase A	Sigma Aldrich	C0130
Collagenase 1a	Sigma Aldrich	C9891
Dispase	Life Technologies	17105-041
BCIP/NBT detection system	Bio-Rad	1706432
BSA (for cell picking)	Sigma Aldrich	B8667-5ml
dNTP Set; 100 mM each A,C,G,T; 4x 24 µM	GE Healthcare/ Roche Diagnostics	28-4065-51/ 11814362001
Ethanol absolute Mol. Bio.Grade 250 ml	VWR Chemicals	437443T
Ethanol absolut ≥99.8%	VWR Chemicals	20821.330
Ethidium Bromide Solution (10 mg/ml)	Sigma-Aldrich	E1510-10ML
Expand Long Template Buffer 1	Roche Diagnostics	11759060001
FastStart DNA Polymerase	Roche Diagnostics	4738420001
Fetal bovine serum (FBS) sera Plus	PAN Biotech	P30-3702
FFPE repair mix	NEB	M6630S
Formaldehyde, 37%	Merck	104003
Hank's balanced salt solution (HBSS) 10x	Biochrom	L2045
Hyaluronidase	Sigma-Aldrich	H4272
Igepal CA-630	Sigma-Aldrich	I3021-50ml
Levamisol	Sigma Aldrich	L9756
Mayer's hematoxylin	Sigma-Aldrich	MHS16
MseI	New England Biolabs	R0525M
NaCl 0.9%	Braun	3570160
Orange G	Sigma-Aldrich	O3756
Percoll™	GE Healthcare	17089101
Pick PBS	Gibco	10-010-023
RPMI 1640	Pan-Biotech	P05-17500
SPRI Select beads	Beckman Coulter	B23317
Tris EDTA 1x pH 8.0 low EDTA for mol. Biology	AppliChem	A8569,0500
Trypan blue	Sigma Aldrich	T8154-20ml
Trypsin/ EDTA (10x)	PAN Biotech	P10-024100
TWEEN 20	Sigma-Aldrich	P9416-50ml
Water, aqua ad iniectabilia (NGS-water)	Braun	2351744
Water for Chromatograhy 1l (PCR-water)	VWR Chemicals	1.153.331.000
Zylene	Roth	9713.3

3.2. List of Antibodies

Antibody	Manufacturer	Catalog No.
Monoclonal Mouse Anti-Human Pan cytokeratin, Clone A45B/B3	Micromet, Germany	N.A.
A45B/B3 biotinylated	in house biotinylated	N.A.
Monoclonal Mouse Anti-vimentin,clone V9	Agilent DAKO	M0725
Monoclonal Mouse Anti-human CD68, clone KP1	Agilent DAKO	GA609
Anti-CD3, T Cell antibody produced in rabbit	Sigma Aldrich	C7930
Alexa Fluor. 488 goat anti-mouse IgG1	Life Technologies	A-21121
IgG1, Kappa from murine myeloma, Clone MOPC 21	Sigma Aldrich	M9269
MOPC21 biotinylated	in house biotinylated	N.A.
Rabbit IgG isotype control	Southern Biotech	0111-01
AP-polymer anti-rabbit	Zytomed Systems	ZUC031-006
AP-polymer anti-mouse	Zytomed Systems	ZUC077-100

2.15.3. List of custom-made buffers

Buffer	Composition
AB serum 10 %	5 ml AB serum human, in 45 ml 1X (PBS) pH 7.4
Citric acid monohydrate 0.1M	2,1g citric acid monohydrate dissolved in 100 ml Milli-Q water
Collagenase solution (2%)	0.1g of Collagenase Ia, 5ml RPMI 1640
Dispase solution (2%)	0.1g of Dispase, 5ml RPMI 1640
Dissociation solution 10X	5ml 2% Dispase solution, 5ml 2% Collagenase solution
HIAR Buffer 10X	0,1M citric acid monohydrate is added drop by drop, to 100ml of 0,1M sodium citrate tribasic dihydrate, under constant swirling conditions, until the pH reaches pH 6,40.
Igepal 10 %	2 ml 100 % Igepal in 18 ml PCR water
Maltose 0.5M	1.7115 g maltose, Milli-Q water (final volume 10 ml)
One Phor All (OPA) buffer 10X	5 ml 1 M Tris acetate, 5 ml 1 M Magnesium acetate ,1 ml 5 M Potassium acetate, PCR-Water (final volume 1l)
PBATw buffer	1,0g of bovine serum albumin (BSA) fraction V in washing buffer (final volume 100 ml)
PBS pH 7.4 10X	450 g Sodium chloride, 71.65 g Disodium phosphate (Na2HPO4), 13.35 g Monopotassium phosphate (KH2PO4), Milli-Q water (final volume 5l)
Percoll 60 %	60 ml Percoll 100 % ,40 ml NaCl (0.9% Braun)
PFA 0.5%	1.35 formaldehyde solution, Milli-Q water (final volume 100 ml)
Sodium citrate tribasic dihydrate 0.1M	2.9g sodium citrate tribasic dihydrate dissolved in 100 ml Milli-Q water
Tris/Borate/EDTA (TBE) buffer 10x	539 g Tris, 275 g Boric acid, 37 g EDTA, Milli-Q water (final volume 5l)
Tween 10 %	2 ml 100 % TWEEN. 20 in 18 ml PCR water
Washing buffer	0,25ml Tween-20 dissolved in 500ml 1X PBS

3.4. List of commercial kits

Kit	Manufacturer	Catalog Number
ABC (Avidin-Biotin Complex)	Linaris	AK-5000
Agilent High Sensitivity DNA Kit	Agilent Technologies	5067-4626
Ampli1TM LowPass Kit (SET A+Set B) 2 x 48 reactions	Menarini Silicon Biosystems	WGLPAB
BCIP/NBT (AP conjugate substrate Kit)	Bio-rad	170-6432
CellSearch Circulating Tumor Cell Kit	Menarini Silicon Biosystems	7900001
CellSearch Circulating Tumor Cell Control Kit	Menarini Silicon Biosystems	7900003
DNeasy Blood & Tissue Kit	Qiagen	69504
Expand Long Template PCR System	Roche Diagnostics	11759060001 (Sigma Aldrich)
FastStart™ Taq DNA Polymerase, dNTPack	Roche Diagnostics	4738420001 (Sigma Aldrich)
KAPA Library Quantification Kit	Roche Diagnostics	07960298001
MiSeq. Reagent Kit v3 (150 cycles)	Illumina	MS-102-3001
Qubit dsDNA BR Assay Kit	Thermo Fisher Scientific	Q32853
Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific	Q32854
QIAamp DNA FFPE Tissue Kit	Qiagen	56404
TruSeq DNA PCR-Free Kit	Illumina	20015962
VECTASTAIN® ABC-AP KIT	Linaris	AK-5000

3.5. List of consumables

Product Name	Company	Catalog Number
AdhesiveCap 200 clear	Carl Zeiss	415190-9191-000
Adhesive sealing sheets	Thermo Fisher Scientific	AB0558
CellSave Preservative Tubes	Menarini Silicon Biosystems	7900005
Cellstar. serological pipette 2 ml	Greiner Bio-one	710180
Cellstar. serological pipette 5 ml	Greiner Bio-one	606180
Cellstar. serological pipette 10 ml	Greiner Bio-one	607180
Cellstar. serological pipette 25 ml	Greiner Bio-one	760180
Centrifuge tube 15 ml	Greiner Bio-One	188271
Centrifuge tube 50 ml	Greiner Bio-One	227261
Diagnostic adhesion slides	Thermo Fisher Scientific	ER-203B-CE24
Eppendorf twin.tec®-PCR-Platte 96 LoBind, skirted,	Eppendorf	0030129512
Eppendorf. protein LoBind tube 0.5ml	Eppendorf	0030 108.094
Eppendorf. protein LoBind tube 1,5ml	Eppendorf	0030108.116
Eppendorf. protein LoBind tube 2ml	Eppendorf	0030 108.132
Eppendorf. protein LoBind tube 5ml	Eppendorf	0030 108.302
LabTek Chamber Slides, glass, 8 fields	Nunc	11367764
MAXYMum Recovery™ 0.2 ml	Axygen Scientific	11370145
Micro-hematocrit capillary	Brand	749321
Microcentrifuge tube 1.5 ml	Greiner Bio-One	616201
Microcentrifuge tube 2 ml	Greiner Bio-One	623201
Multichannel Reagent reservoirs	Integra	4331
Nylon biopsy bag	Thermo scientific	6774010
Nylon mesh filters 30µm	Partec CellTrics.	04-004-2326
PCR tube 0.2 ml	4titude Deutschland	4ti-0795
PCR SingleCap 8er-SoftStrips 0.2 ml	Biozym	710970
SafeSeal Surphob filertips 1250µl	Biozym	VT0270
SafeSeal Surphob filertips 2oo µl	Biozym	VT0240
SafeSeal Surphob filertips 2o µl	Biozym	VT0220
SafeSeal Surphob filertips 1o µl	Biozym	770010
Superfrost® Plus slides	Thermo Fisher Scientific	10149870
Transparent 96-well PCR plate	Biozym	710884

3.6. List of devices

Device	Manufacturer
Bioanalyzer 2100	Agilent Technologies
Cell culture incubator	Heraeus, Hanau
Cell culture laminar flow	Heraeus, Hanau
CellTracks Autoprep® System	Menarini Silicon Biosystems
CellTracks Analyzer II® System	Menarini Silicon Biosystems
Centrifuge 5424	Eppendorf, Hamburg
Centrifuge 5810R	Eppendorf, Hamburg
Centrifuge Rotina 380R	Hettich
Centrifuge Rotofix 32A	Hettich
DMZ Universal Puller	Zeitzt
Gel electrophoresis chamber	Biostep, Jahnsdorf
Gel electrophoresis power supply	MRC Lab
Herasafe™ KS (NSF) biological safety workbench class II	Thermo Fisher Scientific
Inverted microscope	Leica
Laboratory Incubator	Mammert
Laser Microdissection microscope	Zeiss PALM Microbeam
Microscope IX81	Olympus
Magnetic stirrer	VELP Scientifica
Magnetic Rack (FastGene MagnaStand 0.2)	Nippon Genetics Europe
Micromanipulator PatchMan NP2	Eppendorf
MJ Research Peltier Thermal Cycler Tetrad	Bio-Rad
MJResearch Peltier Thermal Cycler PTC-200	Bio-Rad
MiSeq	Illumina
Multipipette Stream	Eppendorf, Hamburg
NanoDrop ND-1000	Thermo Fisher Scientific
Neubauer chamber	Schubert und Weis, Munich
Novaseq 6000	Illumina
PCR bench UVT-S-AR	Thermo Fisher Scientific
PCV-2400 Combined Centrifuge/vortex mixture	Grant-bio
pH-meter	Eutech Instruments, The Netherlands
Pipettes (single and multi-channel)	Gilson
Qubit3 fluorometer	Thermo Fisher Scientific
Screening microscope	Olympus microscope (BX43)
ThemoMixer	Eppednorf, Hamburg
UV illuminator	Intas, Gottingen
Vortex mixers	VELP Scientifica, Italy
Water bath	Memmert, Schwabach

4. Results

Longitudinal tracking of patients required development of robust methods to isolate single cells of high DNA quality from all tissues, which included (i) Laser microdissection to isolate DCCs from diagnostic cytospins; - (ii) Isolation of single cells from flash frozen/FFPE tumor tissue; (iii) Isolation of blood CTCs from CellSearch cartridges. Additionally, for every patient CD3+ T cells, CD68+ macrophages and oral epithelial cells (OECs) were isolated to serve as outgroups for cell lineage tree analysis. The development of these protocols and their successful usage on patient samples are illustrated in the sections 4.1. through 4.5. of this section. The samples were then shipped to our collaborator Dr Ehud Shapiro at the Weizmann Institute of Science, for generating metastatic phylogenies based on STR mutations which determines cellular descent by tracking random mutational events. The results of these phylogenetic analysis are elaborated in Section 4.6. The workflow of my PhD thesis is summarized in Figure 1.

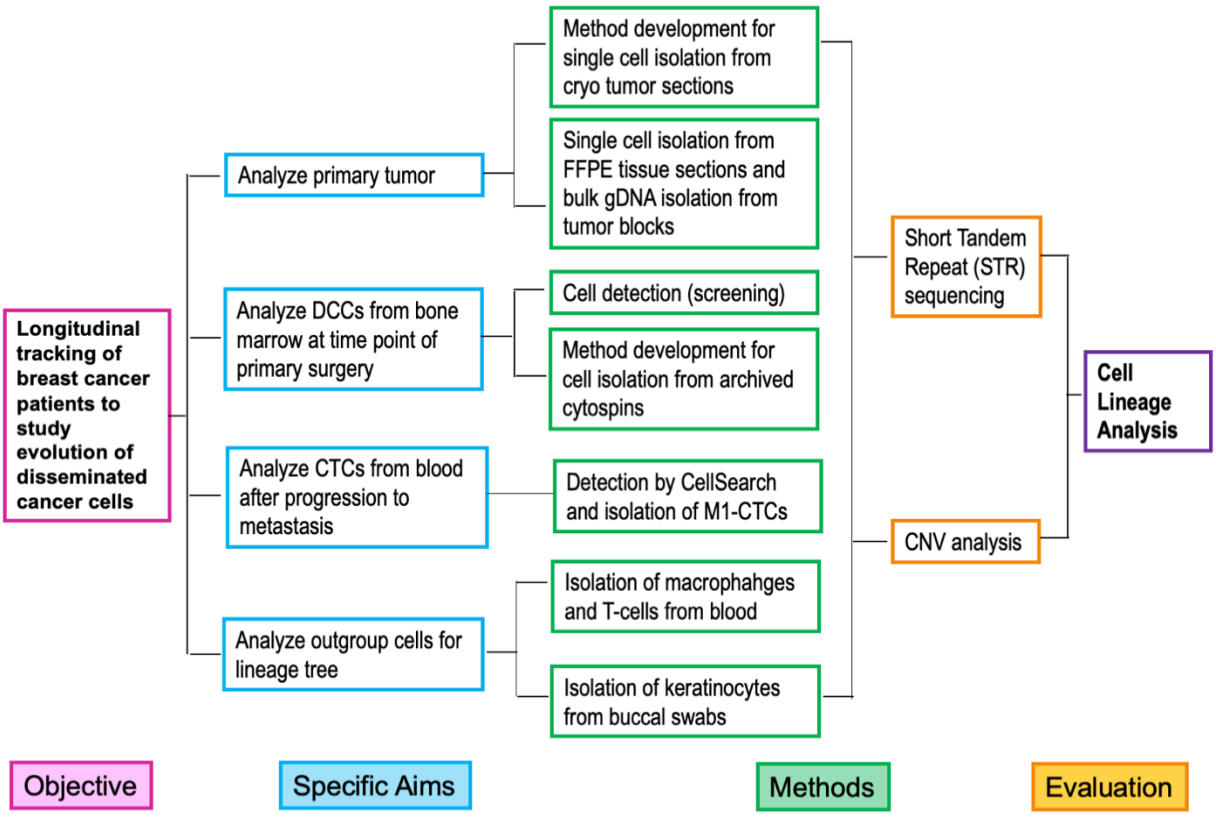


Figure 1 PhD thesis workflow. DCCs, disseminated cancer cells; CTCs, circulating tumor cells.

4.1. Isolation of disseminated cancer cells (DCCs) from archived frozen cytospins

To acquire bone-marrow (BM) DCCs we collaborated with the Department of Gynecology, University of Tübingen and the Department of Gynecology and Obstetrics, LMU Munich. These BM preparations had been collected from breast cancer patients over the last 20 years and hence a long patient follow-up data exists. However, the slides were prepared as cytospins at high cell densities. So far studies involving single cell DNA analysis of DCCs in our laboratory (Klein et al. 1999; Klein, Seidl, et al. 2002) used standard micromanipulation to isolate single cells from adhesion slides on which cells were sedimented at low densities (Figure 2A). This allowed ample space for contamination free isolation of single cells. However, isolation of single cells by standard micromanipulation from cytospins (Figure 2B) is unfeasible due to high cell density and strong adherence of cells to the slide. Consequently, we explored various methods to isolate single cells of good DNA quality from such slides.

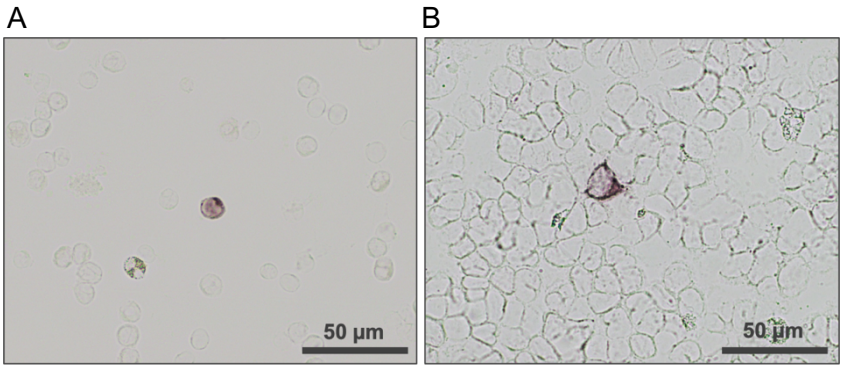


Figure 2: Cell density is crucial for micromanipulator assisted isolation. BM-DCCs can be detected with anit-cytokeratin antibody A45-B/B3 on adhesion slides and cytospins. A: BM mono-nuclear cells sedimented on adhesion slide at a density of $0.16*10^6$ per 14 mm diameter spot are amenable to standard micromanipulation. B: BM mono-nuclear cells centrifuged at a density of $1.5*10^6$ per 16 mm diameter spot on adhesion slide lay flat and adhere strongly to the glass surface making standard micromanipulation unfeasible.

4.1.1. DNA quality assessment of cells isolated from frozen archived cytsopins

At first, we wanted to find out if single cells isolated from old frozen archived cytospins have good DNA quality after whole genome amplification (WGA). Applying standard micromanipulation to cells from low-density areas (Figure 3A) revealed that the DNA quality of aged samples is well suited for WGA and downstream analyses as seen after WGA QC multiplex PCR (Figure 3B). The number of bands in the WGA QC multiplex PCR (methods section, 2.9.3) is referred to as the genome integrity index (GII) which predicts the sample's DNA quality. A GII of 0 (no bands visible in QC-PCR) means the cell was not successfully collected in the tube, while a GII>2 refers to cell with sufficient quality (Polzer et al. 2014). From 123 such patients' BM samples, which had been prepared up to 18 years ago, we found that median 60% of single cells isolated per sample showed a GII>2. Detailed WGA QC multiplex PCR results of all the samples have been summarized in Appendix 1. Dividing the samples by age of preparation in brackets of 5 years and applying non-parametric Kruskal-Wallis Test shows that there is no significant difference among the samples prepared up to 10 years ago (Figure 3C), nevertheless the cell quality slightly drops beyond 10 years of slide preparation. Copy number alteration (CNA) analysis by *Ampli1*™ LowPass (methods section 2.14) of marker negative cells from archived patient samples (8 years and 11 years ago) also

showed comparable balanced profiles to cells picked from fresh adhesion slides, prepared less than a year ago (Figure 3D). Thus, frozen archived cytospins can be used for isolating DCCs of high quality.

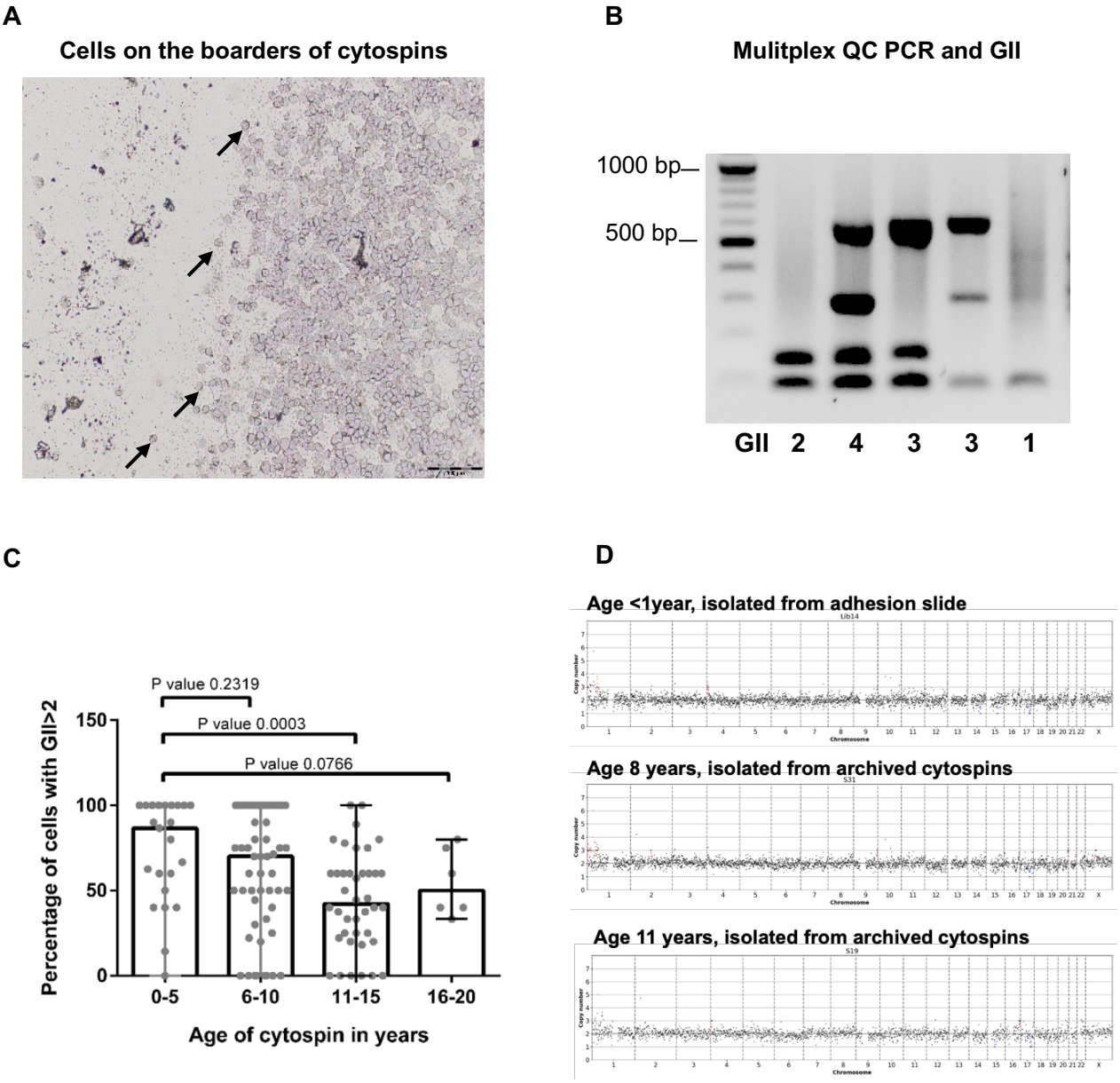
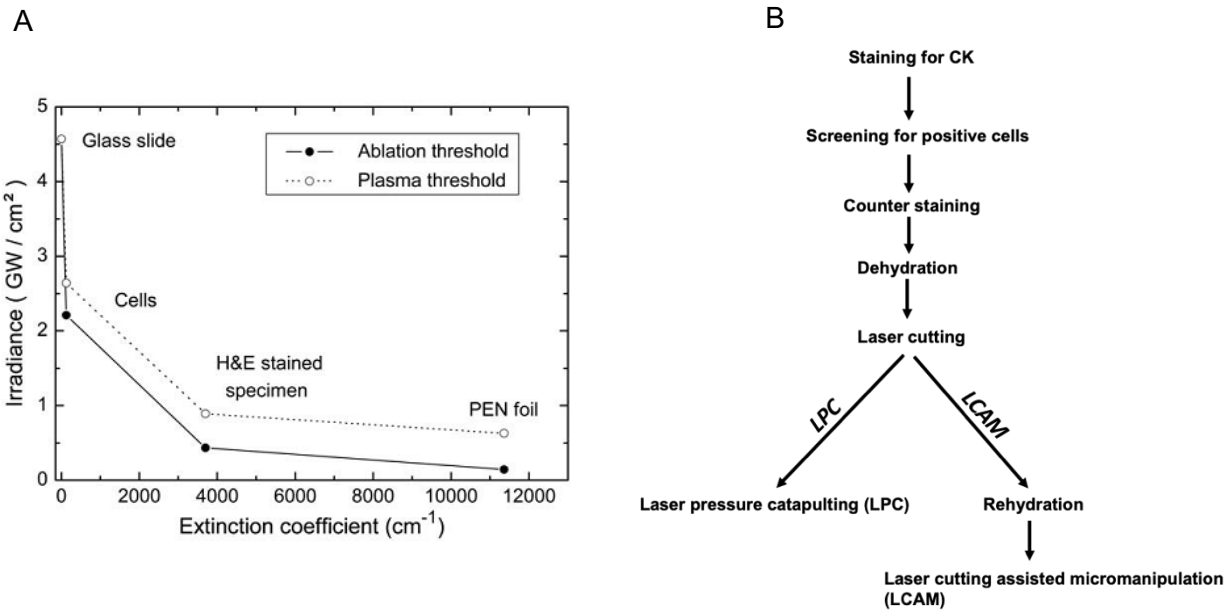


Figure 3: Single cell DNA quality of archived frozen cytospins. A: Example of cells at the boarders of high density cytospins. Marked cells could be easily isolated by standard micromanipulation for testing the single cell quality in such cytospins. B: Representative image of a multiplex QC PCR gel. Genome integrity index (GII) is the number of bands (0-4). C: Cytospin samples were divided according to age of preparation in groups of 5 years and the WGA results of cells isolated by standard micromanipulation are plotted on the ordinate. Kruskal Wallis Test suggests a significant difference (p value=0.0027) between the samples. We obtained a significant difference between the 0-5 years and the 11-15 years group (P value=0.0003) D: *Ampli1*™ Lowpass whole genome CNA sequencing of marker negative cells with GII>2 from <1year old adhesion slides compared to cells from 8 and 11 year old slides.

4.1.2. Laser Microdissection as a solution for DCC isolation

Laser microdissection (LMD) has emerged as a promising technique through the last decades for contamination free isolation of single cells (Ahmed 2006), but in order to obtain good quality nucleic acid post LMD the samples should be prepared on UV absorbing polyethylene nephthalate (PEN) membrane slides (Ahmed 2006; Vogel et al. 2007). The high optical absorption coefficient of the membrane makes LMD possible at lower laser power (Figure 4A), enables sharp laser focusing, acts as a scaffold and protects the specimen from UV laser energy during laser pressure catapulting (LPC) (Vogel et al. 2007). Since, our archived samples were not prepared on membrane slides, we tested various methods to increase laser precision and protect cellular DNA from laser-induced damage. Since unstained cells have the lowest optical absorption coefficient (Figure 4A), we tried to increase the absorption coefficient in the sample by using histological dyes as counterstains. We also tested DNA binding fluorochromes and other UV absorbing molecules in our attempts to compensate for the absence of PEN membrane (Table 3).

Next we optimized the sample dehydration procedure to prepare samples for LMD. For the acquisition of cells, we tested two protocols, 1. Laser cutting followed by laser pressure catapulting (LPC) which allowed rapid contact-free isolation (Figure 4B and C) and 2. Laser cutting assisted micromanipulation which we abbreviated as LCAM (Figure 4B and D).



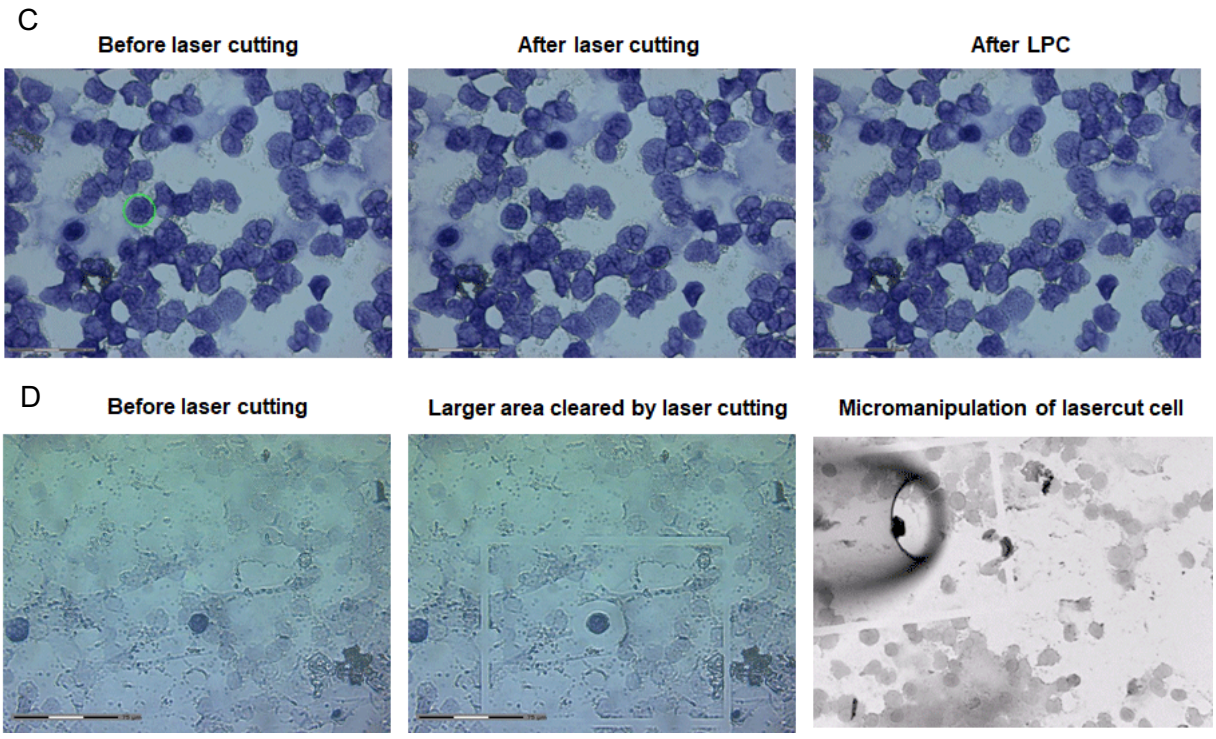


Figure 4: Possibilities of cell isolation by laser microdissection. A: Thresholds values of irradiance (laser power per unit area) for ablation and plasma formation at 337 nm for glass, unstained cells, H and E stained sections and PEN foil; lower optical absorption coefficient (extinction coefficient) relates to higher irradiance thresholds; laser ablation is required for laser cutting and plasma formation is required for LPC; PEN foil offers the lowest threshold for laser ablation followed by stained sample (figure adapted from Vogel et al., 2007). B: The steps and possibilities leading to successful laser microdissection. C: Protocol 1, Laser Pressure Catapulting (LPC) where at first laser cutting is used to isolate the cell from the surroundings followed by acquisition of the cell by laser force from the bottom of the specimen, the cell is collected in a cap of a tube kept above the specimen. D: Protocol 2, Laser cutting assisted micromanipulation (LCAM), where laser cutting is used to clear a larger area around the cell of interest which would make room for standard micromanipulation.

4.1.2.1 Sample preparation

Counterstaining: After screening for positive cells, the specimen had to be counterstained with an intense dye in order to obtain sharp laser focus during laser cutting at minimum laser power. Additionally, the stain should not chemically hinder any step of WGA. As shown in Figure 5, common histological stains were tested by Eva Maria Hecht as part of her master thesis (Hecht 2016). Mayer's hematoxylin performed best in terms of both staining intensity and *Ampli1*™ WGA compatibility (Table 1) and was our choice of counterstain for future experiments (methods section 2.2.2.1).

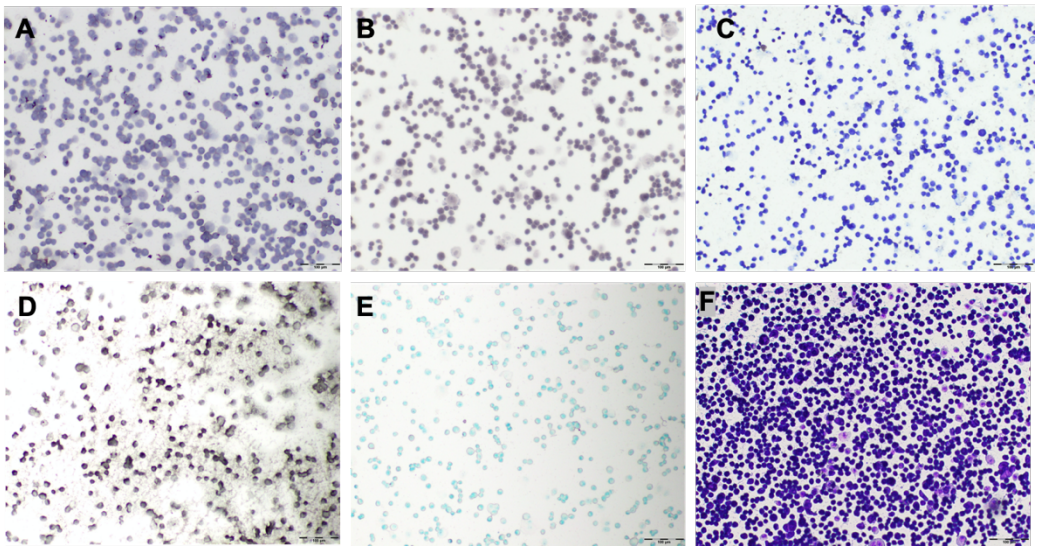


Figure 5: Various hiostological stains tested for LMD. Cells stained with, A: Mayer's Hematoxylin. B: Weigert's Hematoxylin. C: Crystal Violet. D: Chlorazol Black E. E: Methyl Green. F: Giemsa. All images are acquired at 10X magnification. Figure adapted from (Hecht 2016)

Histological Dyes	LPC#	% cells with GII>2, picked manually *
Mayer's hematoxylin	+++	100
Weigert's hematoxylin	+++	0
Crystal Violet	+++	50
Chlorazol Black E	+++	78
Methyl Green	+	80
Giemsa	+++	0

Table 1: Summary of results from Figure 8
#LPC isolation was assessed by the ability to catapult a cell into the cap of a tube with lowest possible laser power, +++ (varied from 55-65%), + (values higher than 70% / inability to LPC intact cells), - (LPC was not possible at all)
* depicts percentage of cells displaying GII>2 from cells isolated by standard micromanipulation; shows that the specific dye or the stain does not interfere with the WGA protocol.

Dehydration: Samples had to be dehydrated for LMD. We used an increasing ethanol series followed by a short dip in acetone to dehydrate the samples, at RT and -20 °C (methods section 2.2.2.1). Patient cytospin samples which were prepared through the last decade were used for the experiment. Figure 6A displays the results of WGA QC PCR of cells picked from slides of various time points treated with the following protocols: 1) fresh (picked right after hematoxylin staining without dehydration); 2) after dehydration at RT and 3) after dehydration at -20 °C. Since the same slide cannot be dehydrated twice, a separate slide from the same BM sample was used for probing the effect of dehydration at -20 °C. The quantification of WGA QC PCR is shown in Figure 6B. We saw that dehydration at RT is detrimental for WGA of single cells for majority of samples, possibly because DNA when exposed to ethanol at RT shows decreased sensitivity to restriction digestion (Svaren et al. 1987) and since *Ampli*1™ WGA is based on *Mse1* restriction digestion, dehydration temperature could be critical at a single cell level. However, dehydration at -20 °C provides results not different to that of fresh samples. Hence, dehydration at -20 °C was used for sample preparation for LMD.

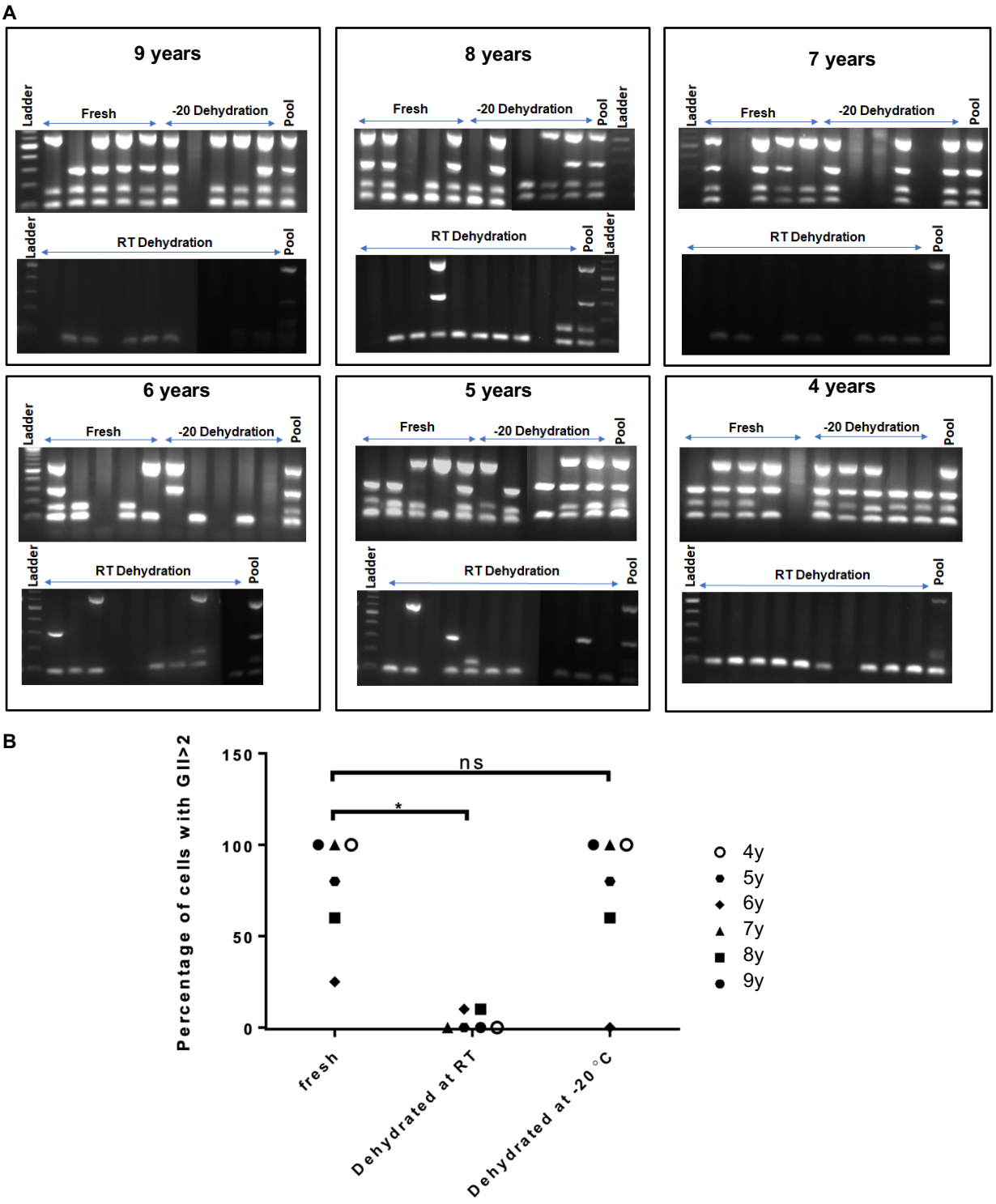


Figure 6 Dehydration temperature is crucial for obtaining good quality DNA from archived frozen cytospins. A: Results of WGA QC PCR, obtained from cells of 6 different BM cytospin samples prepared 9 - 4 years ago. Five cells were isolated per condition. Fresh, Cells picked right after hematoxylin staining; -20 Dehydration, Cells picked after dehydrating the slide at -20 °C; RT Dehydration, Cells picked after dehydrating at RT; Left to right from 9 – 4 years old samples. B: Quantification of A showing the percentage of cells with GII>2 per sample, per treatment, Wilcoxon signed rank test is used to compare the fresh and dehydrated at RT sample pairs (p value 0.03) and fresh and dehydrated at -20 °C sample pairs (p value 0.99, n.s.).

4.1.2.2. Laser pressure catapulting of single cells (LPC)

LPC allows automated and contactless isolation of single cells wherein the cells are catapulted from the glass surface by laser force into the cap of a 200 μl microfuge tube (Figure 7A). We first tried if cells could be removed from glass slides without the presence of a PEN membrane. As, shown in Figure 4A stained cells have a higher UV extinction coefficient, and hence the plasma formation threshold power, which is required for catapulting a cell, is lower. We observed that darker stained cells could be successfully catapulted (Figure 4C). Counterstaining as mentioned in section 4.1.2.1 and seen in Figure 7B, increases the staining intensity of a DCC along with staining the surrounding cells which aids in laser cutting. Therefore, we next tried laser cutting and LPC to isolate DCCs from patient cytospins.

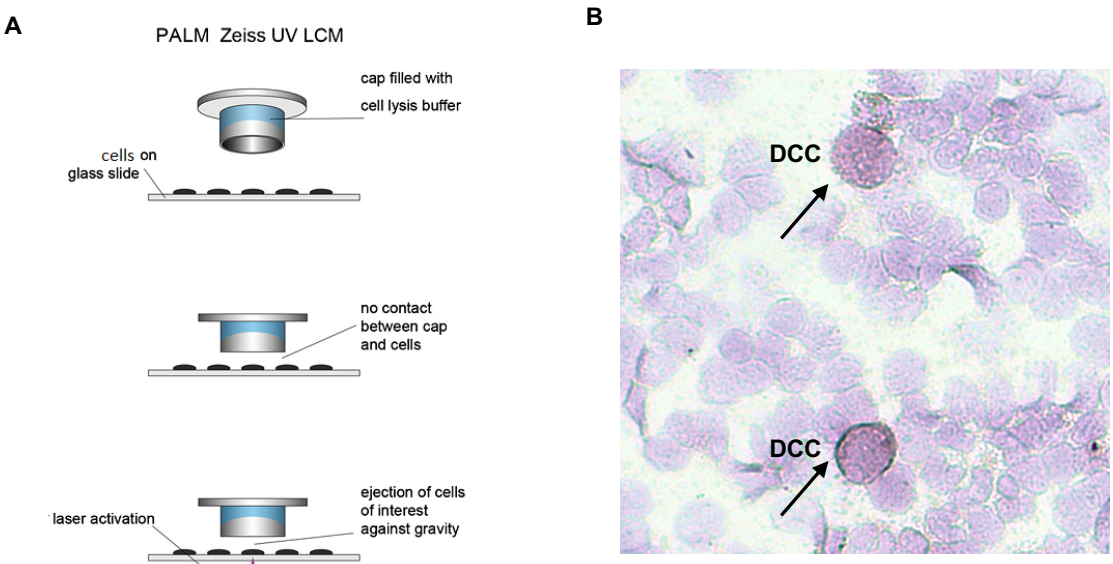


Figure 7: LPC from glass slides
A: The process of LPC (PALM Zeiss), figure adapted from (Vandewoestyne et al. 2013) . Cells are catapulted against gravity onto a cap placed above the specimen B: Cytokeratin positive DCCs developed with BCIP/NBT as substrate and counterstained with Mayer's Hematoxylin, hence the cells are darker than the background (Image captured at 40X magnification).

Initially in order to develop the protocol, we used patient BM cytospins which were irrelevant to the project and contained more than 10 DCCs per slide. We detected the DCCs using a pan cytokeratin antibody A45-B/B3 and developed using BCIP/NBT (described in methods section 2.3.2) The slides were counterstained with hematoxylin (Figure 7B). We performed laser cutting followed by LPC to isolate DCCs from two patient derived cytospins. Surprisingly both the processes worked well despite that there was an absence of PEN membrane on patient slides. Isolated cells were subjected to WGA and thereafter a WGA QC PCR. As controls we isolated some cells from low-density areas of the same cytospins by standard micromanipulation. We found that the quality of cell isolated by LPC was low, as only 35% and 25% of cells from patient T1 and T2 respectively had a GII>2 as opposed to the controls where

the majority (100% and 88% for T1 and T2 respectively) of cells showed good DNA quality. (Figure 8 and Table2)

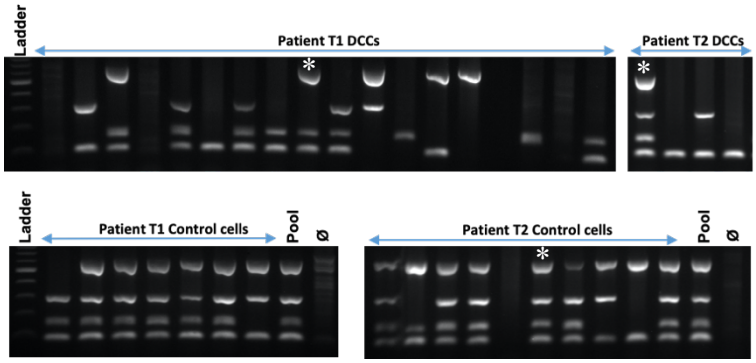


Figure 8: LPC from patient cytospins

Patient T1: Top. WGA QC PCR results of DCCs collected by LPC. Bottom. WGA QC PCR results of control cells collected by standard micromanipulation.
Patient T2: Top. WGA QC PCR results of DCCs collected by LPC. Bottom. WGA QC PCR results of control cells collected by standard micromanipulation
* Cells analyzed by array CGH

	Cells successfully isolated by LPC	Cells with GII>2 (percent)	Cells successfully isolated by micromanipulation	Cells with GII>2 (percent)
T1	14	5 (35%)	7	7 (100%)
T2	4	1 (25%)	9	8 (88%)

Table 2 Summary of results from Figure 8

Next, we analyzed the cells with GII>2 from patient T1 and T2 (Figure 8) by array CGH (comparative genome hybridization) as described in method's section 2.12. We found that the DCCs isolated by LPC had higher DLRS values (which is a measure of hybridization noise) and displayed artefacts in the genomes as compared to the genomes of DCCs isolated by standard micromanipulation (Figure 9). These results lead us to speculate that LPC might have an impact on the DNA quality of the cell, as seen by the artefacts introduced at the array CGH level. Next, we wanted to test if this observation also applies to healthy control cells.

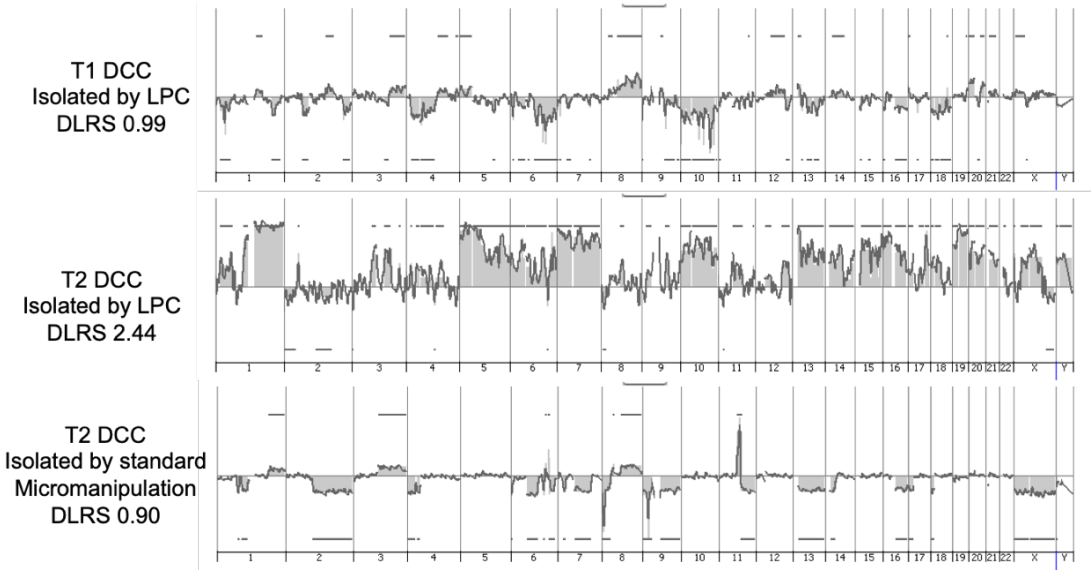


Figure 9: Array CGH of patient DCCs isolated by LPC and by standard micromanipulation. Top to bottom, Patient T1 DCC isolated by LPC displays noisy profile (DLRS spread 0.99); Patient T2 DCC isolated by LPC shows huge aberrations and has a high DLRS spread of 2.44; Patient T2 DCC isolated from the same slide by standard micromanipulation has lower hybridization noise (DLRS spread 0.90).

Since tumor cells intrinsically have aberrant profiles, we wanted to exclude this bias while further testing and developing our protocol. Therefore, we focused on non-tumor cells and used both, patient BM cytospins and healthy donor PBL adhesion slides, for isolating healthy mono-nuclear cells (MNCs). For simulating the scenario of isolating a CK positive DCC we stained patient derived bone marrow cytospins and healthy donor blood PBL adhesion slides against mesenchymal marker vimentin (methods section 2.4), which should be expressed by the mesenchymal cells of BM/blood and should have a normal karyotype. BCIP/NBT was used as a development substrate to mimic the staining of the DCCs and the background was counterstained with Mayer's hematoxylin. After dehydration at -20 °C, test cells and controls were isolated by LPC and standard micromanipulation, respectively, and were subjected to WGA and thereafter WGA QC PCR (Figure 10). We again observed low single cell DNA quality, as only 40% of patient BM MNCs and 25% of healthy donor PBLs isolated by LPC showed a GII>2 compared to 100% of the control cells (Table 3).

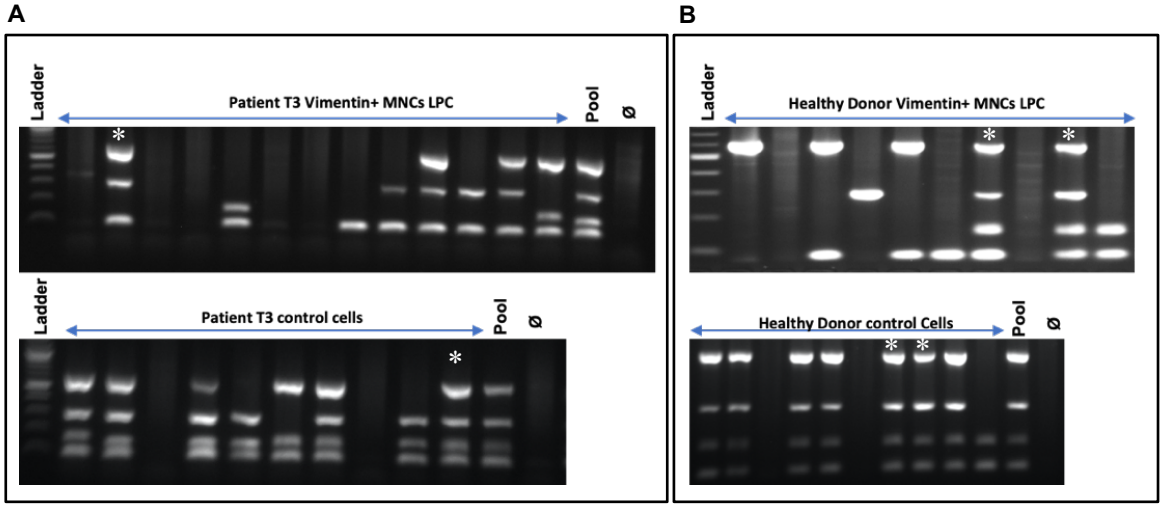


Figure 10: Vimentin positive cells isolated by LPC and standard micromanipulation. A: WGA QC PCR results of vimentin positive MNCs isolated by LPC from patient T3 cytospin, control cells isolated by standard micromanipulation are shown below. B: WGA QC PCR results of vimentin positive MNCs isolated by LPC from healthy donor adhesion slide, control cells isolated by standard micromanipulation are shown below.
* Cells analyzed by array CGH

Patient	Cells successfully isolated by LPC	Cells with GII>2 (percent)	Cells successfully isolated by micromanipulation	% Cells with GII>2 (percent)
T3	8	4 (40%)	8	8 (100%)
*H.D.	8	2 (25%)	8	8 (100%)

Table 3: Summary of results from Figure 10.
*H.D. stands for Healthy Donor

Next, we performed array CGH of high-quality cells (GII>2) isolated by LPC and standard micromanipulation of both patient and healthy donor MNCs. Again, the cells isolated by LPC displayed noisy array CGH profiles and artefacts in otherwise normal genomes when compared to the controls (Figure 11).

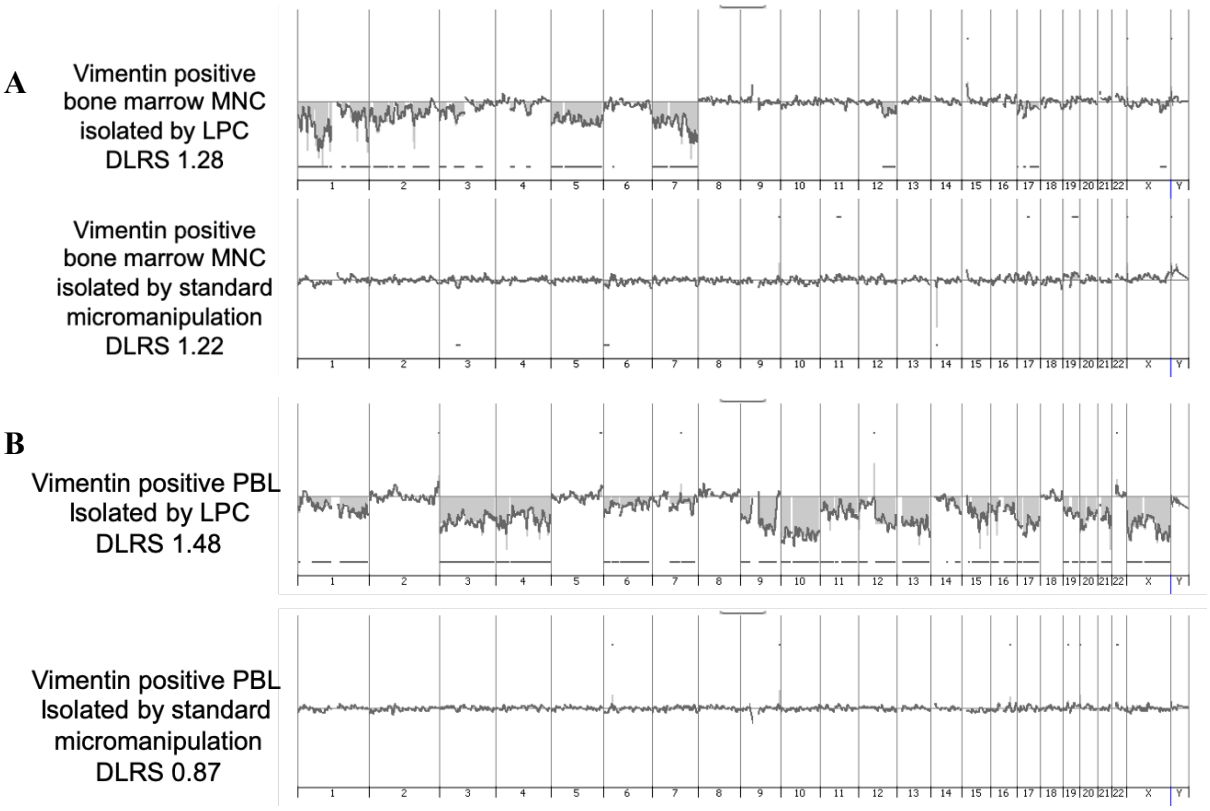


Figure 11: Array CGH of vimentin positive cells isolated by LPC and standard micromanipulation. A: Top, patient BM MNC isolated by LPC displays huge aberrations and has a higher DLRS value despite having a GII of 4 as compared to bottom profile, an MNC isolated by standard micromanipulation. B: Top, healthy donor MNC also shows aberrations and higher DLRS values despite having a GII of 4 as compared to; bottom, an MNC isolated by standard micromanipulation.

Efforts to improve LPC

Since we observed that cells with GII>2 showed artefacts in the DNA at array CGH resolution, we speculated that laser energy could be destroying the DNA. Therefore, we wanted to reduce the laser power required to isolate single cells by LPC, hypothesizing that the lower the power needed to catapult the cell, the better would be the DNA quality. A high optical absorption of the sample would mean usage of lower LPC power to isolate a cell. In order to achieve that we tested various UV absorbing molecules which could increase the optical absorption coefficient of the sample and thereby reduce the threshold energy needed to catapult the cells.

The tests were performed on cytospins prepared with MCF10a cell line cells and the results are summarized in the Table 3. MCF10a cells were used because of their mammary epithelial origin and cytokeratin positivity. The experiments were performed by Eva Maria Hecht, as part of her Masters project (Hecht, E.M. 2016). Briefly, after treating the slides with the known UV absorbing chemicals namely a 1:1 mixture of SYBR Gold:DAPI and DNA intercalating dye 9-amino acridine the percentage of LPC isolated cells with good DNA quality increased as we were able to isolate up to 80% of cells with a GII>2 (Table 4). But unfortunately, the array CGH results revealed that the DNA of treated single MCF10a cells isolated by LPC again had artefacts (Figure 12).

We could finally conclude that DNA quality of cells isolated by LPC directly from glass slides in the absence of PEN membrane is sub-optimal and LPC could not be used for isolation of DCCs from frozen archived cytospins.

Protocol	**LPC#**	**% cells with GII>2 of manually micromanipulated cells***	**% cells with GII>2 of LPC cells****	**ArrayCGH *****
Mayer's hematoxylin (In combination with BCIP/NBT staining)	+++	100	30-60	failed
UV absorbing compounds				
Avobenzone	-	n.d.	n.d.	n.d.
Gumrosine	-	n.d.	n.d.	n.d.
DNA binding fluorochromes with UV-A excitation				
DAPI (In combination with BCIP/NBT staining)	+++	100	60	n.d.
SybrGold (In combination with BCIP/NBT staining)	+++	100	40	n.d.
DAPI+SybrGold (In combination with BCIP/NBT staining)	+++	100	80	failed
9AA	+++	100	50	failed

Table 4: Protocols tested to isolate single DTCs by LPC
#LPC isolation was assessed by the ability to catapult a cell into the cap of a tube with lowest possible laser power, +++ (varied from 55-65%), + (values higher than 70% / inability to LPC intact cells), - LPC was not possible at all
*depicts percentage of cells displaying GII>2 from cells isolated by standard micromanipulation. These controls show i) That the specific dye or the stain does not interfere with the WGA protocol; ii) That the slide that was used for testing have good quality cells to begin with. This step was only performed when LPC was successful.
**depicts % of cells displaying GII>2 from cells isolated by LPC.
***final experimental readout, with normal cells displaying normal genomes as decisive criterion.

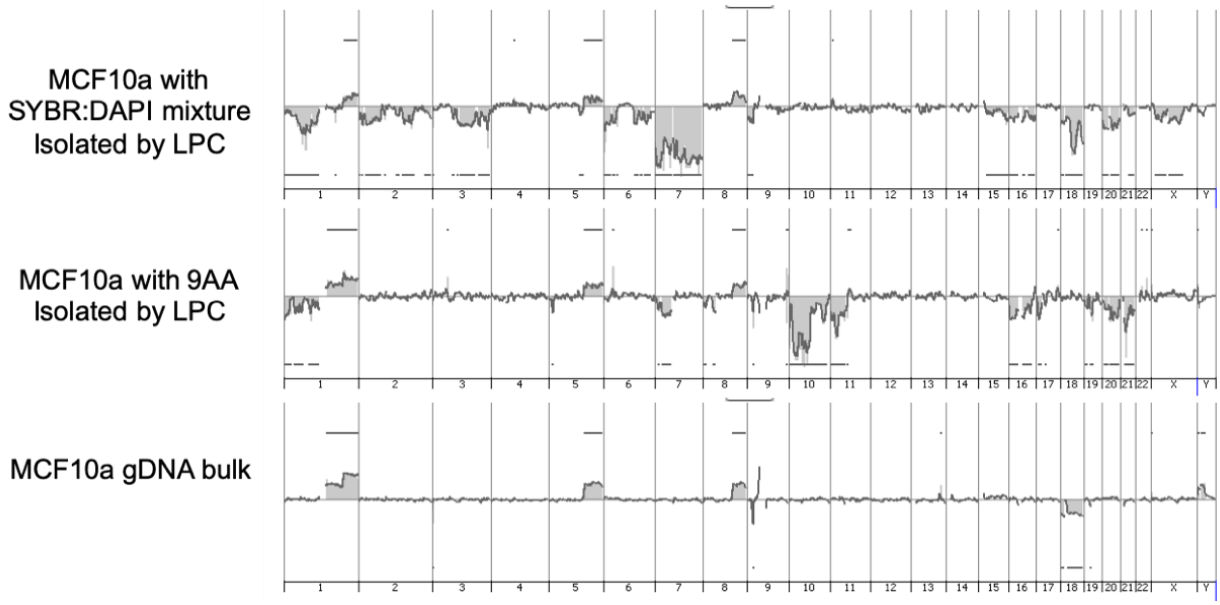


Figure 12 Array CGH of MCF10a cells isolated by improved LPC. Top to bottom. ArrayCGH profile of MCF10a cell stained against cytokeratin and developed with BCIP/NBT and a 1:1 mixture of fluorochrome DAPI and SybrGold isolated by LPC; ArrayCGH profile of MCF10a cell stained against cytokeratin with BCIP/NBT and DNA intercalating acridine dye 9AA isolated by LPC; ArrayCGH of MCF10a bulk genomic DNA shown as control.

4.1.2.3. Laser cutting assisted micromanipulation (LCAM)

Since we failed to obtain cells with intact DNA after isolation by LPC, we tested the next approach - laser cutting assisted standard micromanipulation (LCAM) which is a combination of laser cutting and standard micromanipulation (Figure 4B and D). At first laser cutting is used to isolate the cell of interest from the surrounding cells by destroying a larger area around the cell with laser in order to make room for standard micromanipulation of the cell of interest by a glass capillary. Before standard micromanipulation the slide is rehydrated in a buffer containing detergents to wash any stray cellular elements resulting from laser ablation (described in methods section 2.2.2.2.). WGA quality of MNCs and DCCs isolated from patient BM cytospins having gone through the entire workflow of LCAM is shown in Figure 13. Some cells are isolated by standard micromanipulation only from the low-density regions of the slides, to serve as controls. Up to 85% of cells isolated by LCAM have a GII>2 as compared to 100% of control cells (Table 5).

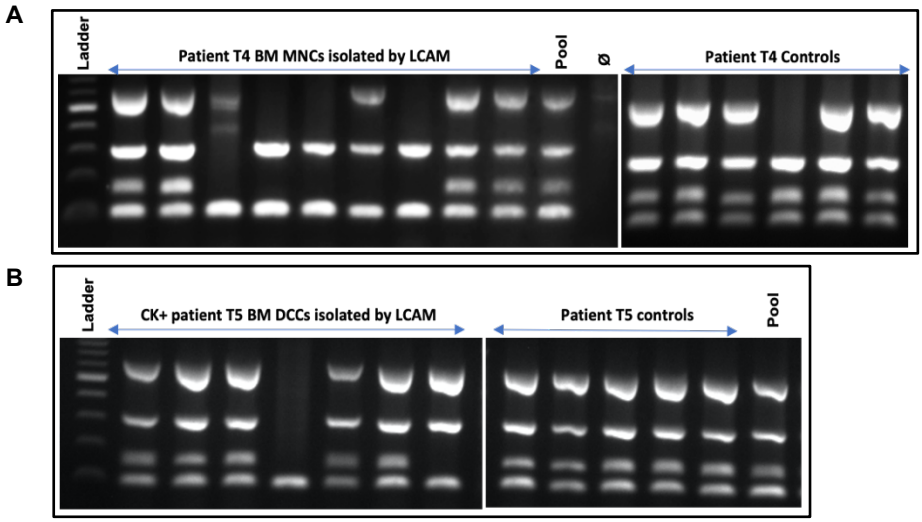


Figure 13: **LCAM from patient cytospins.** A: Left, WGA QC PCR results of marker negative cells collected by LCAM from patient T4. Right- WGA QC PCR results of controls collected by standard micromanipulation from the same slide. B: Left, WGA QC PCR results of DCCs collected by LCAM from Patient T5. Right, WGA QC PCR results of control cells collected by standard micromanipulation.
*Cells were analyzed by array CGH

Patient	Cells successfully isolated by LCAM	Cells with GII>2 (percent)	Cells successfully isolated by micromanipulation	% Cells with GII>2 (percent)
T4	9	5 (55%)	6	6 (100%)
T5	7	6 (85%)	5	5 (100%)

Table 5 Summary of results from Figure 13

Next, we wanted to check the DNA profiles of these cells by array CGH. We saw that marker negative cell isolated from patient T4 cytospin by LCAM displayed balanced array CGH profiles (Figure 14A) similar to that of a control cell picked by standard micromanipulation. We then also tested a DCC from patient T5 for which we could use a DCC cluster as a positive control (Figure 14B) and saw that the samples shared majority of aberrations. As the DNA quality of cells obtained by LCAM was optimal for array CGH analysis, we could conclude that LCAM could be used for isolation of DCCs form frozen archived cytospins.

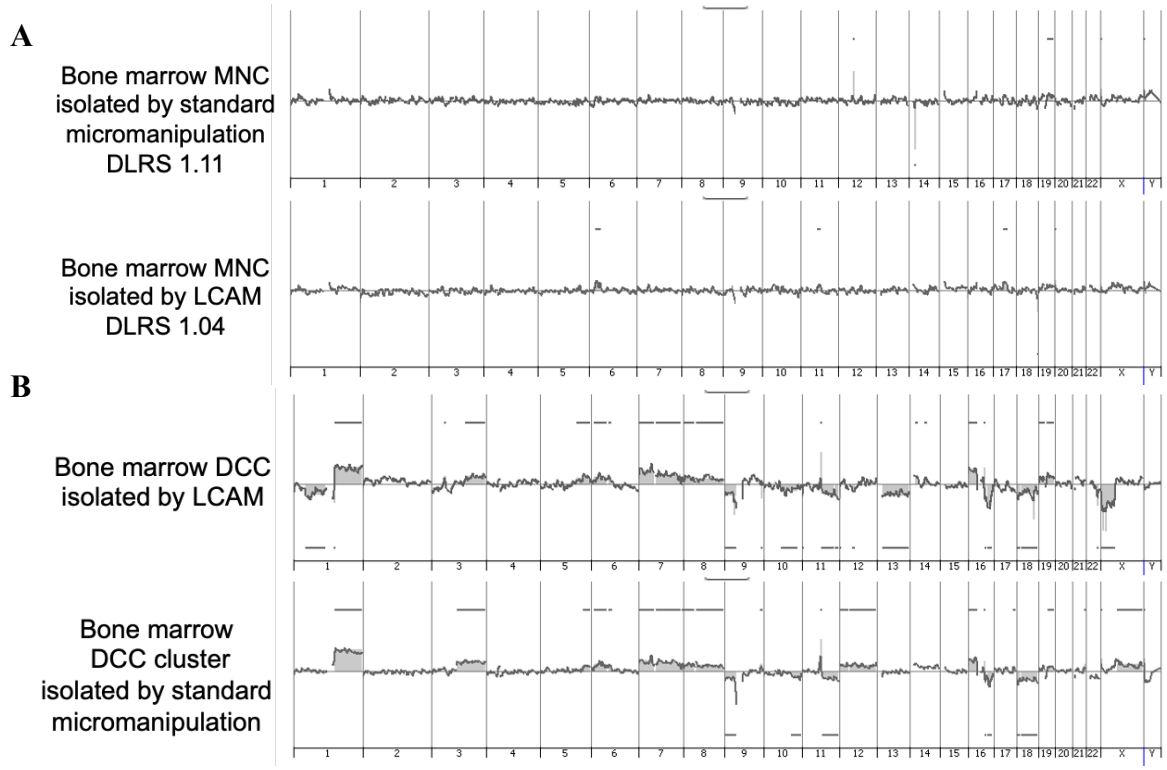


Figure 14: ArrayCGH of cells isolated by LCAM. A. Top: ArrayCGH profile of bone marrow MNC isolated from archived patient cytospins by standard micromanipulation; Bottom: arrayCGH profile of bone marrow MNC isolated from archived patient cytospins by LCAM. B: Top: ArrayCGH profile of CK positive DCC isolated from archived patient cytospins by LCAM; Bottom: ArrayCGH profile of DCC cluster isolated by standard micromanipulation from the same slide as control.

Lastly, we wanted to evaluate the amount of stray DNA left in the surroundings after laser cutting. Presence of DNA contamination could introduce artefacts in downstream analysis. To test that LCAM performs as good as standard micromanipulation, we prepared cytospins of mouse BM MNCs with spiked in MCF10a cells (described in methods section 2.5) and stained for human cytokeratin (described in methods section 2.5.1). MCF10a cells were isolated from these cytospins by LCAM (Figure 15) and also by standard micromanipulation from regions of low density as controls. We also isolated cells from slides containing only MCF10a cells to serve as a pure reference.

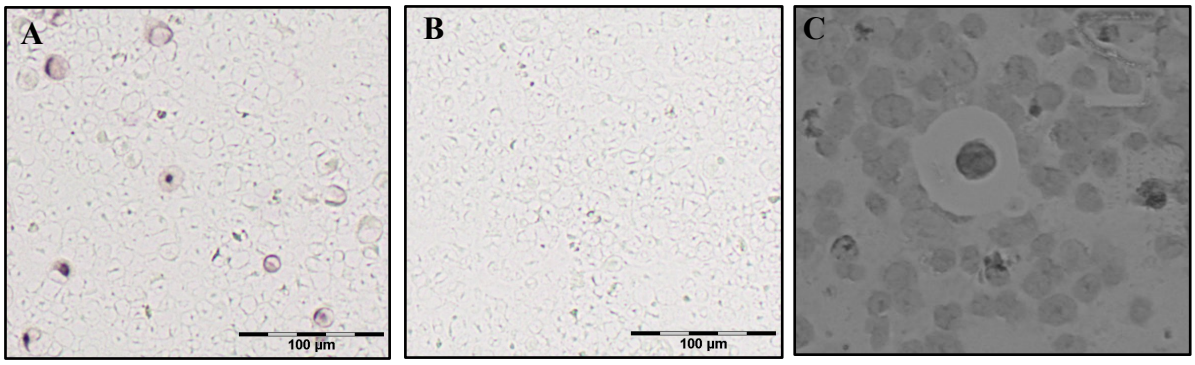


Figure 15: MCF10a cells spiked in BALB/C mouse bone marrow cells. A: MCF10 cells detected after CK staining (with biotinylated A45/Bb3 antibody) in BALB/C mouse bone marrow cells. B: Isotype Control. C: MCF10 cell prepared for LCAM (Image captured at 20X magnification.

In order to investigate DNA contamination shallow whole genome sequencing of the WGA products (Figure 16A) was performed at ITEM-R by Dr Stefan Kirsch's group and analyzed by BioBloom Tools (methods section 2.13). There was no difference in contamination levels for all the samples as evaluated by Mann Whitney test (Figure 16B). We also checked the CNA profiles of cells isolated by LCAM to look for the presence of artefacts by *Ampli*1^{TM} Lowpass sequencing. However, the profiles appeared identical to the cells isolated by standard micromanipulation (Figure 16 C and D). Thus, we could conclude that there is no stray DNA contamination introduced particularly by the LCAM protocol.

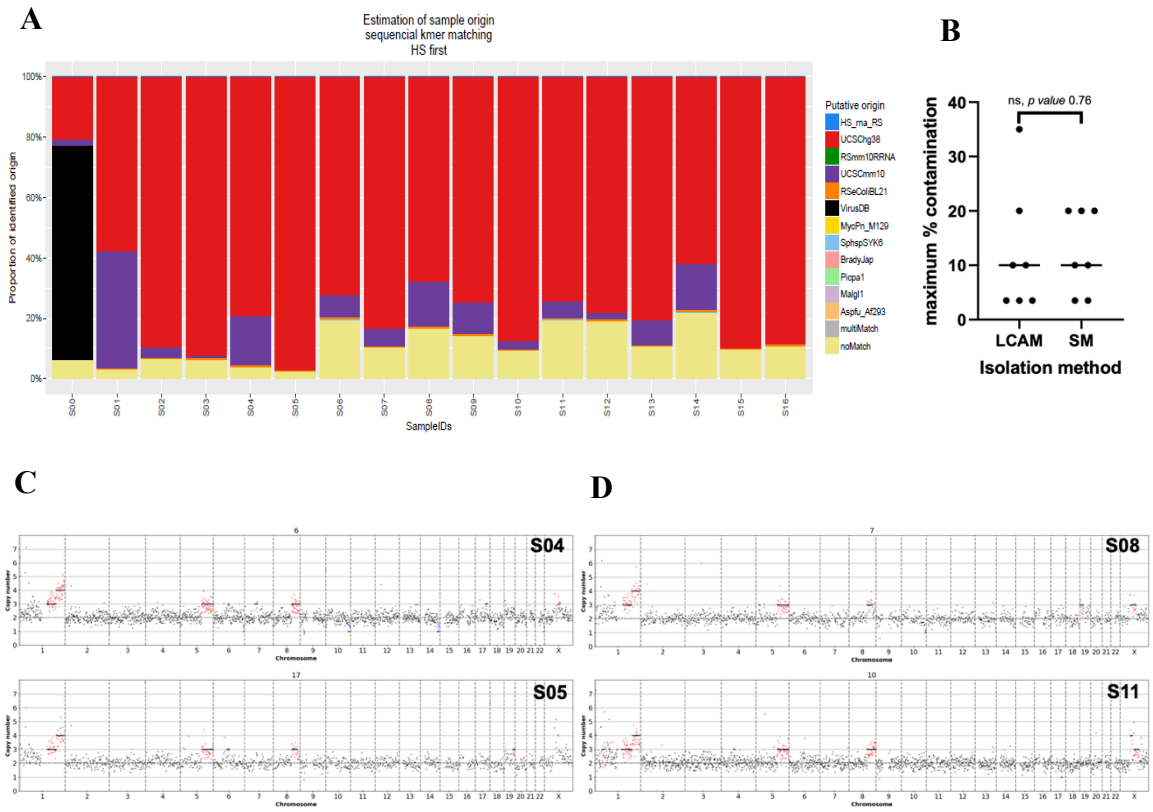


Figure 16: Assessing possible contamination due to LCAM. A: Sequential filtering was performed for the contamination estimation as shown in the legend order. Red, human; violet, mouse; the rest of the legend included common contaminants of molecular biology reagents. Ordinate shows the proportion of identified sequences from different organisms in the sample. Samples S01-S07 and S08-S14 are isolated by LCAM and standard micromanipulation respectively. Samples S15-S16 were isolated by standard micromanipulation from pure MCF10a slides. Samples S15/S16 show no mouse contamination (as expected for pure MCF10a), sample S02/S03/S05/S10/S12 show little contamination (<3.5%), samples S06/S07/S11/S13 show moderate contamination (<10%), samples S04/S08/S09/S14 show strong contamination (<20%), sample S01 shows very strong contamination (>35%), S00 is the "negative" control, i.e. the sequence reads from the spiked in PhiX. B: Median contamination in both LCAM and control group is <10%, Mann Whitney test for the contamination levels in both groups is non-significant, p value 0.76. C: Lowpass whole genome CNA sequencing of sample S04 and S05 isolated by LCAM. D: Lowpass CNA sequencing of sample S08 and S11 isolated by standard micromanipulation.

4.1.3. Summary of DCCs isolated from cytospins

After thorough investigation as seen in Section 4.1.2. LCAM was used as the method for isolation of DCCs from high density BM frozen archived cytospins. A laser microdissected DCC from patient 01 can be seen in Figure 17. In total 182 DCCs were isolated from 23 patients, out of which 143 (78%) had a GII≥2. Table 6 provides a summary of all DCCs isolated by LCAM from the cytospins acquired from the University Hospital Tübingen.

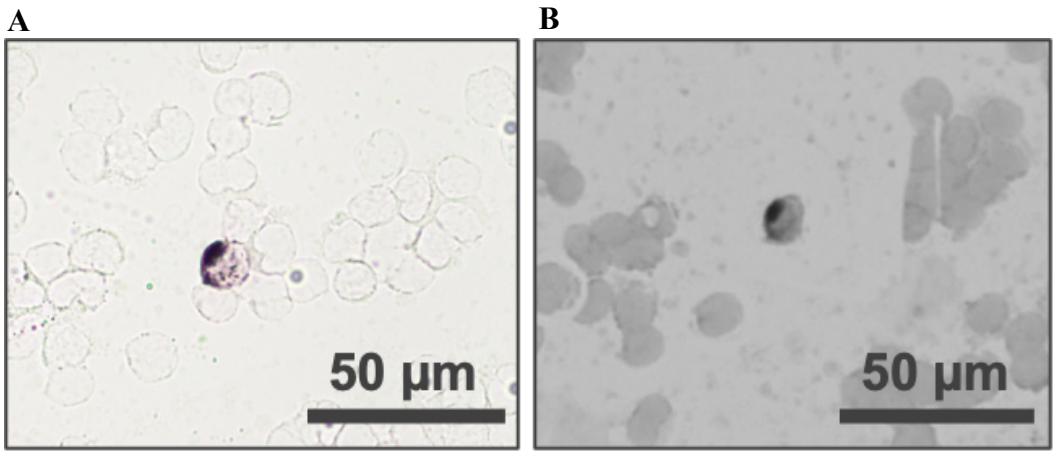


Figure 17: Patient DCC isolated by LCAM. A: DCC detected by A45/Bb3 PAN-cytokeratin staining; B: DCC after laser cutting, ready to be picked by standard micromanipulation.

Patient ID	DCCs successfully isolated by LPC	Cells with GII≥2	*NZs successfully isolated by micromanipulation	Cells with GII≥2
03	12	9	5	3
05	3	2	10	10
04	3	2	8	7
09	10	7	8	8
11	11	11	7	5
14	8	7	6	5
02	2	2	9	8
08	2	2	10	7
01	2	2	10	10
12	2	1	--	--
13	6	3	2	2
15	2	0	2	1
16	4	4	4	4
17	2	2	1	1
18	3	3	2	2
06	49	29	2	0
06-2	8	8	2	2
10	33	33	4	4
19	7	4	9	7
20	5	5	3	3
21	3	3	2	2
22	2	2	2	2
23	3	2	0	0
Total	**182**	**143**	**108**	**93**

Table 6: Summary of DCCs and NZs isolated from patient cytospins from Tübingen cohort.
*NZ stands for marker negative cells isolated by standard micromanipulation

4.2. Isolation of primary tumor DNA

Primary tumors are available as either fresh/frozen or formalin-fixed paraffin-embedded (FFPE). Only a handful of tumor tissue is available as fresh/frozen since their procurement is not part of routine diagnostics in most of the clinics. Consequently, we had three kinds of primary tumor material. 1) For the primary tumors from which cryo tissue was available, we developed a protocol to isolate single cells from tissue sections. 2) For the FFPE tumor blocks we tried to isolate single cells, however in majority cases FFPE tissue blocks did not have good quality. 3) Bulk genomic DNA was isolated from the tumor containing regions of the FFPE tissue block.

4.2.1. Single cell isolation from cryopreserved primary tumor tissue

Method development

In order to obtain single cells from cryopreserved primary tumor tissue, sections of 20-25 μm thickness were fixed in PFA and stained for epithelial marker cytokeratin (CK). After staining, an enzymatic digestion with collagenase and hyaluronidase was performed on the slide. Additionally the slide was incubated with a hypertonic solution of maltose (Goodenough and Gilula 1974; Mitra, Mishra, and Li 2013) based on reports about the use of osmotic pressure to disrupt the zona occludentes and gap junctions. Next, the slide was scratched with a 200 μl pipette tip and the contents were collected in microfuge tube containing maltose solution with pipetting vigorously for several times to obtain a single cell suspension (described in methods section 2.7.1). CK positive single-cells were picked with the manual-micromanipulator.

For testing the protocol, we used healthy mammary breast tissue sections (Figure 18A and B) and tested the DNA quality of cells after WGA with WGA QC-PCR (Figure 18D). We could obtain good quality single cells after the isolation protocol (9 out of 9 cells picked successfully showed GII4). Next we wanted to check if the protocol introduces any artefacts in the genome. We performed array CGH profiling of normal mammary cell as control and found that the genomic profile is balanced (Figure 18E), this means that if we would isolate tumor cells using this protocol, the aberrations detected in the tumor cells would be true.

Hence, we used this protocol for isolation of single cells from cryopreserved tumor tissues of patient breast tumors.

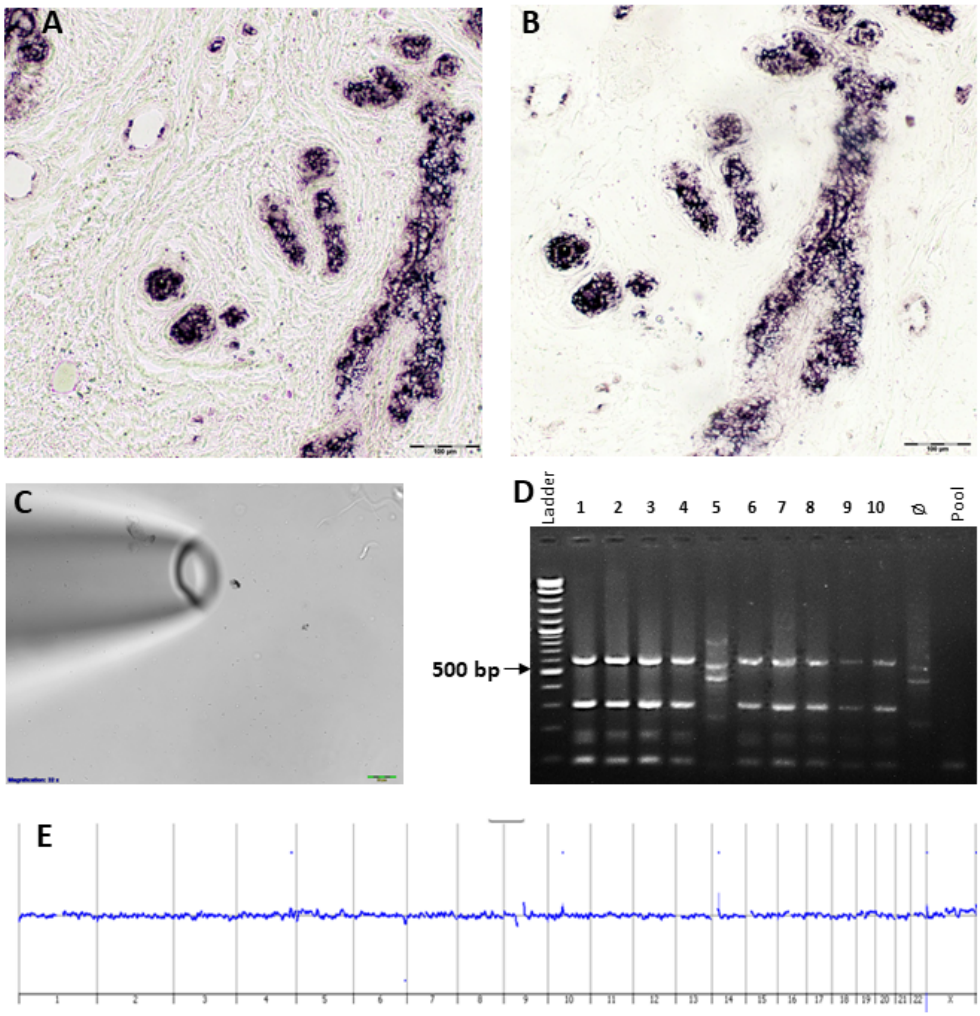


Figure 18: Isolation of single cells from flash frozen tissues. A: CK stained mammary tissue section 25 μm thickness), B: CK stained tissue section after enzymatic digestion, C: Cell isolation from single cell suspension by standard micromanipulation. D: Multiplex QC PCR results of cells picked using the method, the pool is picked up blindly, since the concentration of cells in the picking field is very low and digested/broken stromal cells could be picked up which could explain the GII of 1, E: Array CGH profile of a single cell isolated from healthy mammary tissue using the above protocol

Isolation of single cells from patient cryopreserved tumor sections

Using the method developed for isolation of single cells from frozen tissue sections (Figure 18), we isolated cells from the cryopreserved tumors of patient 01 and patient 02. We found that 57% and 59% of single cells isolated from patient 01 and patient 02 respectively had good DNA quality (GII>2). The results per patient are summarized in Table 7.

Patient	Picked single cells	Cells with GII>2	Success rate
01	89	51	57%
02	66	39	59%

Table 7: Summary of primary tumor single cells isolated from patient cryopreserved sections of breast tumors

4.2.2. FFPE tumor tissue

4.2.2.1. Single cell isolation from FFPE tissues

Single cell isolation was performed from 40-60 μm sections of tumor tissue using a protocol established by Menarini Silicon Biosystems described in detail in methods section 2.7.2. Briefly the tissue sections were deparaffinized, followed by antigen retrieval and enzymatic digestion. The single cell suspension was stained with anti-cytokeratin antibody and DAPI to identify tumor cells with nucleus.

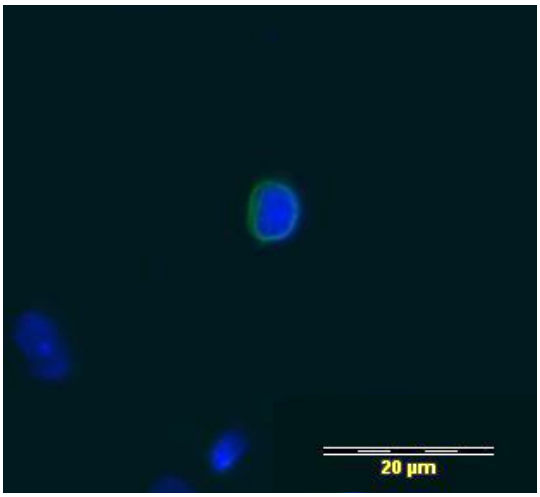


Figure 19: CK positive cell obtained from FFPE tissue.

Although the morphology of FFPE specimen is preserved, the proteins and nucleic acids are crosslinked and denatured during the fixation process and various chemical modifications take place which leads to introduction of artefacts in the DNA. Single cell WGA from such tissues is challenging. Pre-Cr enzyme mix from NEB is reversing those chemical modifications of DNA introduced by formalin fixation (except for DNA fragmentation) and the enzyme mix could be used for WGA of single cells (Martelotto et al. 2017). Dr Giancarlo Feliciello from ITEM-R tested the enzyme mix on single cells derived from FFPE blocks by introducing the enzyme repair step before MseI DNA digestion in the *Ampli*1 WGA protocol (marked in Figure 20A, described in details in methods section 2.9.2). The WGA products were evaluated by a QC PCR (Figure 20A) designed for FFPE samples (Developed by Dr Zbigniew Czyz in the laboratory, details in methods section 2.9.4). It was found that the addition of a repair step improved the DNA quality after WGA.

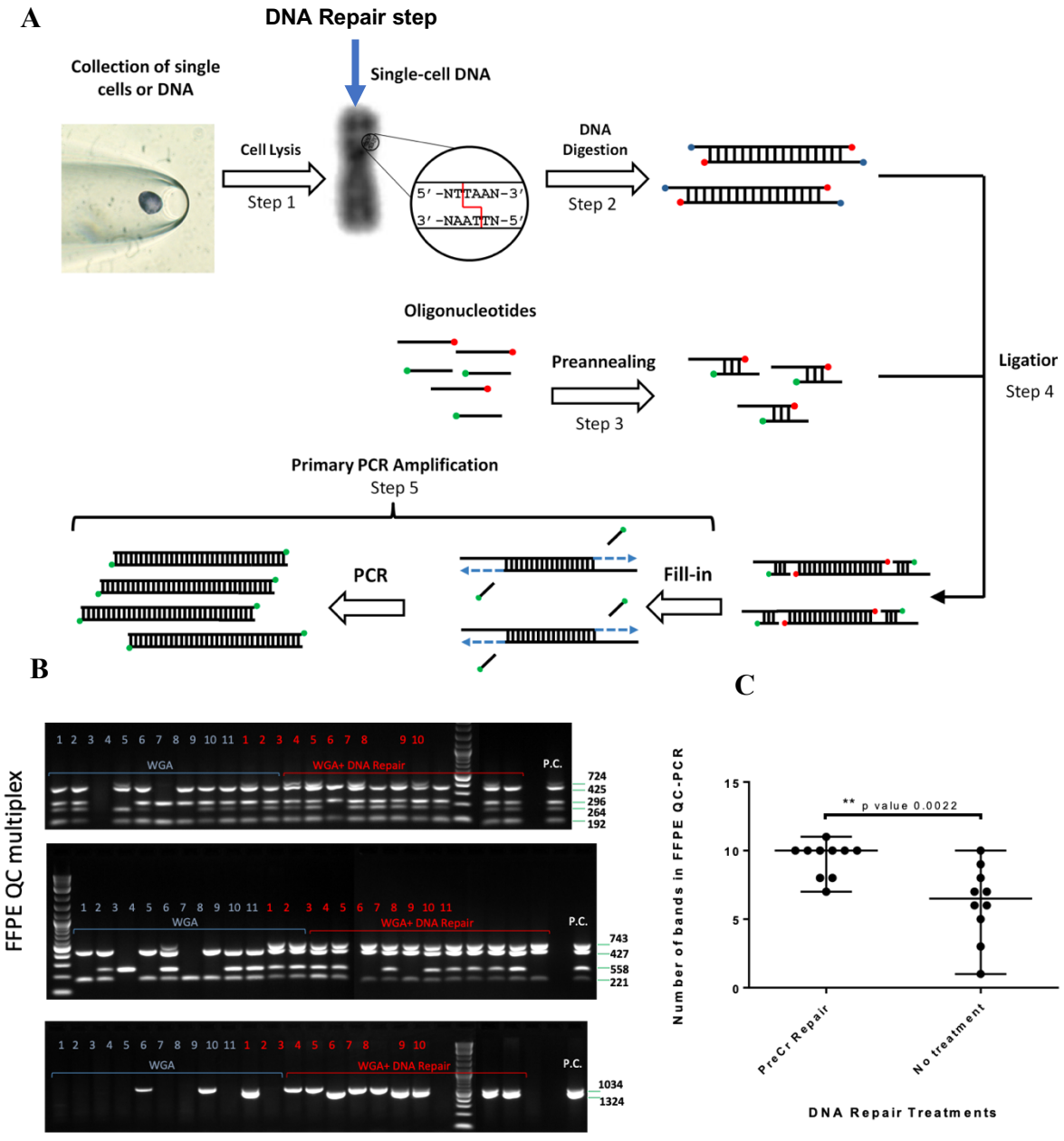


Figure 20: DNA repair prior to WGA. A: After cell lysis, the DNA repaired step is introduced, indicated by the arrow, following which DNA digestion with MSEI is continued, rest of the *Ampli1* WGA protocol remains constant where the repaired and digested DNA is ligated to the adaptors followed by gap filling and PCR amplification similar, figure adapted from (Czyz and Klein 2015). B: FFPE QC multiplex, comprises of set of 3 multiplex PCR; blue, FFPE single cells undergoing standard WGA; red, FFPE single cells undergoing DNA repair prior to WGA. P.C. is the pool of cells as positive control C: Quantification of figure B, there is a significant difference between the DNA quality of cell treated with PreCR Repair and the untreated cells, Mann Whitney Test p value 0.002.

We found that NEB offers the same enzyme mix as PreCR with a buffer optimized for NGS purposes sold as FFPE Repair mix. We tested this mixture on breast cancer single cells isolated from a previously tested good quality FFPE tumor block and found that there were no significant differences in the DNA quality of single cells treated with FFPE Repair mix or PreCR mix used by Dr Giancarlo Feliciello (Figure 21). Hence FFPE Repair mix was the choice of repair enzyme for future experiments.

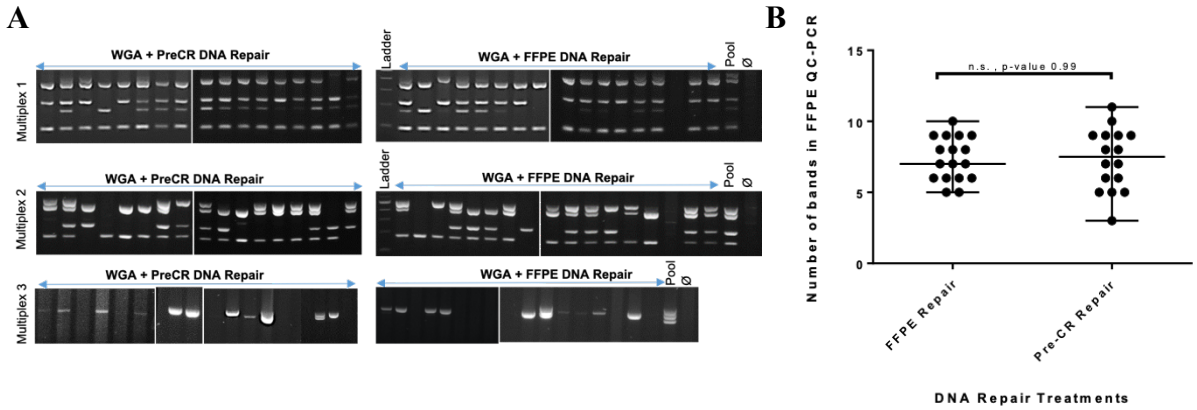


Figure 21: PreCR repair mix vs FFPE Repair mix. A: FFPE QC multiplex, comprises of set of 3 multiplex PCR, left: FFPE single cells undergoing DNA repair with PreCR repair mix before WGA, right: FFPE single cells undergoing DNA repair with FFPE repair mix before WGA. B: Quantification of A by Mann-Whitney Test, p value 0.99

Isolation of single cells from patient FFPE tumor sections

Next, we used the FFPE Repair mix on two patient blocks from the Tübingen breast cancer collective. This experiment was performed by Vincent Walter as part of his medical thesis project. Alas, we did not find any significant improvement facilitated by the FFPE Repair mix over the untreated WGAs when applied to patient tumor cells (Figure 22A). We also performed *Ampli*1^{TM} Lowpass CNA analysis and we were not able to observe any improvement between the cells processed with or without the repair step before WGA (Figure 22B). All profiles are provided in the Appendix 2.

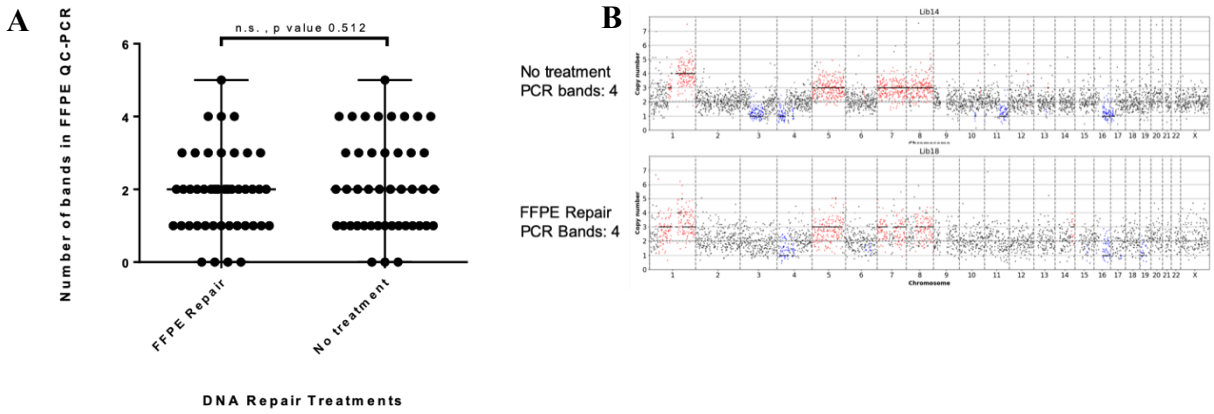


Figure 22: DNA repair on patient samples of Tübingen cohort. A: Quantification of FFPE QC PCR results, no significant difference was found in the single cells undergoing WGA with FFPE repair treatment or without repair treatment, Mann-Whitney Test p value 0.512. B: *Ampli*1^{TM} Lowpass CNA profiles, Top: without treatment, Bottom: FFPE repair treatment

4.2.2.2. Bulk genomic DNA isolation from FFPE tissues

As gDNA isolation of since single cells from FFPE tissue appeared sub-optimal for NGS purposes (indicated by the noisy CNA profiles in Figure 22B), we investigated bulk DNA extraction using the FFPE DNA isolation kit from Qiagen. Details are provided in the methods section 2.7.3. Briefly, tumor areas were marked by pathologist Dr. Florian Weber using hematoxylin and eosin sections, and the respective areas were isolated and digested following the manufacturer's instructions. The gDNA concentrations obtained from FFPE tumor blocks are summarized in Table 7. We were able to obtain good quality DNA from 9 out of 17 (53%) of FFPE tumor blocks tested so far from the Tübingen cohort as implicated from the 260/280 ratios, where a value of ~1.8 indicates good DNA quality.

Block ID	Concentration (ng/μl)	260/280 ratio
13	136.8	1.83
12	42	1.75
08	31.1	2.04
05	32.3	2.01
04 Region 1	67.1	1.87
04 Region 2	51.6	1.87
06-29551	161.6	1.86
06-22516-Region 1	27.7	1.89
06-22516-Region 2	48.5	1.48
06-24920	31.7	1.73
02-Region 1	120.3	2.03
02-Region 2	12.7	1.93
09	53.8	1.89
10 Recurrence	11.5	2.11
10 SN	2.4	2.05
11	29	2.02
03	7.3	2.42

Table 8: Summary of primary tumor gDNA isolated from patient FFPE tumor block from the Tübingen cohort.

4.3. Isolation of M1- Circulating Tumor Cells

To acquire tumor cells from patients with metastatic disease (M1), we obtained blood samples instead of bone marrow because bone marrow isolation is a highly invasive procedure. The CellSearch system was used for enrichment and enumeration of CTCs. For each patient we receive 3 CellSave tubes. The first tube goes directly into CellSearch, from the second tube plasma is isolated from the blood at first and the residual cell fraction goes into CellSearch, for the third tube, MNCs are isolated from the blood and adhesion slides are prepared for the isolation of macrophages and T-cells which serve as outgroups for cell lineage analysis (results in section 4.5.1). The procedure is outlined in Figure 23 (details in methods section 2.6.).

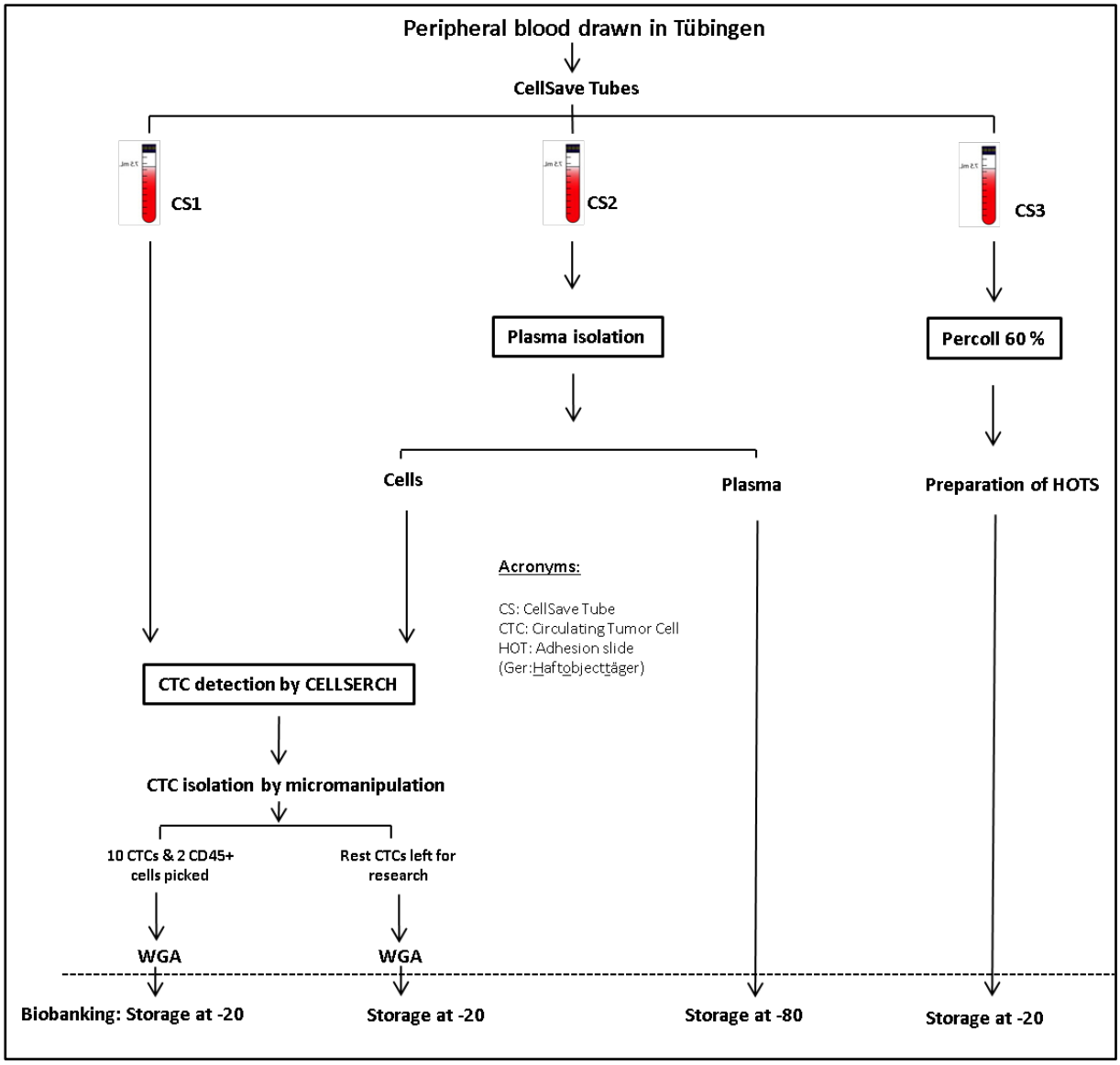


Figure 23: Blood samples from breast cancer patients. Tube CS1: Goes directly into Cellsearch. Tube CS2: Goes into into Cellsearch after plasma isolation. CS3: The blood is put through density gradient separation using Percoll to isolate PBL for preparation of HOTS (adhesion slides)

When a CellSearch run was positive, meaning CTCs were detected in the run (Figure 24A), the cartridge was flushed and the cells were manually screened (method's section 3.5). A CTC was selected when being CK+, DAPI+, CD45- (Figure 24B) and isolated by standard micromanipulation. Isolated CTCs went through *Ampli1* WGA followed by QC PCR (Polzer et al. 2014). DNA quality of CTCs after WGA was good (Figure 24C). CNA analysis of isolated CTCs by *Ampli1* lowpass also displayed good quality (Figure 24D).

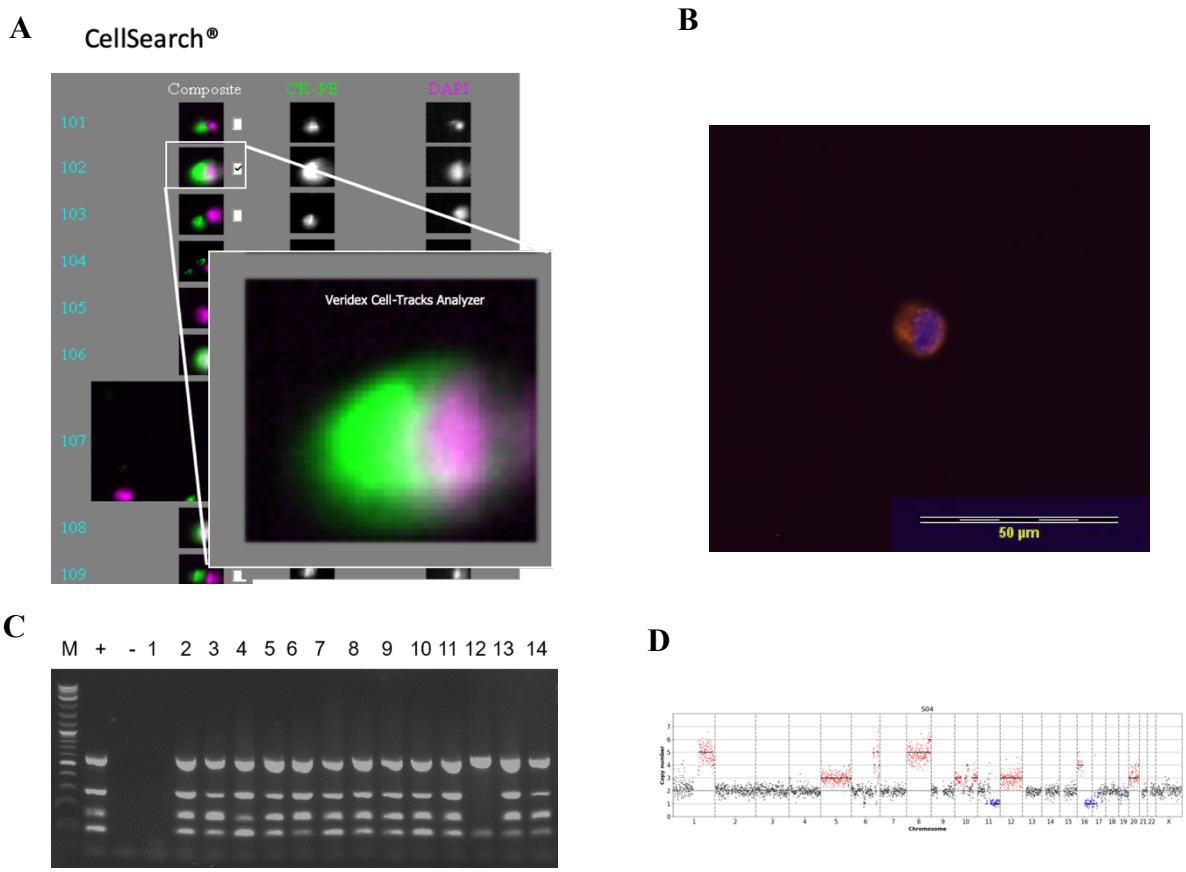


Figure 24: CTCs isolated from M1 breast cancer patients of the Tübingen breast cancer cohort. A: CTC recognized by the CellSearch machine. Cell is positive for cytokeratin (green) and DAPI (purple). B: CTC screened during manual isolation, Red (Cy3) for cytokeratin, Blue (DAPI) for nucleus. C: Cell quality of CTCs isolated from CellSearch cartridges. D: CNA profile of a CTC (patient 01) isolated from the CellSearch cartridge.

When a patient sample is negative for CTCs, blood is collected again at their upcoming doctor's visits. We had 88 samples (from 33 patients) in total, of which 31 (35%) were positive for CTCs (from 17 patients). Out of 33 patients enrolled in the study, during the span of 3 years from 2016 to 2019 we have been able to isolate CTCs from 11 patients (33%). Patient wise number of CTCs isolated are summarized in Table 9.

Patient ID	08	05	06	02	01	04	13	22	03	09	18
CTCs	75	51	48	27	18	9	7	7	3	1	1

Table 9: Summary of CTCs isolated from M1 patients.

4.4. Triplet cohort of patients

Bone marrow (BM) aspirates collected from 4864 M0 patients undergoing curative surgery at the Department of Gynecology, University of Tübingen between 2001 and 2016 were considered for the sample collective. Thirty-three patients (see Methods section 3.1) were finally enrolled in the study with the inclusion criteria of a long follow-up and availability of sample triplets comprising BM DCCs, primary tumor and CTCs/DCCs at a later time point before and after progressing into metastasis. Figure 25 delineates the steps through filtering the patients for the project. So far ten patients have been eligible for cell lineage reconstruction (Table 9).

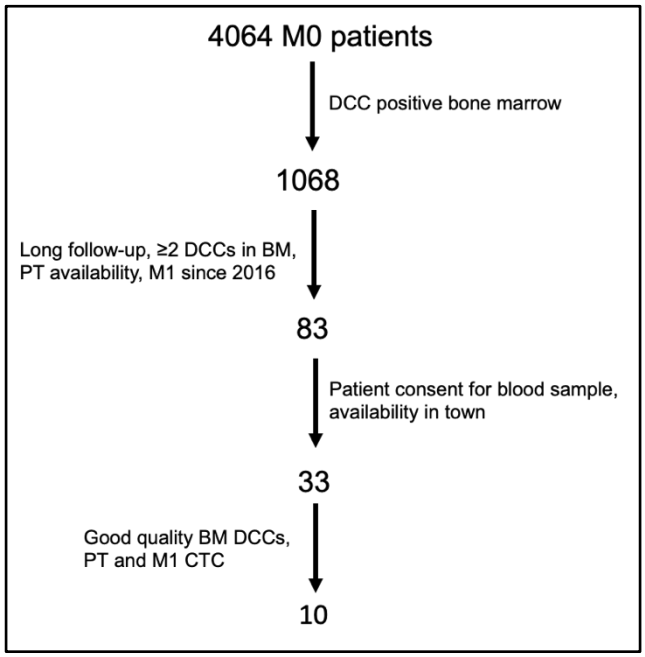


Figure 25: Steps leading to the current number of patients selected for cell lineage analysis. DCC positive M0 patients were selected, next the availability of frozen cytospins for DCC isolation, primary tumors and progression to metastasis were looked at. Fraction of the selected patients were visiting the University Hospital in Tübingen and consented for blood sample for CTC isolation from which till now we have good quality DNA samples from 10 patients.

Patient ID	Primary Tumor Classification	Type	ER	PR	Her2	Metastasis free survival (years)
03	cT3 cN3 M0 G2	NST	1	1	0	3
05	ypT1b ypN0 M0	NST	1	1	0	6
04	ypT3 ypN3 M0 G2	ILC	1	1	0	7
09	rpT1 Nx M0 G2	NST	0	0	1	15
02	pT3 pN3a M0 G2	NST	1	1	0	3
08	ypT3 ypN2a M0 G2	Inflammatory	1	1	0	1
01	pT3 pN1a M0 G2	ILC	1	1	0	7
18	ypT1c ypN2a M0 G3	NST	1	1	0	4
06	pT3 pN3a M0 G2	ILC	0	1	0	5
07	right ypT1 yN1 M0 G2, left: ypT3 ypN1 M0 G2	NST / ILC	1	1	0	14

Table 9 Characteristics of the patients eligible for cell lineage analysis. NST, invasive carcinoma of no special type or invasive ductal carcinoma; ILC, invasive lobular carcinoma; ER, estrogen receptor (1=positive, 0=negative); PR, progesterone receptor (1=positive, 0=negative); Her2, human epidermal growth factor receptor (1=positive, 0=negative)

4.5. Generation of metastatic phylogenies

We have started to generate cell lineage trees from single cells of primary breast cancers, DCCs isolated from bone marrow at the time or surgery, circulating tumour cells derived from metachronous metastases and outgroup cells (i.e. control cells from non-malignant lineages) to identify the metastasis founder cells (MFCs) in individual patients. Single cells from patients from the triplet cohort (Section 4.4.) were used to generate cell lineage trees based on Short Tandem Repeats (STRs) also known as microsatellites. STRs are an abundant source of somatic mutations (Woodworth, Girskis, and Walsh 2017) reflecting cell divisions which determines cellular descent by including random mutational events. Approximately, 12000 STR loci are targeted by duplex Molecular Inversion Probes (MIPs) and sequenced and analyzed by our collaborators in Dr Ehud Shapiro's laboratory at Weizmann Institute of Science, Israel.

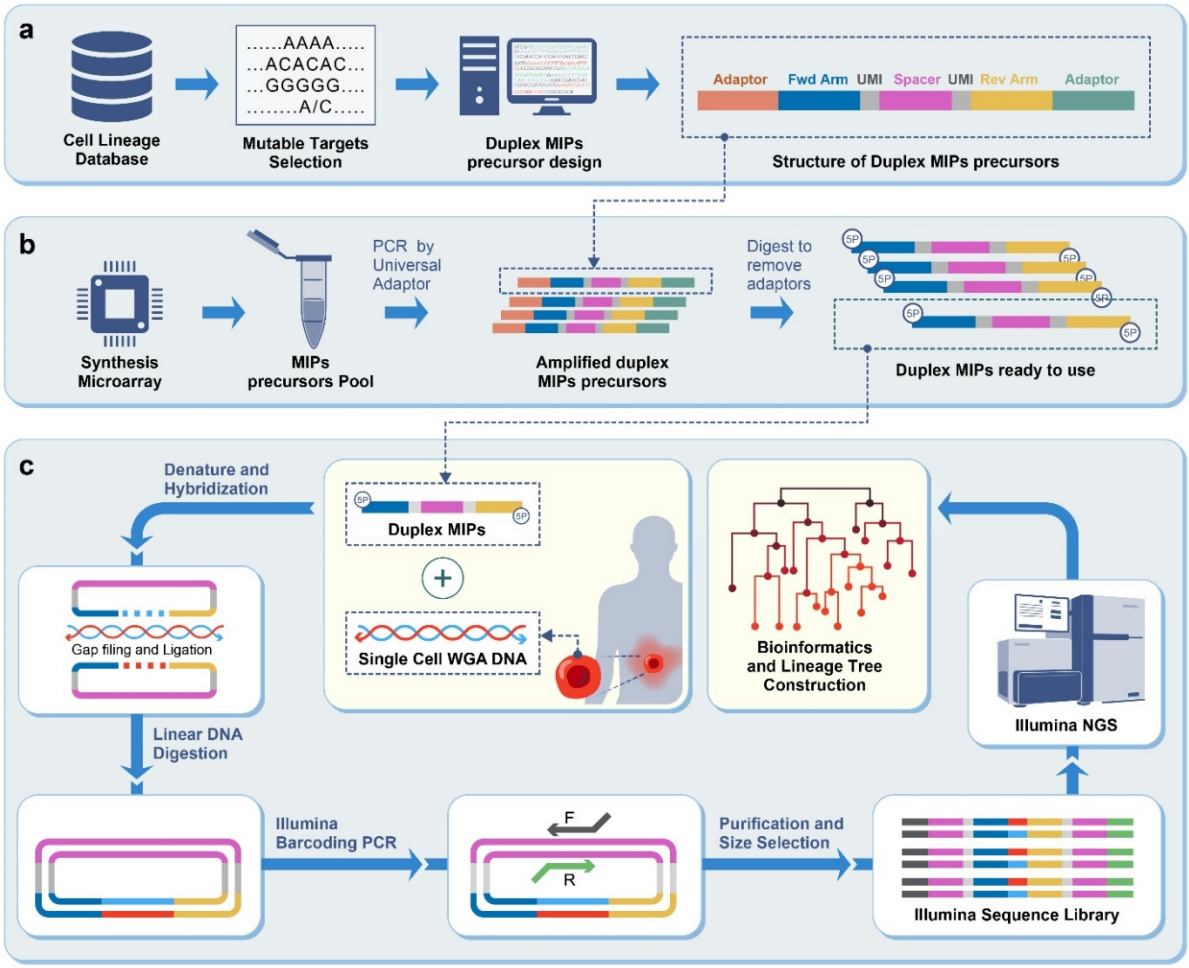


Figure 26: Duplex MIPs based cell lineage workflow. A: Design of duplex MIPs precursor: Precursors for desired STR targets are designed; B: Duplex MIPs preparation: duplex MIPs precursors are synthesized on a microarray, pooled and amplified by PCR, followed by removal of universal adaptors (red and green). The product is purified to obtain active duplex MIPs; C: Duplex MIPs are annealed (blue and yellow) to the flanking regions of the targets of the template, next the products are circularized. Linear DNA (excess MIPs and template DNA) is eliminated by exonucleases digestion. Next, Illumina sequencing libraries are generated for each sample by PCR amplification with respective barcodes and subsequently pooled and sequenced by Illumina NGS platform. Raw reads are analyzed to detect STR mutations for cell lineage tree reconstruction. Figure adapted from (Tao et al., 2018).

Molecular Inversion Probes (MIPs) are single strand DNA molecules with two targeting ends containing a linker between them. MIPs allow high throughput targeting of STR loci and have high specificity (Nilsson et al. 1994). Our collaborators, developed STRs targeting platform with duplex MIPs enabling the capture of thousands of STR loci, produced from precursors synthesized on microarrays (Tao et al. 2018).

Critical parameters beginning from sample preparation of *Ampli*1^{TM} WGA products, selection of outgroups and optimizing the required sequencing depth have been addressed and resolved in conjunction with Prof Ehud Shapiro's group in the recent years. Lowpass CNA analysis was performed for each cell in the lineage tree to authenticate the positioning of a cell in the tree. Bootstrapping analysis was performed to validate the clusters seen on the lineage tree. In this section, the results of optimizations performed on patient 01 samples and the resulting cell lineage tree is described.

4.5.1. Outgroups and root approximating group

In phylogenetics, outgroup rooting is an optional procedure that anchors the beginning of the timescale to the shared ancestor of that outgroup. Some reconstruction algorithms may use this notion of rooting as part of the reconstruction itself; such is the case with the later mentioned triplet max cut (TMC) algorithm. We used two hematopoietic lineage cell types (macrophages and T-cells isolated from blood) and one epithelial lineage cell type (oral epithelial cells isolated from buccal swabs) to serve as outgroups in our cell lineage analysis. The tumor cells are expected to cluster apart from the outgroups. Genomic DNA obtained from peripheral blood was used as root approximating group.

To isolate hematopoetic lineage cells, blood was collected from patients in CellSave tubes and adhesion slides were prepared from blood PBLs for single cell isolation which were stained against CD68 and CD3 to detect macrophages and T-cells respectively (Figure 27 A1, B1), and Method's section 3.5. The stainings were first tested in healthy donor samples. From patient's samples the positive cells were isolated by standard micromanipulation followed by *Ampli1* WGA and WGA-QC. Cells with GII>2 were selected for cell lineage analysis. We could successfully isolate around 25 CD68+ and CD3+ from each patient (Table 10).

Oral epithelial cells were added as an outgroup for serving as an epithelial lineage control which should ideally cluster separately from the hematopoietic lineage cells. To isolate oral epithelial cells (OECs), patients were requested to provide buccal swabs. We developed a protocol to isolate good quality cells from the buccal swabs and tested in healthy donor (method section 2.8.2) and thereafter requested our clinical collaborators at University Hospital Tübingen to execute the protocol. Briefly, the buccal swabs obtained from patients were immediately stirred into PBS and sedimented onto adhesion slides. Frozen slides were shipped to us from Tübingen. The cells were stained with hematoxylin (Figure 27C) and isolated by standard micromanipulation. We could successfully isolate 11-20 OECs from each patient (Table 10).

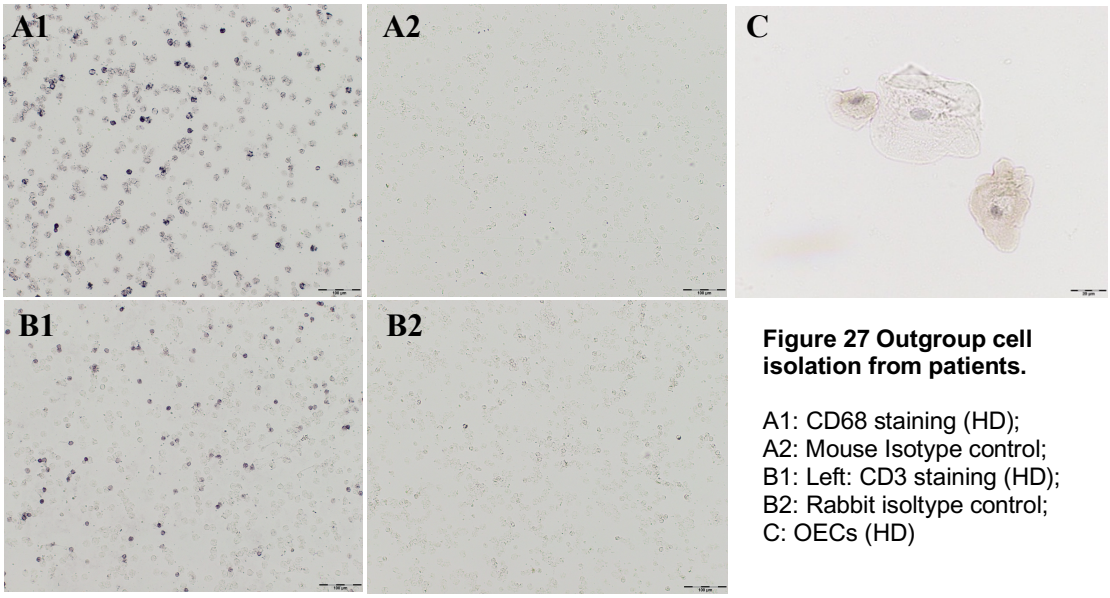


Figure 27 Outgroup cell isolation from patients.

A1: CD68 staining (HD);
A2: Mouse Isotype control;
B1: Left: CD3 staining (HD);
B2: Rabbit isoltype control;
C: OECs (HD)

Patient ID	CD68+ cells	CD3+ cells	OECs
01	25	25	11
02	25	25	22
06	24	21	20
05	28	23	n.a.

Table 10: Outgroup cells isolated from breast cancer patients.

4.5.2. Sample preparation for lineage tree

As primary single cell WGA material is precious, we needed to optimize the use of reamplified WGA products for sequencing. Therefore, various sample preparation protocols were tested. Combinations of original, reamplified and double strand synthesized (ds) WGA products were hybridized with different batches of MIP probes (Table 11). Groups C1-C4 were prepared entirely by our collaborators. We prepared groups C5-C9, with 20 good quality non-tumor single cells. Next, the samples were subjected to duplex MIP hybridization and sequencing at Dr Ehud Shaprio's laboratory.

Label	Reamplification	DSS	AMPure-XP Purification	MIPs preparation
C1	-	+	-	old
C2	-	+	-	new
C3	+	-	-	old
C4	+	-	-	new
C5	-	+	-	new
C6	+	+	-	new
C7	-	-	-	new
C8	+	+	+	new
C9	+	-	+	new

Table 11: Various protocols for sample preparation. C1-C2, original WGA product after DSS hybridized with old and new batches of MIP preparation respectively; C3-C4, reamplified WGA product hybridized with old and new MIP preparation respectively; C5, original WGA product after DSS hybridized with new MIP preparation; C6, reamplified and DSS and hybridized with new MIP preparation; C7, original WGA product hybridized with new MIP preparation; C8, reamplified, DSS and purified with AMPure-XP and hybridized with new MIP preparation, C9, reamplified and purified with AMPure-XP and hybridized with new MIP preparation.

Samples from all the groups C1 through C9 were sequenced and the mapping rates for each of the preparations were calculated (Figure 28). The results indicated that the AMPure XP purification step introduced for control groups C8 and C9 have the highest mapping rates (~93%). C9 was used as the protocol of choice for sample preparation, which involves reamplification and AMPure-XP bead purification.

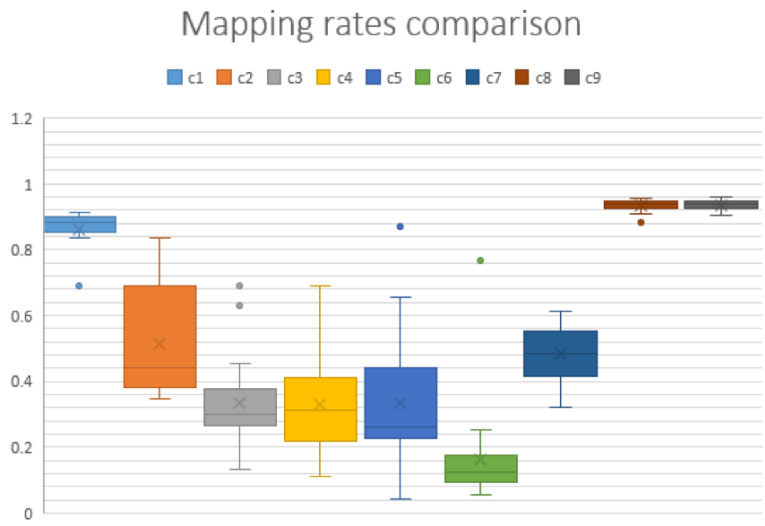


Figure 28 Optimization of sample preparation. Mapping rates of the reads for different types of preparation. The highest amount of mapped reads are obtained in groups C8 and C9.

4.5.3. Sequencing depth

Single cells from primary tumor, CTCs, DCCs, CD68+ macrophages and CD3+ T cells from patient 01 were initially sent for STR sequencing. Following a normalization MiniSeq run 84% of the samples fulfilled the passing criteria (which is >60% of the reads successfully mapped to target STR loci). The samples were sequenced iteratively. At first, the selected cells were sequenced to an average of ~1 million reads/cell, followed by ~4 million reads/cell and ~5 million reads/cell, subsequently. It can be seen in Figure 29, that the number of STR loci covered by >30 reads increases with an increase in sequencing depth, while saturation is reached at around 5 million reads/cell.

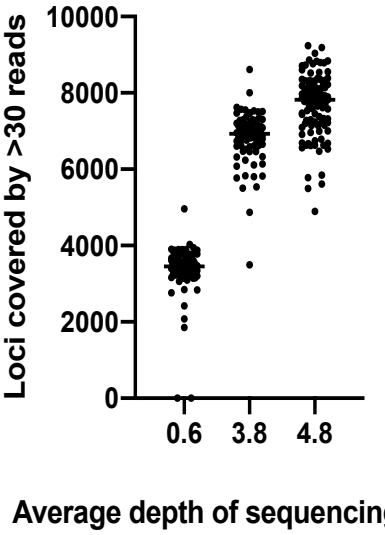


Figure 29: Depth versus Micro-satellite loci covered by more than 30 reads. The first round of NextSeq sequencing resulted in 0.6 x 106 reads per cell, where the median loci covered by >30 reads was 3,452. After two more round of NextSeq sequencing the average depth reached 3.8 x106 reads per cell, where the median loci covered by >30 reads was 6,932. Followed by a NovaSeq run which helped to achieve an average depth of 4.8 x106 reads per cell, where the median loci covered by >30 reads was 7,825.

Lineage trees at ~1 million reads/cell and ~4 million reads/cell are shown in Figure 30A and B, respectively. T-cells are used as the root approximating group in the trees. The trees are plotted using the Triplets Max Cut (TMC) algorithm which puts out all leaves at equal depth (ultrametric). Thereafter, a post process for depth calculation was applied. To calculate the depth, the genomic values of the reconstructed intermediate nodes were inferred for parsimonious values according to STR mutation probabilities derived from somatic mutation rates in the ex-vivo tree of DU-145 prostate cancer cell line (Biezuner et al. 2016). To determine genomic loci that are considered "classifying", a minimum group size of 3 was used, meaning an STR locus is included in the analysis if it presents at least two different signals and each of those two is attributed to at least three cells. Clustering of groups in the tree in Figure 30B is improved over Figure 30A, indicating the role of increased sequencing depth. Interestingly, in both trees it could be seen that the healthy cells (T cells and macrophages) cluster separately from the tumor cells. Within the tumor arm the M1-CTCs including the M0 DCCs tend to form separate cluster from the primary tumor cells.

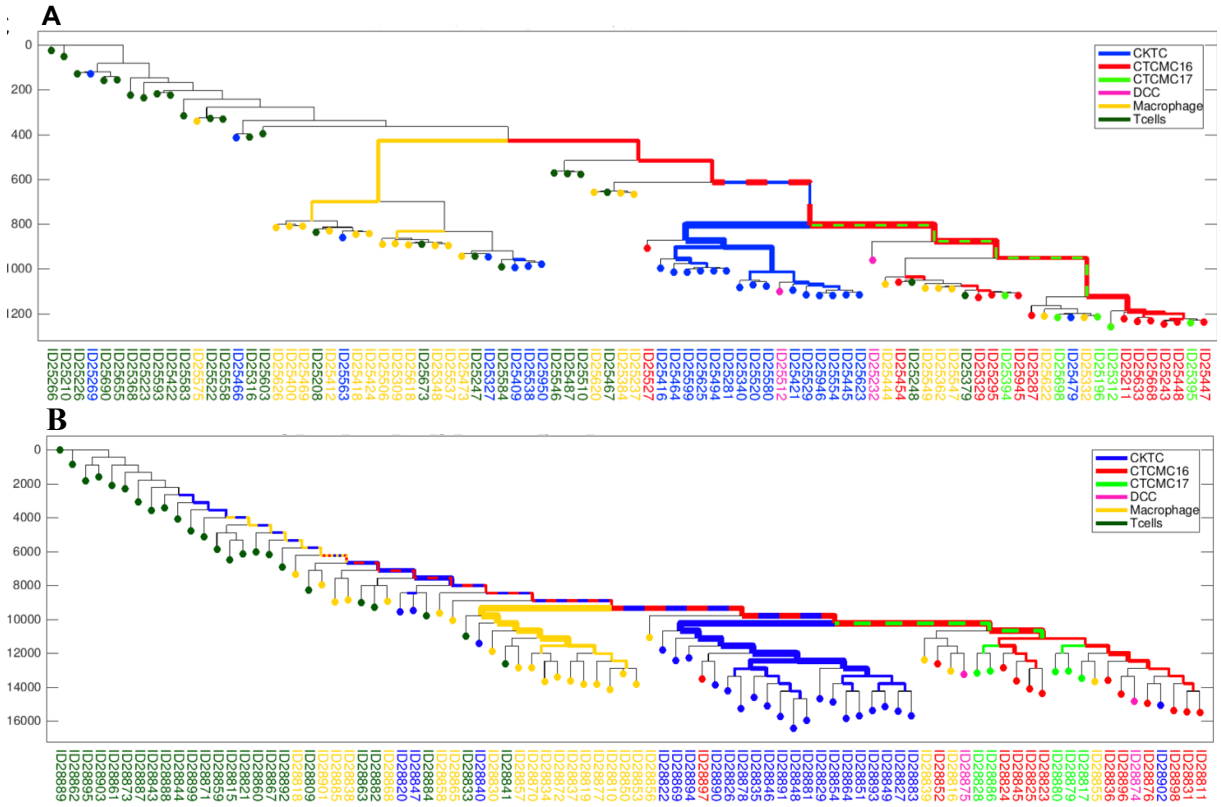


Figure 30: STR based lineage tree of patient 01. Primary tumor (single cell from fresh frozen tissue section) and bone marrow DCC (single cell from cytospin) were sampled in 2009. In 2016 the patient developed metastasis and blood was taken for CTC analysis (single cell from CellSearch). A panel of ~12000 microsatellite loci are sequenced to obtain single cell lineage separation; olive green, T-cells; yellow, macrophages; deep blue, single cells from primary tumor; red, M1-CTCs isolated in year 2016; light green, M1-CTCs isolated in year 2017; pink, bone marrow DCCs. The thickness of matching colors on the edges reflect the hypergeometric score of the edge's descendants, or how unlikely it is to draw X cells of a certain colour out of Y cells in total. A: sequencing depth at ~1 million reads/cell; B: sequencing depth at ~4 million reads/cell.

4.5.4. Addition of germline DNA and outgroups

In order to improve the lineage tree even further, germline DNA derived from blood MNCs was used for root approximation and OECs were introduced as an epithelial lineage outgroup. Simultaneously, improvements in tree reconstruction algorithms continued from our collaborators. A new likelihood-based approach was integrated together with a TMC algorithm. In this approach, likelihood-based heuristics decide on the topology of cells-triplets which are then fed to the TMC algorithm. The depth calculation algorithm remained the same as before. The resulting tree can be seen in Figure 31. We could again see that the healthy cells (T-cells, macrophages and OECs) cluster separately from the tumor cells, and interestingly the OECs form a separate cluster. Within the tumor arm the M1-CTCs including one of the M0 DCCs tend to form separate cluster from the primary tumor cells. However, the other M0 DCC clusters with the healthy cells (see arrow in Figure 31).

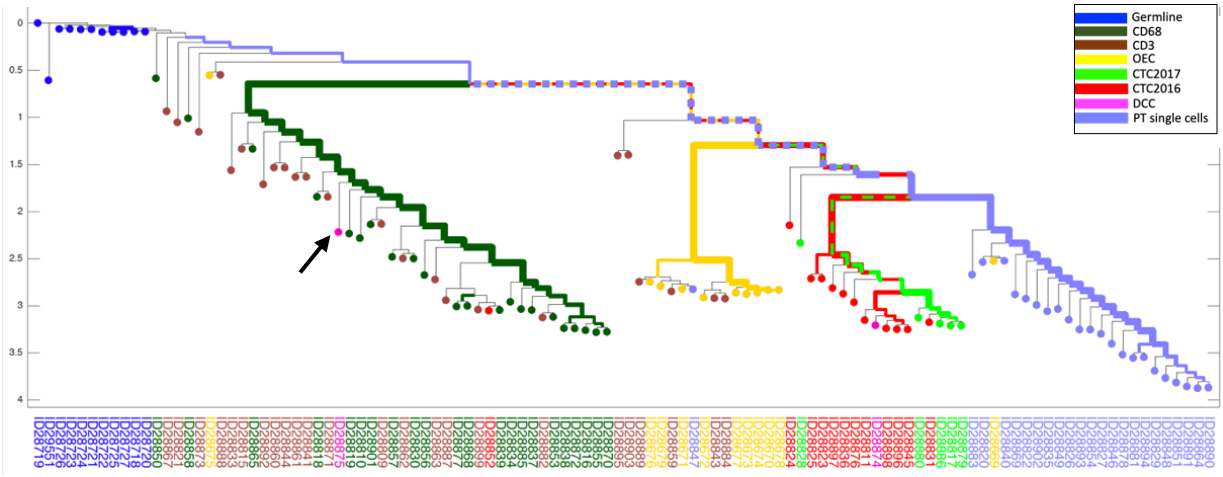


Figure 31: STR based Lineage tree of patient 01 at ~4 million reads/cell with germline DNA as root approximating group and OECs as another outgroup. Brown, T-cells; olive green, macrophages; yellow, OECs (as Kera in the legend); deep blue, Germline (as Bulk in the legend); light blue, single cells from primary tumor; red, M1-CTCs isolated in year 2016; light green, M1-CTCs isolated in year 2017; pink, bone marrow DCCs. The germline DNA samples and the OECs are however shallow sequenced in the tree. The thickness of matching colors on the edges reflect the hypergeometric score of the edge's descendants, or how unlikely it is to draw X cells of a certain colour out of Y cells in total. Arrow indicates DCC clustering with T-cells and macrophages.

To improve the clustering further, another round of deep sequencing was performed to achieve an average of ~5 million reads/cell, which would mean an increased number (median 7,825) of informative STR loci would get covered by greater than 30 reads (Figure 29). Additionally, advancements in the tree reconstruction algorithm used by our collaborators were also made. FastTree2 algorithm was used to generate lineage trees (Price, Dehal, & Arkin, 2010), which is a hybrid approach of the classical neighbor joining and maximum likelihood methods. The resulting tree is shown in Figure 32. Although we see a separation of healthy cells from tumor cells and a separation of the metastatic and primary tumor cells, one of the M0 DCCs still remains with the healthy cells (see arrow in Figure 32) and the separation within the healthy cells (T-cells, macrophages and OECs) does not appear robust. Next, we tried to validate the clusters by CNA profiling of all the cells on the tree.

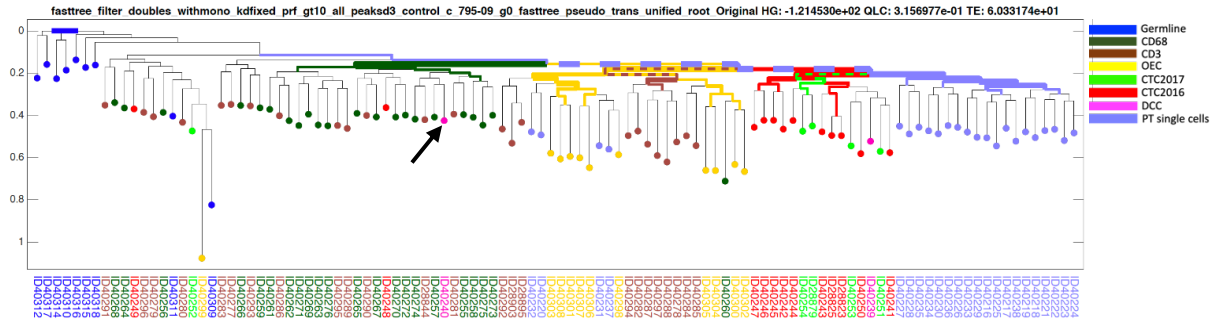


Figure 32: STR based Lineage tree of patient 01 at ~5 million reads/cell. Colours indicate tissue origin of each cell types (deep blue, CD3+ T-cells; red, CD68+ macrophages; green, M1-CTCs isolated in year 2016; yellow, M1-CTCs isolated in year 2017; pink, bone marrow DCCs; olive green, germline gDNA isolated from bulk of blood cells; light blue, OECs; brown, single cells isolated from primary tumor). The thickness of matching colors on the edges reflect the hypergeometric score of the edge's descendants, or how unlikely it is to draw X cells of a certain colour out of Y cells in total. Arrow indicates DCC clustering with T-cells and macrophages.

4.5.5. Validation

Since some tumor cells were clustering with the healthy cells (T-cells, macrophages and OECs) and we were not able to see clear separation of hematopoietic and epithelial lineage cells, we wanted to check the accuracy of the clusters on tree. We used CNA analysis to assess the genomic profiles of the cells in the tree and bootstrapping analysis as a bioinformatical tool to validate the robustness of the tree.

4.5.5.1. CNA profiling of cells on STR based lineage trees

CNA analysis was performed to achieve two aims, (i) To confirm that tumor cells are of cancer origin and outgroup cells don't contain cancer cells: It is possible that single cells isolated from the cryo primary tumor blocks are healthy cells of the mammary epithelium and therefore might be false positives. (ii) To construct cell lineage trees from CNA data. This aim exceeds the scope of the thesis and therefore is not presented here. *Ampli*1™ LowPass-Sequencing Technology as described in the method's section 2.14 was used for the CNA analysis which included all cancer cells and healthy cells in the lineage trees of patient 01. The results of CNA analysis for patient 01 are as follows: (i) Cancer cells: All tumor cells (Primary tumor single cells, DCCs and CTCs) displayed aberrant profiles. Some of the primary tumor cells (3 out of 23 single cells) displayed poor quality in terms of noisy profiles (DLRS values>0.35) and hence were removed from the lineage tree analysis. (ii) Outgroup cells: Macrophages and T-cells exhibit balanced profiles, nevertheless 11% of cells displayed an X-chromosome loss, but these profiles were counted as healthy and were not excluded from the analysis. Three out of 11 OECs (27%) displayed aberrant or noisy profiles (DLRS value>0.35). These three cells were discarded from the lineage tree analysis henceforth. Examples of CNA profiles are shown in Figure 33. All CNA profiles and a hierarchical clustering analysis with CNA data are provided in Appendix 3.

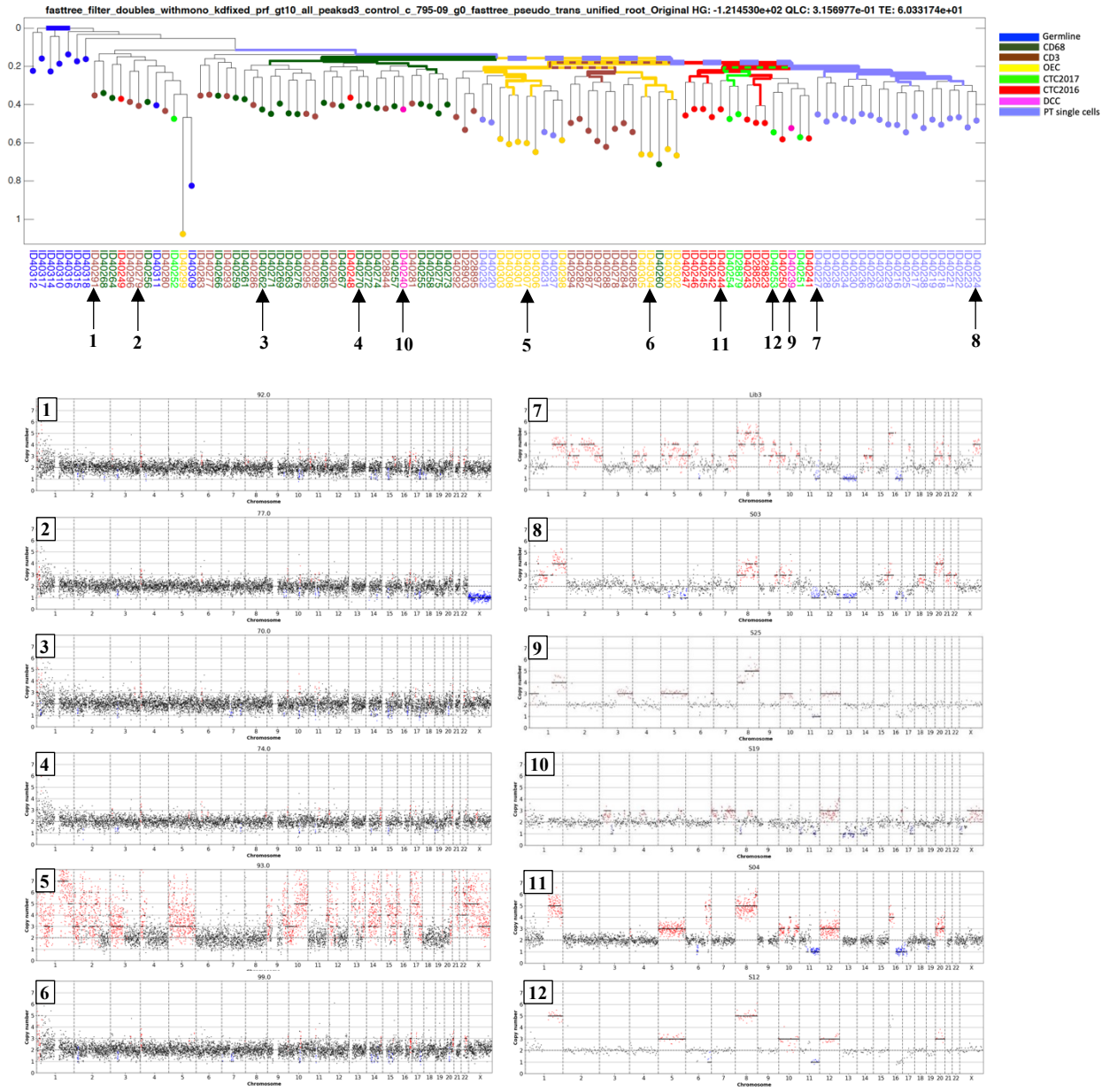


Figure 33: CNA profiles of cells on the STR cell lineage tree. CNA profiles of cells on the STR cell lineage tree. A: Ampli1 lowpass sequencing was conducted for all the cells of the lineage tree to validate the correctness of the tree. 1, 2 are T-cells; 3,4 are macrophages; 5,6 are OECs, 5 is an example of one of the OECs showing an aberrant profile; 7,8 are primary tumor cells; 9,10 are DCCs; 11-12 are CTCs.

The resulting lineage tree after filtering cells on the basis of CNA analysis can be seen in Figure 34. We see a separation of the cancer cells from the healthy cells. OECs form a separate cluster from the hematopoietic cells (CD3+ T-cells, CD68+ macrophages) which originate from distinct embryonic germ layers. Interestingly, we also note a significantly clear separation (as indicated by an asterisk on the tree; explained in the next chapter) of advanced primary tumor cells from metastatic CTCs. One of the M0-DCCs (isolated 7 years before emergence of the metastasis) clusters with the metastatic cells. Furthermore, the other M0-DCC is again grouped with the healthy cells (shown by arrow), but from CNA analysis we can confirm that the DCC shares aberrant regions with the other DCC, M1-CTCs and primary tumor cells (Figure 33, 10). This could suggest that the cell escaped from the primary tumor much earlier than its counterpart and is more 'normal-like'. Next, the significance of the clusters was evaluated by bootstrapping analysis

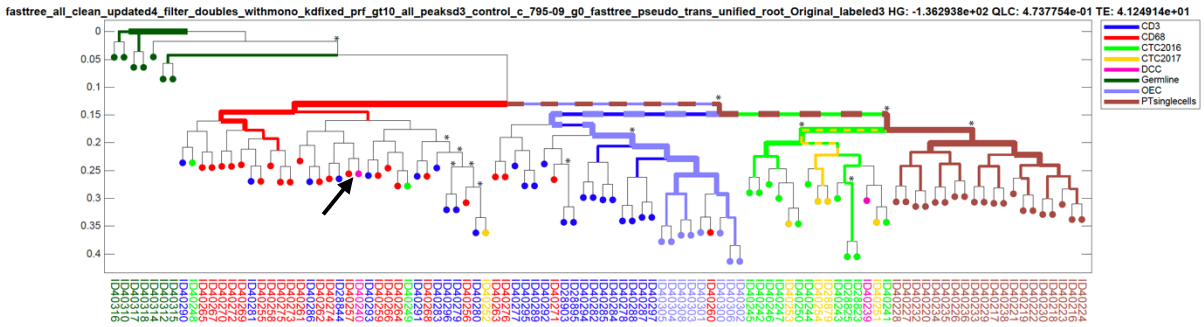


Figure 34: STR based Lineage tree of patient 01 after filtering cells by CNA profiles. Colours indicate tissue origin of each cell types (deep blue, CD3+ T-cells; red, CD68+ macrophages; green, M1-CTCs isolated in year 2016; yellow, M1-CTCs isolated in year 2017; pink, bone marrow DCCs; olive green, germline gDNA isolated from bulk of blood cells; light blue, OECs; brown, single cells isolated from primary tumor). The thickness of matching colors on the edges reflect the hypergeometric score of the edge's descendants, or how unlikely it is to draw X cells of a certain colour out of Y cells in total. Arrow indicates DCC clustering with T-cells and macrophages. *Indicates transfer bootstrap expectation (TBE) value of >70% (bootstrap proportions of ≥70% corresponds to a probability of ≥95% that the corresponding clade is real (Lemoine et al. 2018)).

4.5.5.2. Bootstrapping analysis

On the lineage tree in Figure 34, an asterisk denote that a particular branching is significant using bootstrapping analysis (Lemoine et al. 2018) which suggests that corresponding clades are true. Bootstrapping analysis refers to resampling of a dataset iteratively, conservatively, n loci are sampled with replacement out of n loci. To validate the robustness of clusters seen on the lineage tree, bootstrapping analysis was performed (1000 iterations) to assess if the proximity between cells within designated groups is substantially closer than within a randomly sampled group of cells (Figure 35A). Majority of the groups show consistent clustering, as seen from smaller edge count values (both mean and median) as compared to random sampling, which means the clusters seen in the lineage tree are not created by chance and are robust except for the M0-DCCs (n = 2 cells). We next took a closer look at the DCC cluster by plotting a histogram of distances (here as edge counts) between the two DCCs at all the 1000 bootstrap iterations, Figure 35B. Interestingly, in 8.4% of the bootstrap iterations M0-DCCs had edge count of 2 meaning that the two DCCs were reconstructed in the same branch. Thus, there is a subset of STR loci which support common ancestry, however current data needs to be further evaluated to understand the relationship between the two M0-DCCs.

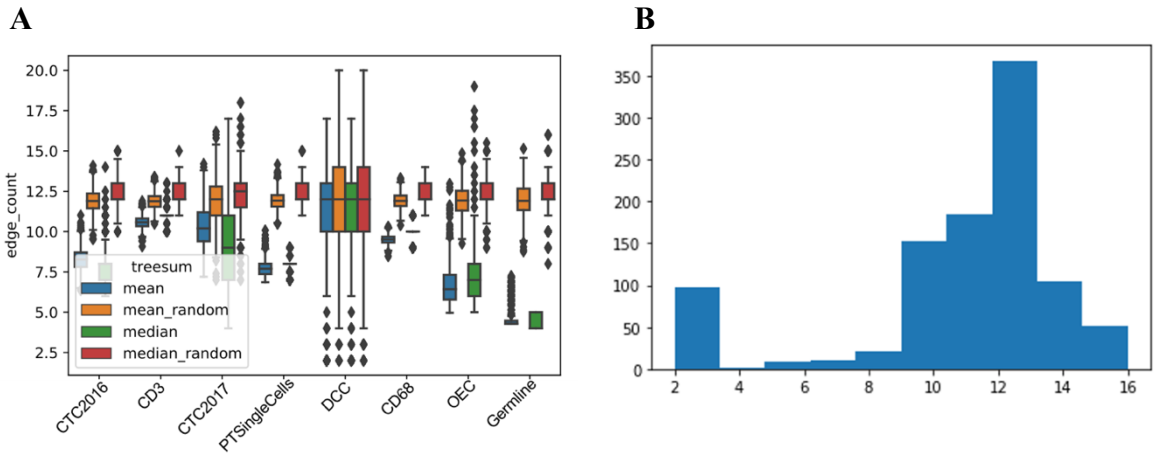


Figure 35: Bootstrapping analysis of patient 01 cells. A: Bootstrapping performed on the cell collective shown in Figure 34, The ordinate indicates the range of mean/median edge counts within group in the bootstrapping iterations. Note that for each group (except the group of DCCs comprising only two samples), the mean and median treesums are lower than random, indicating consistent proximity of the cells in a group. B. Histograms of distances (in edge counts) from the bootstrap iterations (1000 in total) within the two DCC. Edge count 2 would mean two cells are nearest to each other which is seen in 8.4% of the iterations, median edge count is12.

5. Discussion

Various studies from our lab and from around the world have proved that dissemination of cancer cells is an early event as opposed to the classical linear progression model (Klein 2013). However, are these early DCCs capable of giving rise to metastasis? In mouse models it had been proved that 80% of the metastases arose from the early DCCs (Hosseini et al. 2016). Based on the growth rates of tumor and metastasis, mathematical model of early disseminating cells giving rise to metastasis can be deduced (Klein and Holzel 2006), however it is challenging to identify the actual MFCs in patients. In this study, we strived to obtain single-cells by longitudinal sampling of primary tumor, M0-DCCs and M1-CTCs to establish cell lineage trees of breast cancer patients to identify MFCs. We addressed two aims, i) technical: We established a protocol to isolate DCCs from frozen archived cytospins and were able to mine the archive for selecting patients with a long follow-up; ii) biological: We were able to identify MFCs in a breast cancer patient with 7 years of progression using *de novo* STR mutations.

Genomic analysis of paired metastases and primary tumor samples from carcinoma patients in many other studies concluded MFCs disseminated late from the primary tumor, supporting the linear progression model (Leung et al. 2017; Sanborn et al. 2015; Yates et al. 2017). A study involving whole genome sequencing of colorectal cancer samples indicated that 65% of somatic mutations were common in primary tumor and metastases which indicated linear progression (Ishaque et al. 2018). Also a breast cancer study involving whole exome sequencing of paired samples from 6 patients claimed the incidence of linear progression in 5 out of 6 cases (Kroigard et al. 2017). Many facets including technical and clinical differences and also our understanding of the disease could account for such inconsistencies. Bulk sequencing might not be suitable to unravel the heterogeneity of tumors and metastasis. Also, most of these sequencing studies apply the infinite sites assumption as used in species evolution, which means that a genomic location can mutate only once during the lifetime of the cancer which might not be true as seen from single cell sequencing data (Kuipers et al. 2017). Experimental and clinical factors could also influence the studies, for example in the breast cancer study (Kroigard et al. 2017), it was noted that majority of samples analyzed were from synchronous lymph node metastasis rather than distant metachronous metastasis (which is mostly common), and in 3 out of 5 patients showing late dissemination, only one metastasis was analyzed per patient.

To our knowledge this is the first single-cell study where longitudinally collected samples of primary tumors, and individual pairs of DCC/CTC before and after detection of systemic metastasis of one individual patient have been acquired and analyzed by phylogenetics using STR mutations which don't possess selection advantage.

5.1. Patient cohort and method development

Longitudinal tracking of patients is a challenging task which demands availability of good quality patient material at several time points through progression. The inclusion criteria of a long follow-up, M0 at surgery, and availability of sample triplets comprising BM DCCs, primary tumor and CTCs/DCCs at a further time point before and after progressing into metastasis

was used. Bone marrow (BM) aspirates collected from 4864 M0 patients undergoing curative surgery at the Department of Gynaecology, University of Tübingen between 2001 and 2016 were considered for the sample collective, out of which 1068 were DCC positive. Finally, 83 patients were shortlisted for the project and 33 (0.03% of total DCC positive patients) were enrolled for donating blood sample at M1 stage. We have been successful in acquiring triplet samples from 10 patients so far. Reasons for patient drop-out included (i) no progression to M1 within follow-up time (most frequent); (ii) low DCC-DNA quality; (iii) patient lost in follow-up; (iv) patient in M1 stage but no CTCs isolated.

Isolation of DCCs from archived cytospins

In 2006 a consensus protocol for the method was published, which includes the preparation of slides and analysis of 2 million bone marrow cells per patient (Fehm et al. 2006). Since DCCs are so rare, to increase the number of cells screened at minimum cost most investigators (Riethdorf, Wikman, and Pantel 2008) therefore prepared cytospins at high cellular density. Since thousands of slides are still stored and would be amenable for single cell genomic analysis, and a long patient follow-up is available we tried to establish a protocol for single cell isolation from archived, cryo-preserved bone marrow cytospins.

The technical difficulty to isolate DCCs with high quality DNA was overcome so that we now can mine the sample base from various clinics. At first, we confirmed that BM-DCCs obtained from frozen cytospins display high quality DNA irrespective of the age of preparation. We could demonstrate this after micro-manipulator-assisted isolation and WGA of cells from areas of low cell density (Results section 4.1.1, Figure 3). However, isolation of single cells from cytospins by micromanipulation is normally impossible due to high density and strong adherence of cells to the slide. Several methods were set up and tested to isolate single DCCs of good DNA quality after WGA. We compared the Arcturus system which is an infrared based laser microdissection system and the PALM laser microdissection systems for isolation of single cells from cytospins and were only successful using the PALM (Sluka et al. 2008; Vandewoestyne et al. 2013), which is based on UV laser and offers laser pressure catapulting (LPC) for contact free isolation of single cells. The approach was then evaluated by the efficiency to isolate a cell and the quality of the DNA after WGA. Finally, good quality WGA samples were tested by array-CGH.

We standardized sample preparation protocol, (i) Staining of the background was vital for obtaining precise laser cutting, as the sample needs to absorb laser energy in order to allow “getting cut/ablated” (Vogel et al. 2007) and (ii) Sample dehydration was improved for laser microdissection by using reagents at -20 °C, an improvement over the manufacturer’s protocol also shown by other group by using ice-cold reagents (Sturm et al. 2013). Single cells having gone through this sample preparation procedure showed good quality DNA. (Results section 4.1.2.1)

Next, we had to standardize LPC from archived cytospins. We reasoned that lower the laser power needed to catapult the cell, better the DNA quality. High UV extinction coefficient of a sample corresponds to a lower LPC power required to isolate the cell (Vogel et al. 2007). Therefore, we tested several measures to increase the UV extinction coefficient for cells on archived cytospins. These included dark dyes, applying UV absorbing dyes and UV filters. In

total we compared 7 different protocols (Results section 4.1.2.2, Table 4). Unfortunately, we learned that DNA damage by LPC could not be avoided, since array-CGH generated highly artificial profiles (Results section 4.1.2.2, Figure 9,11,12). We finally established a protocol that allows us to obtain high quality DNA from archived cytospins by using a combination of laser microdissection and manual micromanipulation, which we abbreviated as LCAM. Laser cutting was used to ablate the cells in the surroundings so that ample space remains for manual micromanipulation. The genome profiles of cells obtained by array-CGH were as good as cells picked by micromanipulation (Results section 4.1.2.2, Figure 14). We also checked if the method allowed a scope of contamination from stray DNA, and we found that contamination is minimal and at par with the cells isolated manually. (Results section 4.1.2.2, Figure 16).

Finally, LCAM was applied to isolate M0 DCCs from the cohort of Tübingen patients, DCCs from 23 BM samples have been isolated so far out of which 2 patients dropped out due to poor quality.

5.2. Isolation of CTCs from M1 patients

We used M1-CTCs obtained from blood as proxy for the metastases tissues which could be a heterogeneous population of cells arising from different metastatic foci. Since primary tumor had been removed long ago, the CTCs in peripheral blood are most likely arising from the systemic metastases. Although, chances of finding metastatic cells would be higher in the bone marrow (Janni et al. 2000), we obtained blood samples since bone marrow extraction from patients (other than during surgery) would be too invasive and painful.

We used the FDA approved CellSearch system which is the gold standard for detection of CTCs in breast cancer (Cristofanilli et al. 2005). A multi-centre analysis involving 1944 patients showed that nearly 70% of metastatic breast cancer patients were CTC positive, and the CTC numbers although varied with response to treatment, number of metastatic sites, serum tumor bio-markers, did not vary with the tumor subtype (Bidard, Proudhon, and Pierga 2016). Thus, we expected to obtain CTCs from all breast cancer subtypes. A previously published workflow for CTC isolation and single cell genome analysis was used (Polzer et al. 2014).

We found 17 out of 33 M1 patients positive for CTCs which could be because of good therapeutic response of the patient leading to reduced CTCs in circulation. Also, the CellSearch system applies EpCAM based positive selection, which might lead to loss of possible EpCAM negative CTCs in some patients (Sieuwerts et al. 2009). We were successful in isolating CTCs from 11 patients (33%), which could be attributed to patients where few CTCs are detected, the cells could be lost during washing of the cartridge while retrieving the cell suspension to isolate cells.

5.3. Isolation of single cells from primary tumour tissue

For lineage tree analysis we ideally need cells from the primary tumor. In 2 cases out of 10, primary tumor tissue was available as cryopreserved, so high quality DNA could be obtained from such tissues. We succeeded in setting up a protocol to isolate intact single tumor cells from the flash frozen tumor tissues. After WGA, single cell DNA displayed high quality and was also suitable for genomic analyses like arrayCGH (Results section 4.2.1, Figure 18).

However, since the protocol is based on anti-cytokertain staining of the tissue sections, healthy, non-tumor cells from the breast could also be isolated, all cells taking part in cell lineage tree are subsequently analyzed for CNAs, hence confirming the tumor cells.

Primary tumors are mostly available as FFPE (formalin-fixed and paraffin-embedded) tissues which is the standard in surgical pathology, however is characterized by degraded low-quality DNA (Hognas et al. 2018) owing to DNA fragmentation and chemical modifications, prevalence of which varies with the pre-processing conditions, age of preparation etc (Hognas et al. 2018; McDonough et al. 2019). WGA of single cells from FFPE tissue was demonstrated in a study (Martelotto et al. 2017) where they used a cocktail of DNA repair enzymes from NEB (sold as either PreCr repair mix or FFPE DNA repair mix) supposedly to reverse the chemical modifications found in FFPE, however in this study the blocks used were pre-selected by a QC multiplex PCR. Due to the limitation of our triplet sample cohort, we did not have a choice to discard blocks. In the limited blocks we tested from our cohort, single cells performed the same way with or without treatment with FFPE repair mix (Results section 4.2.2.1, Figure 22). Finally, for majority of our tumor blocks we performed bulk DNA isolation from tumor containing areas of the block.

5.4. STR based cell lineage tree analysis

Somatic mutations which spontaneously occur in over a million STR regions during mitosis renders cells with a unique genetic signature, which could be theoretically tracked to obtain cell lineage trees with a probability of 99.95% (Frumkin et al. 2005). In our case, to obtain cell lineage trees from longitudinally acquired samples from breast cancer patients, around 12000 STR loci were sequenced after target enrichment using a patient-generic panel of duplex MIPs. STR loci are prone to stutter noise and it is difficult to correctly genotype short low-complexity reads, however these could be mitigated by a targeted capture using MIPs and a special mapping strategy (Carlson et al. 2015; Raz et al. 2019). This technique was scaled up to a targeted capture of more than 10000 loci by Dr Ehud Shapiro's group (Tao et al. 2018). Next, this MIP panel was tested on our *Ampli*1 WGA samples.

Critical parameters beginning from selection of outgroup, sample preparation of high quality *Ampli*1^{TM} WGA products to optimizing sequencing and tree reconstruction algorithms were addressed and resolved in conjunction with Ehud Shapiro's group in this project. (i) Outgroups were included as ground truth within the cell lineage tree and we selected T-cells and macrophages from blood as hematopoietic lineage outgroup and oral epithelial cells (OECs) as epithelial lineage outgroup (Results section 4.5.1, Figure 27). Acquiring these cell types from patients was relatively easier as they could be requested for blood and buccal swabs at the same time at which blood is collected for M1-CTCs. (ii) We found out that reamplification and an AMpure XP bead purification enhances the performance of *Ampli1* WGA products in the duplex MIP target enrichment pipeline (Results section 4.5.2, Figure 28). (iii) Sequencing depth was a crucial parameter, increasing sequencing depth, increased the number of accurately genotyped STR loci. It was simulated previously that correctness of a triplet tree reconstruction increases with increasing number of STR loci being genotyped as a function of depth in between the cells (Spiro and Shapiro 2016). In other words, when cells are closely related, a greater number of STR loci would be required to reveal the clonal relationships with accuracy. This fact is vital for our early disseminating DCCs and normal outgroup cells, since they are closer to the germline unlike progressed cancer cells. We found that at a sequencing

depth of ~5 million reads/cells, the median genotyped STR loci reach ~8000 (Results section 4.5.3, Figure 29).

5.5. Cell Lineage Tree of patient 01

We wanted to identify the metastasis founder cells in a lobular carcinoma patient, ER/PR positive and Her2 negative, diagnosed with clinical metastases after 7 years of primary surgery. STR based lineage tree analysis should ideally reconstruct the true phylogeny of cells and indeed it is reflected in our data. At first, we see a separation of the cancer cells from the healthy cells, which improves at every iteration of sequencing from ~1 million reads/cell, to ~5 million reads/cell (Results section 4.5.3, Figure 30). Next, OECs form a separate cluster from the hematopoietic cells (CD3+ T cells, CD68+ macrophages) which originate from distinct embryonic germ layers (Results section 4.5.4, Figure 31,32). Interestingly, we also note a clear separation of advanced primary tumor cells from metastatic CTCs. Here, one of the M0-DCCs (isolated 7 years before emergence of the metastasis) clusters with the metastatic cells, not with the advanced cells of the primary tumor.

Furthermore, the other M0-DCC is grouped with the healthy cells, irrespective of the tree reconstruction algorithm used (TMC or FastTree2). One could speculate that the cell escaped from the primary tumour much earlier than its counterpart, and is more ‘normal-like’, owing to which the separation from healthy cells is imperfect. Three M1-CTCs also cluster with the hematopoietic normal cells. Which could also have been dormant DCCs/micro-metastases which started circulating recently.

5.6. Accumulating evidences supporting patient 01 lineage tree

We gathered a substantial amount of information from the CNA analysis (Results section 4.5.5.1, Figure 33 and Appendix 3). We found that outgroup cells were indeed normal (balanced profiles) except for 3 OECs which could be attributed to bad sample quality as seen from the noisy profiles. We could obtain an overview of the tumor cells. (i) M0-DCCs are seen to be heterogeneous, the DCC (ID40240) clustering with the outgroup cells is particularly genomically immature, as has been reported earlier (Klein 2013). (ii) Despite clustering with the hematopoietic outgroup, this cell is a tumor cell since the CNAs are shared with both primary tumor cells and M1-CTCs. (iii) Three M1-CTCs found clustered with the hematopoietic outgroup, are also genomically immature than its counterparts which suggests that there is clonal heterogeneity within metastases, and these cells could be closest to the founders of a slow growing, low abundant clone (see appendix 3).

Next, the robustness of clusters seen on the lineage tree could be quantified by bootstrapping analysis (Shlush et al. 2012) , see Results section 4.5.5. Figure 34, 35.

5.7. Limitations

There are various limitations of this study. More M0 DCCs should have been included in the analysis, but due to the retrospective nature of the study and to obtain patients with a long follow up we had to resort to using archived cytospins which are limited in quantity. However, this can be mitigated by analyzing more patients. We already have 9 more patients in the pipeline for lineage reconstruction. Next, there is a chance that single cells that we isolated

from the primary tumor might not be enough to capture the entire clonal heterogeneity of the primary tumor. Furthermore other methods of CTC isolation like cell morphology based approaches (Guo et al. 2019) could be explored to obtain EpCAM negative CTCs. Our cohort is biased towards luminal cancers (hormone receptor positive), attempts to analyze more patients of each subtype is underway to identify MFCs across all breast cancer subtypes.

From the bioinformatics side, several aspects must also be improved. It is noteworthy that the STR signals could be ambiguous due to the presence of CNAs, especially if signals come from multiple alleles of an STR loci (Biezuner et al. 2016). This bias introduced by the CNAs might also lead to the placing of cancer cells with the hematopoietic lineage cells. Hierarchical clustering analysis with our CNA data (Appendix 4) shows that the M0-DCC and M1-CTCs which have fewer CNAs cluster separately from the rest of the M1-CTCs, interestingly, the same cells are also clustered away from the M1-CTC cluster in the STR based cell lineage tree (see ID40248, ID40249, ID40240 in Results section 4.5.5.1 in Figure 34). Next, bi-allelic STR genotyping could be improved in future. Inclusion of STR loci with mono repeats which have higher mutation rates could also result in better clustering of cells which undergo fewer divisions.

5.8. Conclusion

To conclude, we successfully obtained longitudinal samples from breast cancer patients before and after progressing to metastatic disease, with the help of existing and newly devised protocols. Lineage tracing using STR loci in patient 01 showed that all the metastatic cells cluster apart from the advanced primary tumor cells with a detectable M0-DCC ancestor, thereby contradicting the linear progression model of metastasis. STR based phylogenetics provides a more unbiased approach compared to studies relying on driver mutations or CNA data of paired metastasis and PTs to reconstruct cancer lineage trees (Navin et al. 2011; Ishaque et al. 2018), which do not take convergent evolution into account. Deeper genomic analysis of these M0 DCCs by whole exome sequencing could reveal mutational signatures decisive for progression and help understand the complex mechanisms of tumor evolution for improving therapeutic interventions.

6. References

Abdelwahab Yousef, A. J. 2017. 'Male Breast Cancer: Epidemiology and Risk Factors', *Semin Oncol*, 44: 267-72.

Ades, F., D. Zardavas, I. Bozovic-Spasojevic, L. Pugliano, D. Fumagalli, E. de Azambuja, G. Viale, C. Sotiriou, and M. Piccart. 2014. 'Luminal B breast cancer: molecular characterization, clinical management, and future perspectives', *J Clin Oncol*, 32: 2794-803.

Ahmed, F. E. 2006. 'Laser Microdissection: Application to Carcinogenesis', *Cancer Genomics Proteomics*, 3: 217-25.

Al-Mahmood, S., J. Sapiezynski, O. B. Garbuzenko, and T. Minko. 2018. 'Metastatic and triple-negative breast cancer: challenges and treatment options', *Drug Deliv Transl Res*, 8: 1483-507.

Allard, W. J., J. Matera, M. C. Miller, M. Repollet, M. C. Connelly, C. Rao, A. G. Tibbe, J. W. Uhr, and L. W. Terstappen. 2004. 'Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases', *Clin Cancer Res*, 10: 6897-904.

Arpino, G., V. J. Bardou, G. M. Clark, and R. M. Elledge. 2004. 'Infiltrating lobular carcinoma of the breast: tumor characteristics and clinical outcome', *Breast Cancer Res*, 6: R149-56.

Banys, M., I. Gruber, N. Krawczyk, S. Becker, R. Kurth, D. Wallwiener, J. Jakubowska, J. Hoffmann, R. Rothmund, A. Staebler, and T. Fehm. 2012. 'Hematogenous and lymphatic tumor cell dissemination may be detected in patients diagnosed with ductal carcinoma in situ of the breast', *Breast Cancer Res Treat*, 131: 801-8.

Bauer, K. D., J. de la Torre-Bueno, I. J. Diel, D. Hawes, W. J. Decker, C. Priddy, B. Bossy, S. Ludmann, K. Yamamoto, A. S. Masih, F. P. Espinoza, and D. S. Harrington. 2000. 'Reliable and sensitive analysis of occult bone marrow metastases using automated cellular imaging', *Clin Cancer Res*, 6: 3552-9.

Behjati, S., M. Huch, R. van Boxtel, W. Karthaus, D. C. Wedge, A. U. Tamuri, I. Martincorena, M. Petljak, L. B. Alexandrov, G. Gundem, P. S. Tarpey, S. Roerink, J. Blokker, M. Maddison, L. Mudie, B. Robinson, S. Nik-Zainal, P. Campbell, N. Goldman, M. van de Wetering, E. Cuppen, H. Clevers, and M. R. Stratton. 2014. 'Genome sequencing of normal cells reveals developmental lineages and mutational processes', *Nature*, 513: 422-25.

Bidard, F. C., C. Proudhon, and J. Y. Pierga. 2016. 'Circulating tumor cells in breast cancer', *Mol Oncol*, 10: 418-30.

Biezuner, T., A. Spiro, O. Raz, S. Amir, L. Milo, R. Adar, N. Chapal-Ilani, V. Berman, Y. Fried, E. Ainbinder, G. Cohen, H. M. Barr, R. Halaban, and E. Shapiro. 2016. 'A generic, cost-effective, and scalable cell lineage analysis platform', *Genome Res*, 26: 1588-99.

Bill, R., and G. Christofori. 2015. 'The relevance of EMT in breast cancer metastasis: Correlation or causality?', *FEBS Lett*, 589: 1577-87.

Blomberg, O. S., L. Spagnuolo, and K. E. de Visser. 2018. 'Immune regulation of metastasis: mechanistic insights and therapeutic opportunities', *Dis Model Mech*, 11.

Brabletz, T., D. Lyden, P. S. Steeg, and Z. Werb. 2013. 'Roadblocks to translational advances on metastasis research', *Nat Med*, 19: 1104-9.

Braun, S., F. D. Vogl, B. Naume, W. Janni, M. P. Osborne, R. C. Coombes, G. Schlimok, I. J. Diel, B. Gerber, G. Gebauer, J. Y. Pierga, C. Marth, D. Oruzio, G. Wiedswang, E. F.

Solomayer, G. Kundt, B. Strobl, T. Fehm, G. Y. Wong, J. Bliss, A. Vincent-Salomon, and K. Pantel. 2005. 'A pooled analysis of bone marrow micrometastasis in breast cancer', *N Engl J Med*, 353: 793-802.

Bray, F., J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal. 2018. 'Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries', *CA Cancer J Clin*, 68: 394-424.

Brenton, J. D., L. A. Carey, A. A. Ahmed, and C. Caldas. 2005. 'Molecular classification and molecular forecasting of breast cancer: ready for clinical application?', *J Clin Oncol*, 23: 7350-60.

Carlson, K. D., P. H. Sudmant, M. O. Press, E. E. Eichler, J. Shendure, and C. Queitsch. 2015. 'MIPSTR: a method for multiplex genotyping of germline and somatic STR variation across many individuals', *Genome Res*, 25: 750-61.

Chambers, A. F., A. C. Groom, and I. C. MacDonald. 2002. 'Dissemination and growth of cancer cells in metastatic sites', *Nat Rev Cancer*, 2: 563-72.

Chatterjee, M., and K. L. van Golen. 2011. 'Farnesyl transferase inhibitor treatment of breast cancer cells leads to altered RhoA and RhoC GTPase activity and induces a dormant phenotype', *Int J Cancer*, 129: 61-9.

Chu, J. E., and A. L. Allan. 2012. 'The Role of Cancer Stem Cells in the Organ Tropism of Breast Cancer Metastasis: A Mechanistic Balance between the "Seed" and the "Soil"?', *Int J Breast Cancer*, 2012: 209748.

Chu, J., S. Sadeghi, A. Raymond, S. D. Jackman, K. M. Nip, R. Mar, H. Mohamadi, Y. S. Butterfield, A. G. Robertson, and I. Birol. 2014. 'BioBloom tools: fast, accurate and memory-efficient host species sequence screening using bloom filters', *Bioinformatics*, 30: 3402-4.

Collins, V. P., R. K. Loeffler, and H. Tivey. 1956. 'Observations on growth rates of human tumors', *Am J Roentgenol Radium Ther Nucl Med*, 76: 988-1000.

Cristofanilli, M., D. F. Hayes, G. T. Budd, M. J. Ellis, A. Stopeck, J. M. Reuben, G. V. Doyle, J. Matera, W. J. Allard, M. C. Miller, H. A. Fritsche, G. N. Hortobagyi, and L. W. Terstappen. 2005. 'Circulating tumor cells: a novel prognostic factor for newly diagnosed metastatic breast cancer', *J Clin Oncol*, 23: 1420-30.

Cserni, G., E. Chmielik, B. Cserni, and T. Tot. 2018. 'The new TNM-based staging of breast cancer', *Virchows Arch*, 472: 697-703.

Curtis, C., S. P. Shah, S. F. Chin, G. Turashvili, O. M. Rueda, M. J. Dunning, D. Speed, A. G. Lynch, S. Samarajiwa, Y. Yuan, S. Graf, G. Ha, G. Haffari, A. Bashashati, R. Russell, S. McKinney, Metabric Group, A. Langerod, A. Green, E. Provenzano, G. Wishart, S. Pinder, P. Watson, F. Markowetz, L. Murphy, I. Ellis, A. Purushotham, A. L. Borresen-Dale, J. D. Brenton, S. Tavare, C. Caldas, and S. Aparicio. 2012. 'The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups', *Nature*, 486: 346-52.

Czyz, Z. T., M. Hoffmann, G. Schlimok, B. Polzer, and C. A. Klein. 2014. 'Reliable single cell array CGH for clinical samples', *PLoS One*, 9: e85907.

Czyz, Z. T., and C. A. Klein. 2015. 'Deterministic Whole-Genome Amplification of Single Cells', *Methods Mol Biol*, 1347: 69-86.

Domschke, C., I. J. Diel, S. Englert, S. Kalteisen, L. Mayer, J. Rom, J. Heil, C. Sohn, and F. Schuetz. 2013. 'Prognostic value of disseminated tumor cells in the bone marrow of patients with operable primary breast cancer: a long-term follow-up study', *Ann Surg Oncol*, 20: 1865-71.

Eckhardt, B. L., P. A. Francis, B. S. Parker, and R. L. Anderson. 2012. 'Strategies for the discovery and development of therapies for metastatic breast cancer', *Nat Rev Drug Discov*, 11: 479-97.

Endo, H., and M. Inoue. 2019. 'Dormancy in cancer', *Cancer Sci*, 110: 474-80.

Engel, J., R. Eckel, J. Kerr, M. Schmidt, G. Furstenberger, R. Richter, H. Sauer, H. J. Senn, and D. Holzel. 2003. 'The process of metastasisation for breast cancer', *Eur J Cancer*, 39: 1794-806.

Evrony, G. D., E. Lee, B. K. Mehta, Y. Benjamini, R. M. Johnson, X. Cai, L. Yang, P. Haseley, H. S. Lehmann, P. J. Park, and C. A. Walsh. 2015. 'Cell lineage analysis in human brain using endogenous retroelements', *Neuron*, 85: 49-59.

Eyles, J., A. L. Puaux, X. Wang, B. Toh, C. Prakash, M. Hong, T. G. Tan, L. Zheng, L. C. Ong, Y. Jin, M. Kato, A. Prevost-Blondel, P. Chow, H. Yang, and J. P. Abastado. 2010. 'Tumor cells disseminate early, but immunosurveillance limits metastatic outgrowth, in a mouse model of melanoma', *J Clin Invest*, 120: 2030-9.

Fallahpour, S., T. Navaneelan, P. De, and A. Borgo. 2017. 'Breast cancer survival by molecular subtype: a population-based analysis of cancer registry data', *CMAJ Open*, 5: E734-E39.

Fehm, T., S. Braun, V. Muller, W. Janni, G. Gebauer, C. Marth, C. Schindlbeck, D. Wallwiener, E. Borgen, B. Naume, K. Pantel, and E. Solomayer. 2006. 'A concept for the standardized detection of disseminated tumor cells in bone marrow from patients with primary breast cancer and its clinical implementation', *Cancer*, 107: 885-92.

Foulkes, W. D., I. E. Smith, and J. S. Reis-Filho. 2010. 'Triple-negative breast cancer', *N Engl J Med*, 363: 1938-48.

Frieda, K. L., J. M. Linton, S. Hormoz, J. Choi, K. K. Chow, Z. S. Singer, M. W. Budde, M. B. Elowitz, and L. Cai. 2017. 'Synthetic recording and in situ readout of lineage information in single cells', *Nature*, 541: 107-11.

Frumkin, D., A. Wasserstrom, S. Itzkovitz, T. Stern, A. Harmelin, R. Eilam, G. Rechavi, and E. Shapiro. 2008. 'Cell lineage analysis of a mouse tumor', *Cancer Res*, 68: 5924-31.

Frumkin, D., A. Wasserstrom, S. Kaplan, U. Feige, and E. Shapiro. 2005. 'Genomic variability within an organism exposes its cell lineage tree', *PLoS Comput Biol*, 1: e50.

Gay, L. J., and B. Felding-Habermann. 2011. 'Contribution of platelets to tumour metastasis', *Nat Rev Cancer*, 11: 123-34.

Ghajar, C. M., H. Peinado, H. Mori, I. R. Matei, K. J. Evason, H. Brazier, D. Almeida, A. Koller, K. A. Hajjar, D. Y. Stainier, E. I. Chen, D. Lyden, and M. J. Bissell. 2013. 'The perivascular niche regulates breast tumour dormancy', *Nat Cell Biol*, 15: 807-17.

Giannakeas, V., V. Sopik, and S. A. Narod. 2018. 'A comparison of two models for breast cancer mortality for women with ductal carcinoma in situ: an SEER-based analysis', *Breast Cancer Res Treat*, 169: 587-94.

Giuliano, A. E., S. B. Edge, and G. N. Hortobagyi. 2018. 'Eighth Edition of the AJCC Cancer Staging Manual: Breast Cancer', *Ann Surg Oncol*, 25: 1783-85.

Gomis, R. R., and S. Gawrzak. 2017. 'Tumor cell dormancy', *Mol Oncol*, 11: 62-78.

Goodenough, D. A., and N. B. Gilula. 1974. 'The splitting of hepatocyte gap junctions and zonulae occludentes with hypertonic disaccharides', *J Cell Biol*, 61: 575-90.

Grujovic, Ana. 2019. 'Regulation of cellular dormancy in disseminated breast cancer cells', Doctoral Thesis, University of Regensburg.

Guo, T., E. Stankiewicz, X. Mao, and Y. J. Lu. 2019. 'The Isolation and Analysis of Circulating Tumor Cells', *Methods Mol Biol*, 2054: 115-28.

Guzvic, M., B. Braun, R. Ganzer, M. Burger, M. Nerlich, S. Winkler, M. Werner-Klein, Z. T. Czyz, B. Polzer, and C. A. Klein. 2014. 'Combined genome and transcriptome analysis of single disseminated cancer cells from bone marrow of prostate cancer patients reveals unexpected transcriptomes', *Cancer Res*, 74: 7383-94.

Haffty, B. G., Q. Yang, M. Reiss, T. Kearney, S. A. Higgins, J. Weidhaas, L. Harris, W. Hait, and D. Toppmeyer. 2006. 'Locoregional relapse and distant metastasis in conservatively managed triple negative early-stage breast cancer', *J Clin Oncol*, 24: 5652-7.

Hammond, M. E., D. F. Hayes, M. Dowsett, D. C. Allred, K. L. Hagerty, S. Badve, P. L. Fitzgibbons, G. Francis, N. S. Goldstein, M. Hayes, D. G. Hicks, S. Lester, R. Love, P. B. Mangu, L. McShane, K. Miller, C. K. Osborne, S. Paik, J. Perlmutter, A. Rhodes, H. Sasano, J. N. Schwartz, F. C. Sweep, S. Taube, E. E. Torlakovic, P. Valenstein, G. Viale, D. Visscher, T. Wheeler, R. B. Williams, J. L. Wittliff, and A. C. Wolff. 2010. '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*, 28: 2784-95.

Harbeck, N., and M. Gnant. 2017. 'Breast cancer', *Lancet*, 389: 1134-50.

Harper, K. L., M. S. Sosa, D. Entenberg, H. Hosseini, J. F. Cheung, R. Nobre, A. Avivar-Valderas, C. Nagi, N. Girnius, R. J. Davis, E. F. Farias, J. Condeelis, C. A. Klein, and J. A. Aguirre-Ghiso. 2016. 'Mechanism of early dissemination and metastasis in Her2(+) mammary cancer', *Nature*, 540: 588-92.

Hartkopf, A. D., F. A. Taran, M. Wallwiener, M. Hahn, S. Becker, E. F. Solomayer, S. Y. Brucker, T. N. Fehm, and D. Wallwiener. 2014. 'Prognostic relevance of disseminated tumour cells from the bone marrow of early stage breast cancer patients - results from a large single-centre analysis', *Eur J Cancer*, 50: 2550-9.

Hartkopf AD, Brucker SY, Taran F-A, Harbeck N, von Au A, Naume B, Pierga J-Y, Hoffmann O, Beckmann MW, Rydén L, Fehm T, Aft R, Montserrat S, Walter V, Rack B, Schuetz F, Borgen E, Ta M-H, Bittner A-K, Fasching P, Fernö M, Krawczyk N, Weilbaecher K, Margelí M, Hahn M, Jueckstock J, Domschke C, Bidard F-C, Kasimir-Bauer S, Schoenfisch B, Kurt AG, Wallwiener M, Gebauer G, Wallwiener D, Janni W, Pantel K. 2018 'Abstract GS5-07: International pooled analysis of the prognostic impact of disseminated tumor cells from the bone marrow in early breast cancer: Results from the PADDY study', *Proceedings of the 2018 San Antonio Breast Cancer Symposium*, 79.

Havas, K. M., V. Milchevskaya, K. Radic, A. Alladin, E. Kafkia, M. Garcia, J. Stolte, B. Klaus, N. Rotmensz, T. J. Gibson, B. Burwinkel, A. Schneeweiss, G. Pruneri, K. R. Patil, R. Sotillo, and M. Jechlinger. 2017. 'Metabolic shifts in residual breast cancer drive tumor recurrence', *J Clin Invest*, 127: 2091-105.

Hecht, Eva-Maria. 2016. 'Optimizing Laser Pressure Catapulting Of Single Cells While Preserving The DNA Integrity', Master Thesis, University of Regensburg

Heitzer, E., M. Auer, C. Gasch, M. Pichler, P. Ulz, E. M. Hoffmann, S. Lax, J. Waldispuehl-Geigl, O. Mauermann, C. Lackner, G. Hofler, F. Eisner, H. Sill, H. Samonigg, K. Pantel, S. Riethdorf, T. Bauernhofer, J. B. Geigl, and M. R. Speicher. 2013. 'Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing', *Cancer Res*, 73: 2965-75.

Hogan, M. P., D. A. Goldman, B. Dashevsky, C. C. Riedl, M. Gonen, J. R. Osborne, M. Jochelson, C. Hudis, M. Morrow, and G. A. Ulaner. 2015. 'Comparison of 18F-FDG

PET/CT for Systemic Staging of Newly Diagnosed Invasive Lobular Carcinoma Versus Invasive Ductal Carcinoma', *J Nucl Med*, 56: 1674-80.

Hognas, G., K. Kivinummi, H. M. L. Kallio, R. Hieta, P. Ruusuvuori, A. Koskenalho, J. Kesseli, T. L. J. Tammela, J. Riikonen, J. Ilvesaro, S. Kares, P. P. Hirvikoski, M. Laurila, T. Mirtti, M. Nykter, P. M. Kujala, T. Visakorpi, T. Tolonen, and G. S. Bova. 2018. 'Feasibility of Prostate PAXgene Fixation for Molecular Research and Diagnostic Surgical Pathology: Comparison of Matched Fresh Frozen, FFPE, and PFPE Tissues', *Am J Surg Pathol*, 42: 103-15.

Hosseini, H., M. M. Obradovic, M. Hoffmann, K. L. Harper, M. S. Sosa, M. Werner-Klein, L. K. Nanduri, C. Werno, C. Ehrl, M. Maneck, N. Patwary, G. Haunschild, M. Guzvic, C. Reimelt, M. Grauvogl, N. Eichner, F. Weber, A. D. Hartkopf, F. A. Taran, S. Y. Brucker, T. Fehm, B. Rack, S. Buchholz, R. Spang, G. Meister, J. A. Aguirre-Ghiso, and C. A. Klein. 2016. 'Early dissemination seeds metastasis in breast cancer', *Nature*.

Howlader, N., S. F. Altekruse, C. I. Li, V. W. Chen, C. A. Clarke, L. A. Ries, and K. A. Cronin. 2014. 'US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status', *J Natl Cancer Inst*, 106.

Husemann, Y., J. B. Geigl, F. Schubert, P. Musiani, M. Meyer, E. Burghart, G. Forni, R. Eils, T. Fehm, G. Riethmuller, and C. A. Klein. 2008. 'Systemic spread is an early step in breast cancer', *Cancer Cell*, 13: 58-68.

Inman, J. L., C. Robertson, J. D. Mott, and M. J. Bissell. 2015. 'Mammary gland development: cell fate specification, stem cells and the microenvironment', *Development*, 142: 1028-42.

Irlbeck, Christoph. 2019. 'Molecular characterization of disseminated cancer cells isolated from patients with luminal B type breast cancer', PhD Thesis, University of Regensburg.

Ishaque, N., M. L. Abba, C. Hauser, N. Patil, N. Paramasivam, D. Huebschmann, J. H. Leupold, G. P. Balasubramanian, K. Kleinheinz, U. H. Toprak, B. Hutter, A. Benner, A. Shavinskaya, C. Zhou, Z. Gu, J. Kerssemakers, A. Marx, M. Moniuszko, M. Kozlowski, J. Reszec, J. Niklinski, J. Eils, M. Schlesner, R. Eils, B. Brors, and H. Allgayer. 2018. 'Whole genome sequencing puts forward hypotheses on metastasis evolution and therapy in colorectal cancer', *Nat Commun*, 9: 4782.

Jaffer, S., and I. J. Bleiweiss. 2002. 'Histologic classification of ductal carcinoma in situ', *Microsc Res Tech*, 59: 92-101.

Janni, W., S. Gastroph, F. Hepp, C. Kentenich, D. Rjosk, C. Schindlbeck, T. Dimpfl, H. Sommer, and S. Braun. 2000. 'Prognostic significance of an increased number of micrometastatic tumor cells in the bone marrow of patients with first recurrence of breast carcinoma', *Cancer*, 88: 2252-9.

Janni, W., F. D. Vogl, G. Wiedswang, M. Synnestvedt, T. Fehm, J. Juckstock, E. Borgen, B. Rack, S. Braun, H. Sommer, E. Solomayer, K. Pantel, J. Nesland, K. Friese, and B. Naume. 2011. 'Persistence of disseminated tumor cells in the bone marrow of breast cancer patients predicts increased risk for relapse--a European pooled analysis', *Clin Cancer Res*, 17: 2967-76.

Javed, A., and A. Lteif. 2013. 'Development of the human breast', *Semin Plast Surg*, 27: 5-12.

Klein, C. A. 2000. 'The biology and analysis of single disseminated tumour cells', *Trends Cell Biol*, 10: 489-93.

———. 2008. 'Cancer. The metastasis cascade', *Science*, 321: 1785-7.

———. 2009. 'Parallel progression of primary tumours and metastases', *Nat Rev Cancer*, 9: 302-12.
———. 2013. 'Selection and adaptation during metastatic cancer progression', *Nature*, 501: 365-72.
Klein, C. A., T. J. Blankenstein, O. Schmidt-Kittler, M. Petronio, B. Polzer, N. H. Stoecklein, and G. Riethmuller. 2002. 'Genetic heterogeneity of single disseminated tumour cells in minimal residual cancer', *Lancet*, 360: 683-9.
Klein, C. A., and D. Holzel. 2006. 'Systemic cancer progression and tumor dormancy: mathematical models meet single cell genomics', *Cell Cycle*, 5: 1788-98.
Klein, C. A., O. Schmidt-Kittler, J. A. Schardt, K. Pantel, M. R. Speicher, and G. Riethmuller. 1999. 'Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells', *Proc Natl Acad Sci U S A*, 96: 4494-9.
Klein, C. A., S. Seidl, K. Petat-Dutter, S. Offner, J. B. Geigl, O. Schmidt-Kittler, N. Wendler, B. Passlick, R. M. Huber, G. Schlimok, P. A. Baeuerle, and G. Riethmuller. 2002. 'Combined transcriptome and genome analysis of single micrometastatic cells', *Nat Biotechnol*, 20: 387-92.
Kretzschmar, K., and F. M. Watt. 2012. 'Lineage tracing', *Cell*, 148: 33-45.
Kroigard, A. B., M. J. Larsen, C. Brasch-Andersen, A. V. Laenkholm, A. S. Knoop, J. D. Jensen, M. Bak, J. Mollenhauer, M. Thomassen, and T. A. Kruse. 2017. 'Genomic Analyses of Breast Cancer Progression Reveal Distinct Routes of Metastasis Emergence', *Sci Rep*, 7: 43813.
Kuipers, J., K. Jahn, B. J. Raphael, and N. Beerenwinkel. 2017. 'Single-cell sequencing data reveal widespread recurrence and loss of mutational hits in the life histories of tumors', *Genome Res*, 27: 1885-94.
Lambert, A. W., D. R. Pattabiraman, and R. A. Weinberg. 2017. 'Emerging Biological Principles of Metastasis', *Cell*, 168: 670-91.
Lemoine, F., J. B. Domelevo Entfellner, E. Wilkinson, D. Correia, M. Davila Felipe, T. De Oliveira, and O. Gascuel. 2018. 'Renewing Felsenstein's phylogenetic bootstrap in the era of big data', *Nature*, 556: 452-56.
Leung, M. L., A. Davis, R. Gao, A. Casasent, Y. Wang, E. Sei, E. Vilar, D. Maru, S. Kopetz, and N. E. Navin. 2017. 'Single-cell DNA sequencing reveals a late-dissemination model in metastatic colorectal cancer', *Genome Res*, 27: 1287-99.
Li, C. I., B. O. Anderson, J. R. Daling, and R. E. Moe. 2003. 'Trends in incidence rates of invasive lobular and ductal breast carcinoma', *JAMA*, 289: 1421-4.
Li, C. I., D. J. Uribe, and J. R. Daling. 2005. 'Clinical characteristics of different histologic types of breast cancer', *Br J Cancer*, 93: 1046-52.
Linde, N., G. Fluegen, and J. A. Aguirre-Ghiso. 2016. 'The Relationship Between Dormant Cancer Cells and Their Microenvironment', *Adv Cancer Res*, 132: 45-71.
Lodato, M. A., M. B. Woodworth, S. Lee, G. D. Evrony, B. K. Mehta, A. Karger, S. Lee, T. W. Chittenden, A. M. D'Gama, X. Cai, L. J. Luquette, E. Lee, P. J. Park, and C. A. Walsh. 2015. 'Somatic mutation in single human neurons tracks developmental and transcriptional history', *Science*, 350: 94-98.
Luker, K. E., and G. D. Luker. 2006. 'Functions of CXCL12 and CXCR4 in breast cancer', *Cancer Lett*, 238: 30-41.
Lyons, W. R. 1958. 'Hormonal synergism in mammary growth', *Proc R Soc Lond B Biol Sci*, 149: 303-25.

Magbanua, M. J., E. V. Sosa, R. Roy, L. E. Eisenbud, J. H. Scott, A. Olshen, D. Pinkel, H. S. Rugo, and J. W. Park. 2013. 'Genomic profiling of isolated circulating tumor cells from metastatic breast cancer patients', *Cancer Res*, 73: 30-40.

Magbanua, M. J., E. V. Sosa, J. H. Scott, J. Simko, C. Collins, D. Pinkel, C. J. Ryan, and J. W. Park. 2012. 'Isolation and genomic analysis of circulating tumor cells from castration resistant metastatic prostate cancer', *BMC Cancer*, 12: 78.

Malhotra, G. K., X. Zhao, H. Band, and V. Band. 2010. 'Histological, molecular and functional subtypes of breast cancers', *Cancer Biol Ther*, 10: 955-60.

Malladi, S., D. G. Macalinao, X. Jin, L. He, H. Basnet, Y. Zou, E. de Stanchina, and J. Massague. 2016. 'Metastatic Latency and Immune Evasion through Autocrine Inhibition of WNT', *Cell*, 165: 45-60.

Martelotto, L. G., T. Baslan, J. Kendall, F. C. Geyer, K. A. Burke, L. Spraggon, S. Piscuoglio, K. Chadalavada, G. Nanjangud, C. K. Ng, P. Moody, S. D'Italia, L. Rodgers, H. Cox, A. da Cruz Paula, A. Stepansky, M. Schizas, H. Y. Wen, T. A. King, L. Norton, B. Weigelt, J. B. Hicks, and J. S. Reis-Filho. 2017. 'Whole-genome single-cell copy number profiling from formalin-fixed paraffin-embedded samples', *Nat Med*, 23: 376-85.

McDonough, S. J., A. Bhagwate, Z. Sun, C. Wang, M. Zschunke, J. A. Gorman, K. J. Kopp, and J. M. Cunningham. 2019. 'Use of FFPE-derived DNA in next generation sequencing: DNA extraction methods', *PLoS One*, 14: e0211400.

McKenna, A., G. M. Findlay, J. A. Gagnon, M. S. Horwitz, A. F. Schier, and J. Shendure. 2016. 'Whole-organism lineage tracing by combinatorial and cumulative genome editing', *Science*, 353: aaf7907.

McMaster Pathophysiology Review. 2012-2018. 'Breast cancer'. http://www.pathophys.org/breast-cancer/.

Mitra, A., L. Mishra, and S. Li. 2013. 'Technologies for deriving primary tumor cells for use in personalized cancer therapy', *Trends Biotechnol*, 31: 347-54.

Muller, P., D. Weckermann, G. Riethmuller, and G. Schlimok. 1996. 'Detection of genetic alterations in micrometastatic cells in bone marrow of cancer patients by fluorescence in situ hybridization', *Cancer Genet Cytogenet*, 88: 8-16.

Muzumdar, M. D., K. J. Dorans, K. M. Chung, R. Robbins, T. Tammela, V. Gocheva, C. M. Li, and T. Jacks. 2016. 'Clonal dynamics following p53 loss of heterozygosity in Kras-driven cancers', *Nat Commun*, 7: 12685.

Narod, S. A., J. Iqbal, V. Giannakeas, V. Sopik, and P. Sun. 2015. 'Breast Cancer Mortality After a Diagnosis of Ductal Carcinoma In Situ', *JAMA Oncol*, 1: 888-96.

Navin, N., J. Kendall, J. Troge, P. Andrews, L. Rodgers, J. McIndoo, K. Cook, A. Stepansky, D. Levy, D. Esposito, L. Muthuswamy, A. Krasnitz, W. R. McCombie, J. Hicks, and M. Wigler. 2011. 'Tumour evolution inferred by single-cell sequencing', *Nature*, 472: 90-4.

Nguyen, D. X., P. D. Bos, and J. Massague. 2009. 'Metastasis: from dissemination to organ-specific colonization', *Nat Rev Cancer*, 9: 274-84.

Nilsson, M., H. Malmgren, M. Samiotaki, M. Kwiatkowski, B. P. Chowdhary, and U. Landegren. 1994. 'Padlock probes: circularizing oligonucleotides for localized DNA detection', *Science*, 265: 2085-8.

O'Shaughnessy, J. 2005. 'Extending survival with chemotherapy in metastatic breast cancer', *Oncologist*, 10 Suppl 3: 20-9.

O'Sullivan, B., J. Brierley, D. Byrd, F. Bosman, S. Kehoe, C. Kossary, M. Pineros, E. Van Eycken, H. K. Weir, and M. Gospodarowicz. 2017. 'The TNM classification of malignant

tumours-towards common understanding and reasonable expectations', *Lancet Oncol*, 18: 849-51.

Oh, H., C. E. Boeke, R. M. Tamimi, S. A. Smith-Warner, M. Wang, W. C. Willett, and A. H. Eliassen. 2015. 'The interaction between early-life body size and physical activity on risk of breast cancer', *Int J Cancer*, 137: 571-81.

Paget, S. 1989. 'The distribution of secondary growths in cancer of the breast. 1889', *Cancer Metastasis Rev*, 8: 98-101.

Pan, H., R. Gray, J. Braybrooke, C. Davies, C. Taylor, P. McGale, R. Peto, K. I. Pritchard, J. Bergh, M. Dowsett, D. F. Hayes, and Ebctcg. 2017. '20-Year Risks of Breast-Cancer Recurrence after Stopping Endocrine Therapy at 5 Years', *N Engl J Med*, 377: 1836-46.

Pantel, K., and M. von Knebel Doeberitz. 2000. 'Detection and clinical relevance of micrometastatic cancer cells', *Curr Opin Oncol*, 12: 95-101.

Perou, C. M., T. Sorlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. Lonning, A. L. Borresen-Dale, P. O. Brown, and D. Botstein. 2000. 'Molecular portraits of human breast tumours', *Nature*, 406: 747-52.

Polyak, K., and O. Metzger Filho. 2012. 'SnapShot: breast cancer', *Cancer Cell*, 22: 562-62 e1.

Polzer, B., G. Medoro, S. Pasch, F. Fontana, L. Zorzino, A. Pestka, U. Andergassen, F. Meier-Stiegen, Z. T. Czyz, B. Alberter, S. Treitschke, T. Schamberger, M. Sergio, G. Bregola, A. Doffini, S. Gianni, A. Calanca, G. Signorini, C. Bolognesi, A. Hartmann, P. A. Fasching, M. T. Sandri, B. Rack, T. Fehm, G. Giorgini, N. Manaresi, and C. A. Klein. 2014. 'Molecular profiling of single circulating tumor cells with diagnostic intention', *EMBO Mol Med*, 6: 1371-86.

Prat, A., M. C. Cheang, M. Martin, J. S. Parker, E. Carrasco, R. Caballero, S. Tyldesley, K. Gelmon, P. S. Bernard, T. O. Nielsen, and C. M. Perou. 2013. 'Prognostic significance of progesterone receptor-positive tumor cells within immunohistochemically defined luminal A breast cancer', *J Clin Oncol*, 31: 203-9.

Prat, A., E. Pineda, B. Adamo, P. Galvan, A. Fernandez, L. Gaba, M. Diez, M. Viladot, A. Arance, and M. Munoz. 2015. 'Clinical implications of the intrinsic molecular subtypes of breast cancer', *Breast*, 24 Suppl 2: S26-35.

Price, M. N., P. S. Dehal, and A. P. Arkin. 2010. 'FastTree 2--approximately maximum-likelihood trees for large alignments', *PLoS One*, 5: e9490.

Psaila, B., R. N. Kaplan, E. R. Port, and D. Lyden. 2006. 'Priming the 'soil' for breast cancer metastasis: the pre-metastatic niche', *Breast Dis*, 26: 65-74.

Psaila, B., and D. Lyden. 2009. 'The metastatic niche: adapting the foreign soil', *Nat Rev Cancer*, 9: 285-93.

Rabbani, S. A., and A. P. Mazar. 2007. 'Evaluating distant metastases in breast cancer: from biology to outcomes', *Cancer Metastasis Rev*, 26: 663-74.

Rakha, E. A., J. S. Reis-Filho, F. Baehner, D. J. Dabbs, T. Decker, V. Eusebi, S. B. Fox, S. Ichihara, J. Jacquemier, S. R. Lakhani, J. Palacios, A. L. Richardson, S. J. Schnitt, F. C. Schmitt, P. H. Tan, G. M. Tse, S. Badve, and I. O. Ellis. 2010. 'Breast cancer prognostic classification in the molecular era: the role of histological grade', *Breast Cancer Res*, 12: 207.

Raz, O., T. Biezuner, A. Spiro, S. Amir, L. Milo, A. Titelman, A. Onn, N. Chapal-Ilani, L. Tao, T. Marx, U. Feige, and E. Shapiro. 2019. 'Short tandem repeat stutter model inferred from direct measurement of in vitro stutter noise', *Nucleic Acids Res*, 47: 2436-45.

Reizel, Y., N. Chapal-Ilani, R. Adar, S. Itzkovitz, J. Elbaz, Y. E. Maruvka, E. Segev, L. I. Shlush, N. Dekel, and E. Shapiro. 2011. 'Colon stem cell and crypt dynamics exposed by cell lineage reconstruction', *PLoS Genet*, 7: e1002192.

Reizel, Y., S. Itzkovitz, R. Adar, J. Elbaz, A. Jinich, N. Chapal-Ilani, Y. E. Maruvka, N. Nevo, Z. Marx, I. Horovitz, A. Wasserstrom, A. Mayo, I. Shur, D. Benayahu, K. Skorecki, E. Segal, N. Dekel, and E. Shapiro. 2012. 'Cell lineage analysis of the mammalian female germline', *PLoS Genet*, 8: e1002477.

Riethdorf, S., H. Wikman, and K. Pantel. 2008. 'Review: Biological relevance of disseminated tumor cells in cancer patients', *Int J Cancer*, 123: 1991-2006.

Robson, M., S. A. Im, E. Senkus, B. Xu, S. M. Domchek, N. Masuda, S. Delaloge, W. Li, N. Tung, A. Armstrong, W. Wu, C. Goessl, S. Runswick, and P. Conte. 2017. 'Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation', *N Engl J Med*, 377: 523-33.

Romond, E. H., E. A. Perez, J. Bryant, V. J. Suman, C. E. Geyer, Jr., N. E. Davidson, E. Tan-Chiu, S. Martino, S. Paik, P. A. Kaufman, S. M. Swain, T. M. Pisansky, L. Fehrenbacher, L. A. Kutteh, V. G. Vogel, D. W. Visscher, G. Yothers, R. B. Jenkins, A. M. Brown, S. R. Dakhil, E. P. Mamounas, W. L. Lingle, P. M. Klein, J. N. Ingle, and N. Wolmark. 2005. 'Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer', *N Engl J Med*, 353: 1673-84.

Rosen, J. M. 2012. 'On hormone action in the mammary gland', *Cold Spring Harb Perspect Biol*, 4.

Rosen, R. D., and A. Sapra. 2020. 'TNM Classification.' in, *StatPearls* (Treasure Island (FL)).

Salipante, S. J., A. Kas, E. McMonagle, and M. S. Horwitz. 2010. 'Phylogenetic analysis of developmental and postnatal mouse cell lineages', *Evol Dev*, 12: 84-94.

Sanborn, J. Z., J. Chung, E. Purdom, N. J. Wang, H. Kakavand, J. S. Wilmott, T. Butler, J. F. Thompson, G. J. Mann, L. E. Haydu, R. P. Saw, K. J. Busam, R. S. Lo, E. A. Collisson, J. S. Hur, P. T. Spellman, J. E. Cleaver, J. W. Gray, N. Huh, R. Murali, R. A. Scolyer, B. C. Bastian, and R. J. Cho. 2015. 'Phylogenetic analyses of melanoma reveal complex patterns of metastatic dissemination', *Proc Natl Acad Sci U S A*, 112: 10995-1000.

Sanger, N., K. E. Effenberger, S. Riethdorf, V. Van Haasteren, J. Gauwerky, I. Wiegratz, K. Strebhardt, M. Kaufmann, and K. Pantel. 2011. 'Disseminated tumor cells in the bone marrow of patients with ductal carcinoma in situ', *Int J Cancer*, 129: 2522-6.

Schardt, J. A., M. Meyer, C. H. Hartmann, F. Schubert, O. Schmidt-Kittler, C. Fuhrmann, B. Polzer, M. Petronio, R. Eils, and C. A. Klein. 2005. 'Genomic analysis of single cytokeratin-positive cells from bone marrow reveals early mutational events in breast cancer', *Cancer Cell*, 8: 227-39.

Schlimok, G., I. Funke, B. Holzmann, G. Gottlinger, G. Schmidt, H. Hauser, S. Swierkot, H. H. Warnecke, B. Schneider, H. Koprowski, and G. Riethmuller. 1987. 'Micrometastatic cancer cells in bone marrow: in vitro detection with anti-cytokeratin and in vivo labeling with anti-17-1A monoclonal antibodies', *Proc Natl Acad Sci U S A*, 84: 8672-6.

Schmidt-Kittler, O., T. Ragg, A. Daskalakis, M. Granzow, A. Ahr, T. J. Blankenstein, M. Kaufmann, J. Diebold, H. Arnholdt, P. Muller, J. Bischoff, D. Harich, G. Schlimok, G. Riethmuller, R. Eils, and C. A. Klein. 2003. 'From latent disseminated cells to overt metastasis: genetic analysis of systemic breast cancer progression', *Proc Natl Acad Sci U S A*, 100: 7737-42.

Schwartz, A. M., D. E. Henson, D. Chen, and S. Rajamarthandan. 2014. 'Histologic grade remains a prognostic factor for breast cancer regardless of the number of positive lymph nodes and tumor size: a study of 161 708 cases of breast cancer from the SEER Program', *Arch Pathol Lab Med*, 138: 1048-52.

Seely, J. M., and T. Alhassan. 2018. 'Screening for breast cancer in 2018-what should we be doing today?', *Curr Oncol*, 25: S115-S24.

Senft, D., and Z. E. Ronai. 2016. 'Adaptive Stress Responses During Tumor Metastasis and Dormancy', *Trends Cancer*, 2: 429-42.

Shapiro, E. 2018. 'On the journey from nematode to human, scientists dive by the zebrafish cell lineage tree', *Genome Biol*, 19: 63.

Shariff-Marco, S., J. Yang, E. M. John, A. W. Kurian, I. Cheng, R. Leung, J. Koo, K. R. Monroe, B. E. Henderson, L. Bernstein, Y. Lu, M. L. Kwan, R. Sposto, C. L. Vigen, A. H. Wu, T. H. Keegan, and S. L. Gomez. 2015. 'Intersection of Race/Ethnicity and Socioeconomic Status in Mortality After Breast Cancer', *J Community Health*, 40: 1287-99.

Shlush, L. I., N. Chapal-Ilani, R. Adar, N. Pery, Y. Maruvka, A. Spiro, R. Shouval, J. M. Rowe, M. Tzukerman, D. Bercovich, S. Izraeli, G. Marcucci, C. D. Bloomfield, T. Zuckerman, K. Skorecki, and E. Shapiro. 2012. 'Cell lineage analysis of acute leukemia relapse uncovers the role of replication-rate heterogeneity and microsatellite instability', *Blood*, 120: 603-12.

Sieuwerts, A. M., J. Kraan, J. Bolt, P. van der Spoel, F. Elstrodt, M. Schutte, J. W. Martens, J. W. Gratama, S. Sleijfer, and J. A. Foekens. 2009. 'Anti-epithelial cell adhesion molecule antibodies and the detection of circulating normal-like breast tumor cells', *J Natl Cancer Inst*, 101: 61-6.

Slamon, D., W. Eiermann, N. Robert, T. Pienkowski, M. Martin, M. Press, J. Mackey, J. Glaspy, A. Chan, M. Pawlicki, T. Pinter, V. Valero, M. C. Liu, G. Sauter, G. von Minckwitz, F. Visco, V. Bee, M. Buyse, B. Bendahmane, I. Tabah-Fisch, M. A. Lindsay, A. Riva, J. Crown, and Group Breast Cancer International Research. 2011. 'Adjuvant trastuzumab in HER2-positive breast cancer', *N Engl J Med*, 365: 1273-83.

Sluka, P., L. O'Donnell, R. I. McLachlan, and P. G. Stanton. 2008. 'Application of laser-capture microdissection to analysis of gene expression in the testis', *Prog Histochem Cytochem*, 42: 173-201.

Sorlie, T., C. M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, T. Thorsen, H. Quist, J. C. Matese, P. O. Brown, D. Botstein, P. E. Lonning, and A. L. Borresen-Dale. 2001. 'Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications', *Proc Natl Acad Sci U S A*, 98: 10869-74.

Sosa, M. S., P. Bragado, and J. A. Aguirre-Ghiso. 2014. 'Mechanisms of disseminated cancer cell dormancy: an awakening field', *Nat Rev Cancer*, 14: 611-22.

Sotiriou, C., and L. Pusztai. 2009. 'Gene-expression signatures in breast cancer', *N Engl J Med*, 360: 790-800.

Spiro, A., and E. Shapiro. 2016. 'Accuracy of Answers to Cell Lineage Questions Depends on Single-Cell Genomics Data Quality and Quantity', *PLoS Comput Biol*, 12: e1004983.

Stoecklein, N. H., A. Erbersdobler, O. Schmidt-Kittler, J. Diebold, J. A. Schardt, J. R. Izbicki, and C. A. Klein. 2002. 'SCOMP is superior to degenerated oligonucleotide primed-polymerase chain reaction for global amplification of minute amounts of DNA from microdissected archival tissue samples', *Am J Pathol*, 161: 43-51.

Stoecklein, N. H., S. B. Hosch, M. Bezler, F. Stern, C. H. Hartmann, C. Vay, A. Siegmund, P. Scheunemann, P. Schurr, W. T. Knoefel, P. E. Verde, U. Reichelt, A. Erbersdobler, R. Grau, A. Ullrich, J. R. Izbicki, and C. A. Klein. 2008. 'Direct genetic analysis of single disseminated cancer cells for prediction of outcome and therapy selection in esophageal cancer', *Cancer Cell*, 13: 441-53.

Sturm, D., L. Marselli, F. Ehehalt, D. Richter, M. Distler, S. Kersting, R. Grutzmann, K. Bokvist, P. Froguel, R. Liechti, A. Jorns, P. Meda, G. B. Baretton, H. D. Saeger, A. M. Schulte, P. Marchetti, and M. Solimena. 2013. 'Improved protocol for laser microdissection of human pancreatic islets from surgical specimens', *J Vis Exp*.

Suhail, Y., M. P. Cain, K. Vanaja, P. A. Kurywchak, A. Levchenko, R. Kalluri, and Kshitiz. 2019. 'Systems Biology of Cancer Metastasis', *Cell Syst*, 9: 109-27.

Sulston, J. E., and H. R. Horvitz. 1977. 'Post-embryonic cell lineages of the nematode, Caenorhabditis elegans', *Dev Biol*, 56: 110-56.

Sulston, J. E., E. Schierenberg, J. G. White, and J. N. Thomson. 1983. 'The embryonic cell lineage of the nematode Caenorhabditis elegans', *Dev Biol*, 100: 64-119.

Svaren, J., S. Inagami, E. Lovegren, and R. Chalkley. 1987. 'DNA denatures upon drying after ethanol precipitation', *Nucleic Acids Res*, 15: 8739-54.

Tao, Liming, Ofir Raz, Zipora Marx, Tamir Biezuner, Shiran Amir, Lilach Milo, Rivka Adar, Amos Onn, Noa Chapal-Ilani, Veronika Berman, Ron Levy, Barak Oron, Ruth Halaban, and Ehud Shapiro. 2018. 'A biological-computational human cell lineage discovery platform based on duplex molecular inversion probes', *bioRxiv*: 191296.

Tjensvoll, K., O. Nordgard, M. Skjaeveland, S. Oltedal, E. A. M. Janssen, and B. Gilje. 2019. 'Detection of disseminated tumor cells in bone marrow predict late recurrences in operable breast cancer patients', *BMC Cancer*, 19: 1131.

Tjensvoll, K., S. Oltedal, R. K. Farmen, F. V. Shammas, R. Heikkila, J. T. Kvaloy, B. Gilje, R. Smaaland, and O. Nordgard. 2010. 'Disseminated tumor cells in bone marrow assessed by TWIST1, cytokeratin 19, and mammaglobin A mRNA predict clinical outcome in operable breast cancer patients', *Clin Breast Cancer*, 10: 378-84.

Torre, L. A., R. L. Siegel, E. M. Ward, and A. Jemal. 2016. 'Global Cancer Incidence and Mortality Rates and Trends--An Update', *Cancer Epidemiol Biomarkers Prev*, 25: 16-27.

Valastyan, S., and R. A. Weinberg. 2011. 'Tumor metastasis: molecular insights and evolving paradigms', *Cell*, 147: 275-92.

van de Moosdijk, A. A., N. Y. Fu, A. C. Rios, J. E. Visvader, and R. van Amerongen. 2017. 'Lineage Tracing of Mammary Stem and Progenitor Cells', *Methods Mol Biol*, 1501: 291-308.

van de Wouw, A. J., M. L. Janssen-Heijnen, J. W. Coebergh, and H. F. Hillen. 2002. 'Epidemiology of unknown primary tumours; incidence and population-based survival of 1285 patients in Southeast Netherlands, 1984-1992', *Eur J Cancer*, 38: 409-13.

van der Toom, E. E., J. E. Verdone, and K. J. Pienta. 2016. 'Disseminated tumor cells and dormancy in prostate cancer metastasis', *Curr Opin Biotechnol*, 40: 9-15.

Vandewoestyne, M., K. Goossens, C. Burvenich, A. Van Soom, L. Peelman, and D. Deforce. 2013. 'Laser capture microdissection: should an ultraviolet or infrared laser be used?', *Anal Biochem*, 439: 88-98.

Visvader, J. E., and J. Stingl. 2014. 'Mammary stem cells and the differentiation hierarchy: current status and perspectives', *Genes Dev*, 28: 1143-58.

Vogel, A., K. Lorenz, V. Horneffer, G. Huttmann, D. von Smolinski, and A. Gebert. 2007. 'Mechanisms of laser-induced dissection and transport of histologic specimens', *Biophys J*, 93: 4481-500.

von Minckwitz, G., A. du Bois, M. Schmidt, N. Maass, T. Cufer, F. E. de Jongh, E. Maartense, C. Zielinski, M. Kaufmann, W. Bauer, K. H. Baumann, M. R. Clemens, R. Duerr, C. Uleer, M. Andersson, R. C. Stein, V. Nekljudova, and S. Loibl. 2009. 'Trastuzumab beyond progression in human epidermal growth factor receptor 2-positive advanced breast cancer: a german breast group 26/breast international group 03-05 study', *J Clin Oncol*, 27: 1999-2006.

Vondeling, G. T., G. L. Menezes, E. P. Dvortsin, F. G. A. Jansman, I. R. Konings, M. J. Postma, and M. H. Rozenbaum. 2018. 'Burden of early, advanced and metastatic breast cancer in The Netherlands', *BMC Cancer*, 18: 262.

Waks, A. G., and E. P. Winer. 2019. 'Breast Cancer Treatment: A Review', *JAMA*, 321: 288-300.

Weckermann, D., B. Polzer, T. Ragg, A. Blana, G. Schlimok, H. Arnholdt, S. Bertz, R. Harzmann, and C. A. Klein. 2009. 'Perioperative activation of disseminated tumor cells in bone marrow of patients with prostate cancer', *J Clin Oncol*, 27: 1549-56.

Weigelt, B., F. L. Baehner, and J. S. Reis-Filho. 2010. 'The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective of the last decade', *J Pathol*, 220: 263-80.

Wen, H. Y., and E. Brogi. 2018. 'Lobular Carcinoma In Situ', *Surg Pathol Clin*, 11: 123-45.

Werner-Klein, M., S. Scheitler, M. Hoffmann, I. Hodak, K. Dietz, P. Lehnert, V. Naimer, B. Polzer, S. Treitschke, C. Werno, A. Markiewicz, K. Weidele, Z. Czyz, U. Hohenleutner, C. Hafner, S. Haferkamp, M. Berneburg, P. Rummele, A. Ulmer, and C. A. Klein. 2018. 'Genetic alterations driving metastatic colony formation are acquired outside of the primary tumour in melanoma', *Nat Commun*, 9: 595.

Willems, T., M. Gymrek, G. D. Poznik, C. Tyler-Smith, Y. Group Genomes Project Chromosome, and Y. Erlich. 2016. 'Population-Scale Sequencing Data Enable Precise Estimates of Y-STR Mutation Rates', *Am J Hum Genet*, 98: 919-33.

Woelfle, U., E. Breit, K. Zafrakas, M. Otte, F. Schubert, V. Muller, J. R. Izbicki, T. Loning, and K. Pantel. 2005. 'Bi-specific immunomagnetic enrichment of micrometastatic tumour cell clusters from bone marrow of cancer patients', *J Immunol Methods*, 300: 136-45.

Wolff, A. C., M. E. Hammond, D. G. Hicks, M. Dowsett, L. M. McShane, K. H. Allison, D. C. Allred, J. M. Bartlett, M. Bilous, P. Fitzgibbons, W. Hanna, R. B. Jenkins, P. B. Mangu, S. Paik, E. A. Perez, M. F. Press, P. A. Spears, G. H. Vance, G. Viale, D. F. Hayes, Oncology American Society of Clinical, and Pathologists College of American. 2013. 'Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update', *J Clin Oncol*, 31: 3997-4013.

Woodworth, M. B., K. M. Girskis, and C. A. Walsh. 2017. 'Building a lineage from single cells: genetic techniques for cell lineage tracking', *Nat Rev Genet*, 18: 230-44.

Yates, L. R., S. Knappskog, D. Wedge, J. H. R. Farmery, S. Gonzalez, I. Martincorena, L. B. Alexandrov, P. Van Loo, H. K. Haugland, P. K. Lilleng, G. Gundem, M. Gerstung, E. Pappaemmanuil, P. Gazinska, S. G. Bhosle, D. Jones, K. Raine, L. Mudie, C. Latimer, E. Sawyer, C. Desmedt, C. Sotiriou, M. R. Stratton, A. M. Sieuwerts, A. G. Lynch, J. W. Martens, A. L. Richardson, A. Tutt, P. E. Lonning, and P. J. Campbell. 2017. 'Genomic Evolution of Breast Cancer Metastasis and Relapse', *Cancer Cell*, 32: 169-84 e7.

8. Acknowledgements

First of all, I would like to express my sincere gratitude towards my PI Dr. Christoph A. Klein for allowing me to work in his laboratory and for supervising my doctoral research. I would like to thank him for being such a great mentor, under his guidance I have acquired valuable insights on how to address scientific questions and plan experiments with a reductionist approach. Also, I am grateful to him for allowing me to design my experiments independently but also guiding me whenever I needed. I would like to thank him for helping me sharpen my presentation skills and also allowing me to visit various conferences. A special thanks for organizing the FOR2127 “Meet the Experts” Seminar Series, where we were able to closely interact with renowned scientists in the field. I would like to mention that I have learnt to be logically optimistic and to be persistent with my efforts from him which he explained with an example of the human genome project that “Progress curve is exponential” and I found out that it actually is like that!

I would like to thank Stefanie Paush for helping me settle in Regensburg and till date helping me with my dumb administrative queries.

Next, I would like to thank DFG for financing me and my project. A bundle of gratitude towards the patients whose consent made this research possible. Many thanks to our clinical partner Dr Andreas Hartkopf and especially Vincent Walter without whose dedicated efforts I could not have imagined building the triplet cohort of patient samples.

I would like to thank our collaborators, Dr Ehud Shapiro and his team, especially Ofir Raz and Tzipora Marx for helping us with our research and promptly addressing our naive questions. Furthermore, I would like to thank our lineage tree team for their cooperation, Julia G. Junghans, Sandra Huber, Dr Melanie Werner Klein, and especially Dr. Zbigniew Czyż for the technical help.

Next, I want to thank Dr. Elisabeth Schneider, Oliver Schmidt, Dr. Miodrag Gužvić, Dr Sebastian Scheitler, Dr. Hedayatollah Hosseini, Dr Ana Grujovic for their insights and guidance during the course of my study. I would like to thank Dr. Giancarlo Feliciello for training me for LowPass analysis and helping me whenever I needed. Many thanks to ITEM-R members especially Dr. Bernhard Polzer and Dr. Stefan Kirsch for helping me with the contamination experiment, Dr Martin Hoffman for the CNA analysis from my data.

I would like to thank Dr. Courtney König, Dr Christoph Irlbeck and Dr Christoph Klein for proof reading parts of my thesis and my examiners Dr Gunter Meister and Dr Stephan Schneuwly.

I would also like to thank our SL-DTC team for introducing me to the methods in the lab and helping me with the patient samples. Special thanks to Manfred Meyer for his help with the arrayCGH experiments, and Isa always till date helps me with my DCC screening related questions (it is sometimes difficult to identify DCCs on cytospins!).

A special thanks to my students Eva Maria Hecht, Vincent Walter, Annina Grieff without whom I would have never finished my PhD.

I would like to thank all members of LEX and ITEM-R for the friendly atmosphere.
I am grateful to my RIGEL mentors Dr Gunter Meister and Nikolas Stoecklein for their time and valuable suggestions during progress reports.

From personal life, I would like to thank my friends Dr Amrita Das, Swarupa Chatterjee, Dr Rohit Menon, Dr Somnath Das, Dr. Zbigniew Czyż, Regina Hoheisel who made life fun in Regensburg. Special thanks to my Zumba instructor Virginia. Many thanks to my long-distance friends Poonam Kumari, Trishtina Hembram, Sujith Naik, Kireeti Bodduna, Dr Harshit Lakhotia, Ankan Bag, Artitra Mukhopadhyaya, John Jose, Mamata Vats and Shipra Gupta for always being there.

Lastly, I would like to thank my lovely, parents; Rupali and Ashutosh Ghosh, brother; Anirban Ghosh and husband; Paramveer for their unconditional love and support throughout my life.

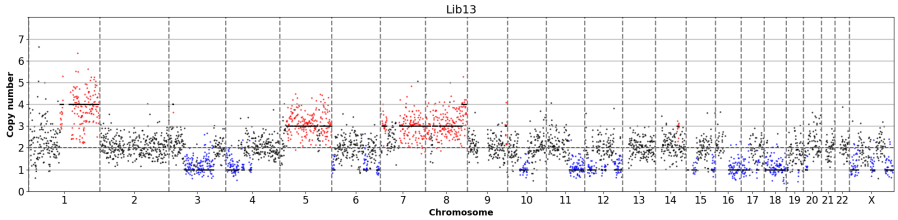
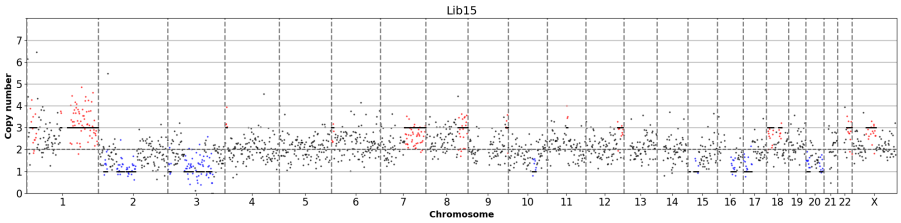
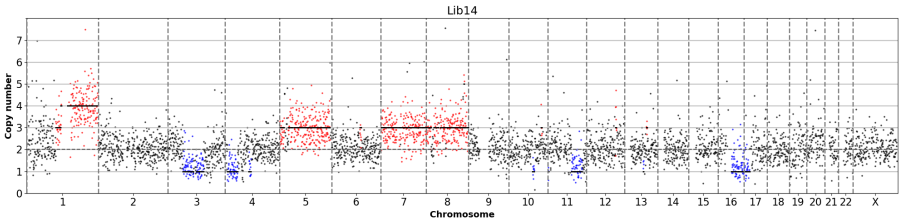
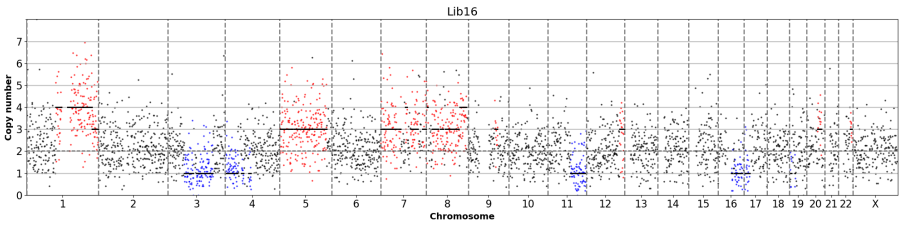
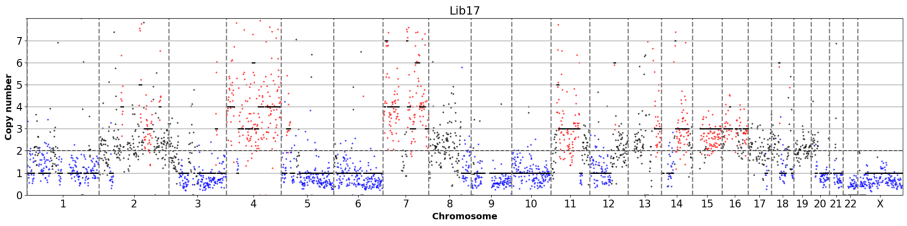
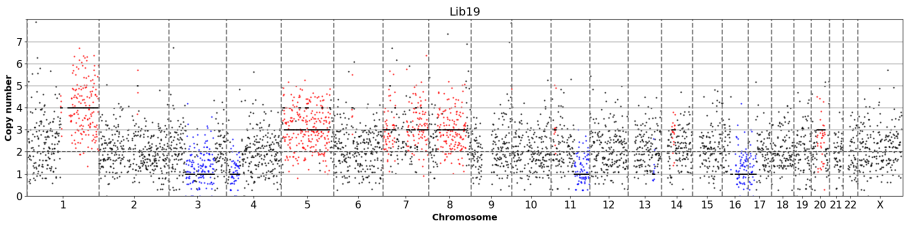
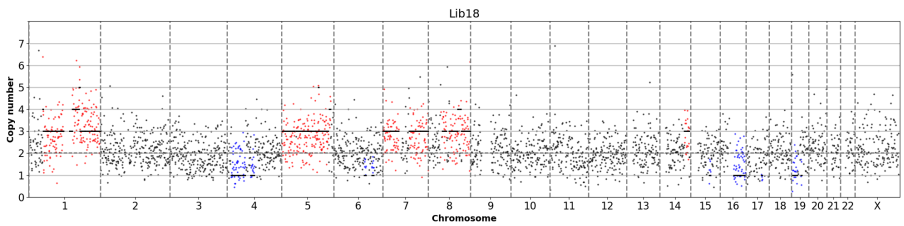
Appendix

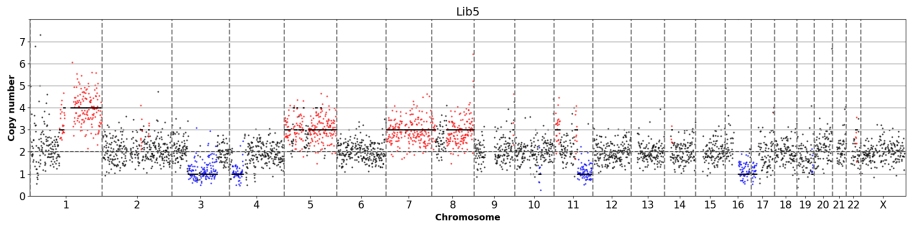
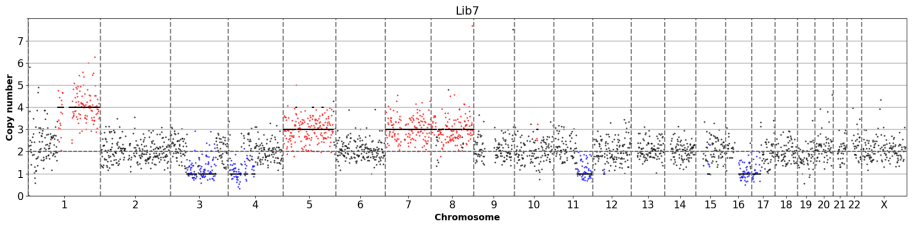
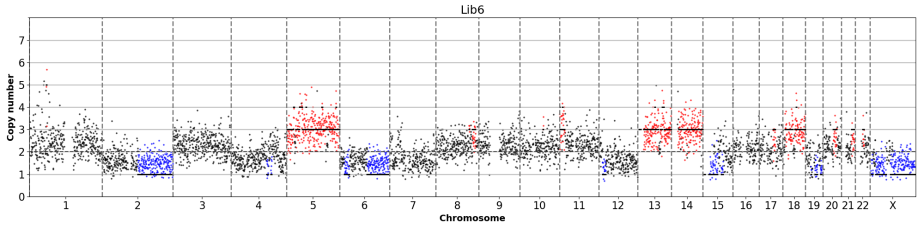
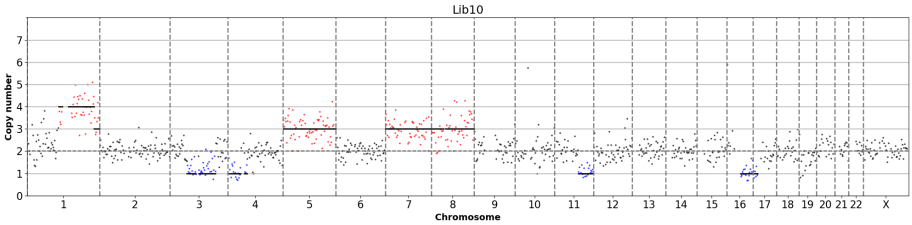
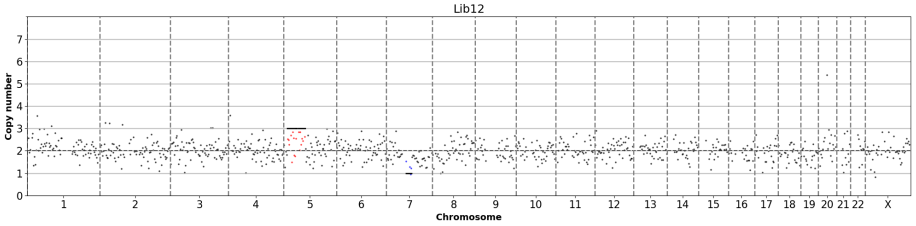
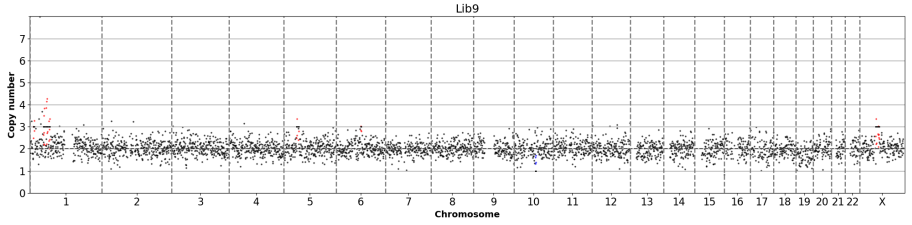
1. Single cell DNA quality of archived frozen cytospins. (see chapter 4.1.1.)

BM Sample ID	BM isolation date	Age of cytospin at isolation	No. of picked cells	Cells with GII>2	Percentage of cells with GII>2
113	01.06.18	0	10	9	90
114	01.06.18	0	5	3	60
115	01.06.18	0	10	9	90
116	01.06.18	0	4	4	100
123	01.06.15	0	8	5	62.5
120	01.06.14	1	5	5	100
122	01.06.14	1	5	5	100
92	01.06.13	2	3	3	100
78	01.06.15	3	5	2	40
98	01.06.12	3	9	9	100
107	01.06.13	3	10	4	40
109	01.06.15	3	2	2	100
119	01.06.12	3	5	5	100
1	11.09.13	4	5	5	100
71	13.05.11	4	10	10	100
83	01.06.12	4	9	6	66.6
97	01.06.11	4	10	8	80
104	01.06.12	4	10	6	60
110	01.06.14	4	2	0	0
79	01.06.13	5	10	4	40
88	01.06.10	5	4	2	50
96	01.06.10	5	7	1	14
2	24.08.11	6	4	1	25
3	16.08.11	6	1	1	100
77	01.06.10	6	8	8	100
80	01.06.12	6	10	5	50
87	01.06.09	6	5	5	100
90	01.06.09	6	1	1	100
91	01.06.09	6	5	3	60
93	01.06.09	6	8	8	100
99	01.06.10	6	14	10	71.4
100	01.06.10	6	9	2	20
103	01.06.10	6	5	0	0
105	01.06.10	6	18	18	100
111	01.06.12	6	2	1	50
112	01.06.12	6	4	4	100
81	01.06.11	7	9	4	44.4
85	01.06.09	7	10	9	90
95	01.06.08	7	10	6	60
101	01.06.09	7	10	9	90
102	01.06.09	7	10	5	50
108	01.06.10	7	10	7	70
117	01.06.11	7	2	1	50
4	12.05.10	8	4	3	75
5	28.04.10	8	5	5	100
6	20.01.10	8	4	3	75
7	14.07.09	8	4	0	0
75	09.07.08	8	4	0	0
82	01.06.10	8	10	5	50
84	01.06.10	8	10	7	70
89	01.06.07	8	3	3	100
94	01.06.07	8	9	9	100
118	01.06.10	8	4	2	50
121	01.06.08	8	9	9	100

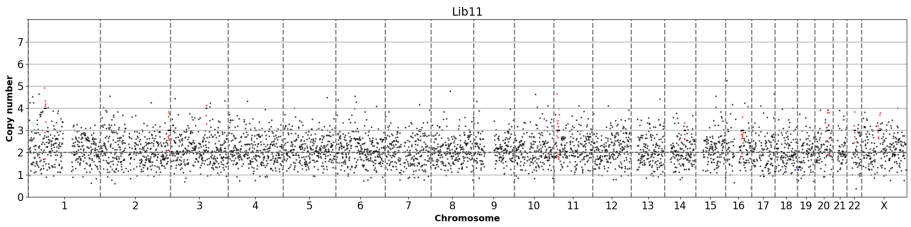
BM Sample ID	BM isolation date	Age of cytospin at isolation	No. of picked cells	Cells with GII>2	Percentage of cells with GII>2
8	22.04.09	9	3	0	0
9	21.04.09	9	4	0	0
10	21.04.09	9	4	4	100
11	13.01.09	9	4	0	0
12	09.12.08	9	15	5	33
13	03.12.08	9	4	2	50
14	25.11.08	9	4	0	0
15	05.11.08	9	3	0	0
16	14.10.08	9	11	11	100
17	15.07.08	9	5	3	60
18	04.06.08	9	3	3	100
72	20.05.08	9	8	6	75
86	01.06.09	9	10	7	70
19	29.01.08	10	5	5	100
20	16.01.08	10	5	4	80
21	15.01.08	10	10	4	40
22	05.12.07	10	8	6	75
23	30.10.07	10	5	5	100
24	23.10.07	10	5	1	20
25	19.06.07	10	4	1	25
73	22.05.07	10	8	6	75
74	14.03.07	10	10	3	30
26	16.05.07	11	5	4	80
27	15.05.07	11	5	0	0
28	18.04.07	11	5	5	100
29	04.04.07	11	5	3	60
30	27.03.07	11	5	3	60
31	21.03.07	11	5	2	40
32	28.02.07	11	5	3	60
33	23.01.07	11	9	7	78
34	17.01.07	11	3	3	100
35	06.12.06	11	5	3	60
36	21.11.06	11	5	3	60
37	21.11.06	11	5	3	60
38	15.11.06	11	3	1	33
39	26.09.06	11	10	5	50
40	20.09.06	11	4	0	0
41	18.08.06	11	9	8	89
42	04.08.06	11	8	3	37
43	04.04.06	12	5	3	60
44	10.02.06	12	3	1	33
45	29.07.05	12	4	0	0
46	27.07.05	12	8	3	37.5
47	01.07.05	12	11	5	45.5
48	29.06.05	12	11	2	18
49	28.06.05	12	4	3	75
50	21.06.05	12	8	2	25
106	01.06.04	12	7	4	57
51	04.05.05	13	4	1	25
52	28.04.05	13	4	1	25
53	23.02.05	13	5	1	20
54	21.02.05	13	5	4	80
55	26.01.05	13	5	1	20
56	12.08.04	13	3	0	0
57	18.05.04	14	9	2	22
76	01.06.04	14	10	4	40
58	04.12.02	15	4	3	75
59	24.09.02	15	9	4	44.5
60	17.09.02	15	4	0	0
61	06.09.02	15	5	2	40
62	16.07.02	15	10	6	60
63	27.06.02	15	5	4	80
64	12.06.02	15	5	0	0
65	04.06.02	16	5	3	60
66	15.05.02	16	6	2	33
67	08.05.02	16	4	3	75
68	06.05.02	16	5	2	40
69	10.04.02	16	5	2	40
70	08.02.00	18	5	4	80

2. CNA profiles of single cells isolated from FFPE tumor sections (see chapter 4.2.2.1).

Treatment	CNA Profile
No treatment	
No treatment	
No treatment	
No treatment	
FFPE repair	
FFPE repair	
FFPE repair	

Treatment	CNA Profile
No treatment	
No treatment	
No treatment	
No treatment	
FFPE repair	
FFPE repair	

FFPE repair



3. CNA profiles of all cells from patient 01.

Tree IDs are the numbers corresponding to cells on the lineage trees shown in in Chapter 4.5.5.1, Figure 33 and Figure 34. Cell IDs are unique IDs for each single cell. Cells IDs corresponding to cells from respective groups are mentioned below.

Primary tumor cells: Cell IDs 6119-6143

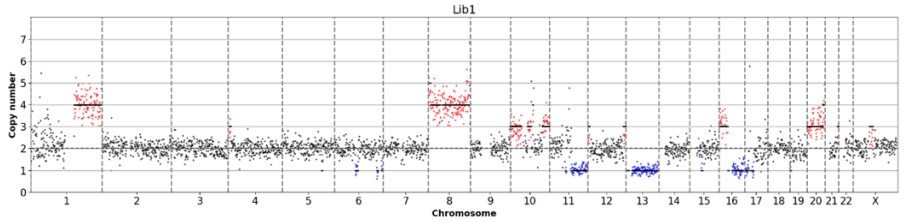
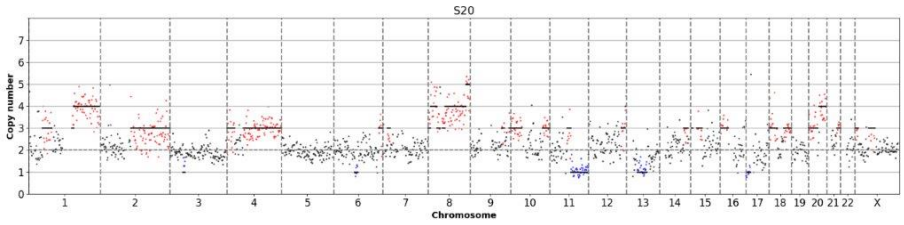
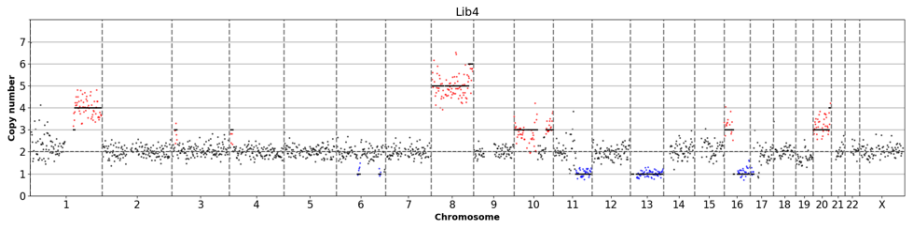
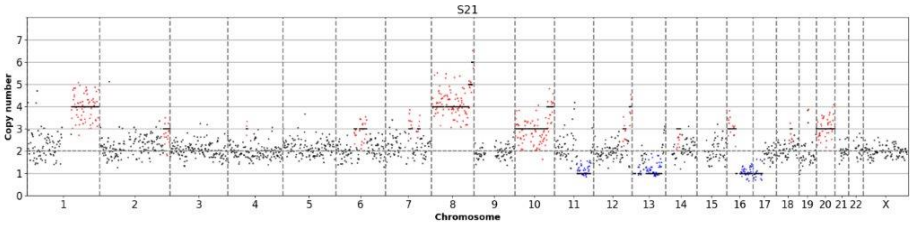
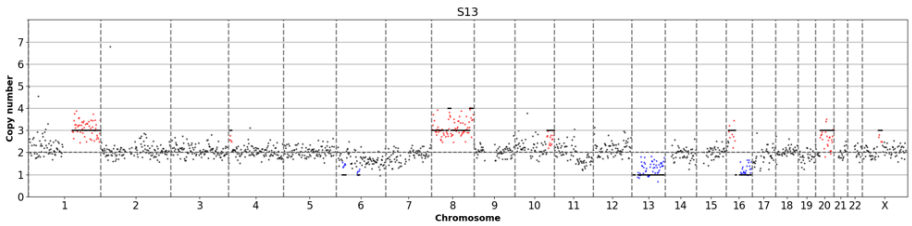
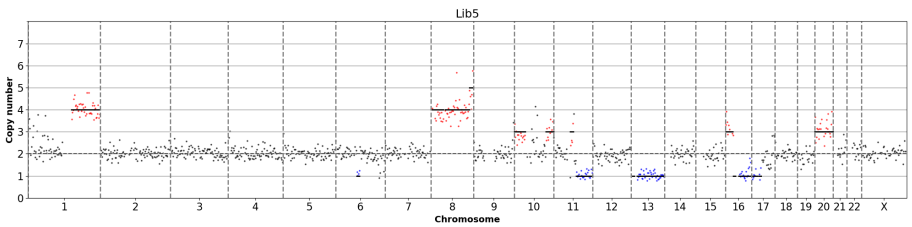
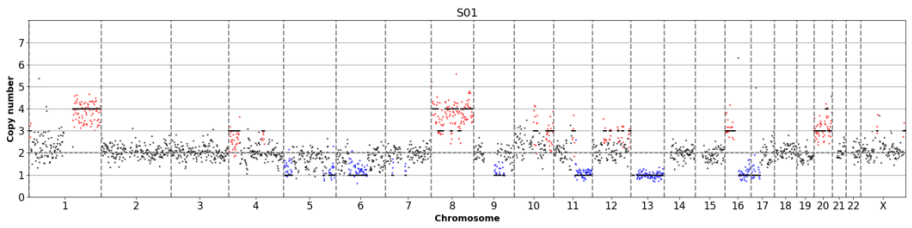
M0-DCC: Cell IDs 6144-6145

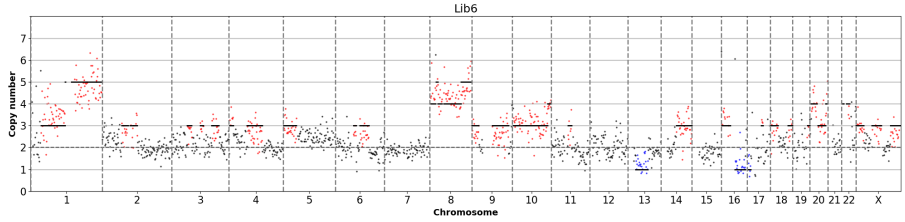
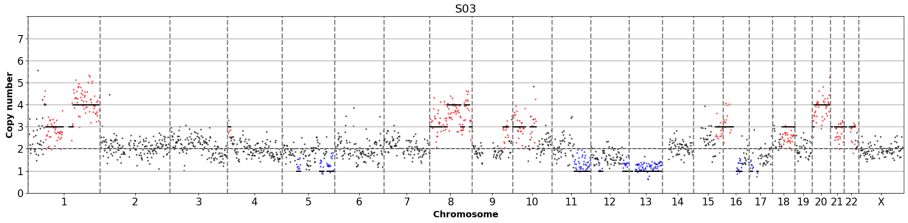
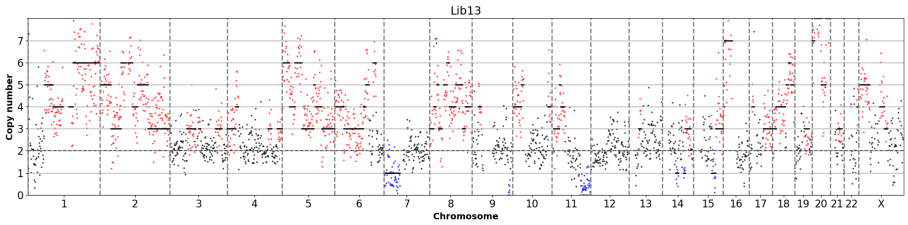
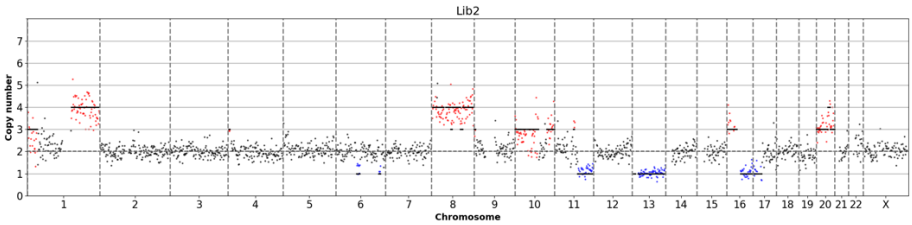
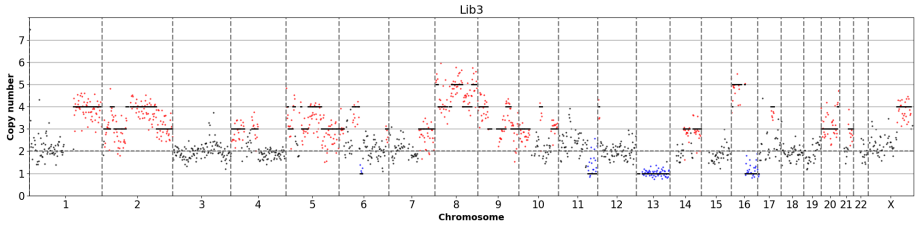
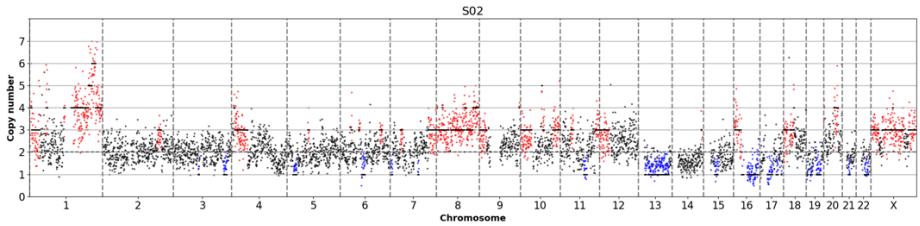
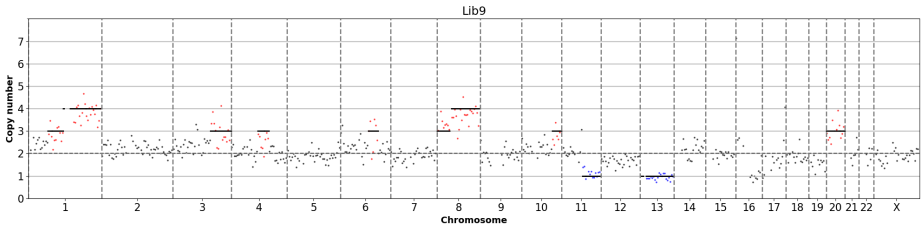
M1-CTC: Cell IDs 6146-6163

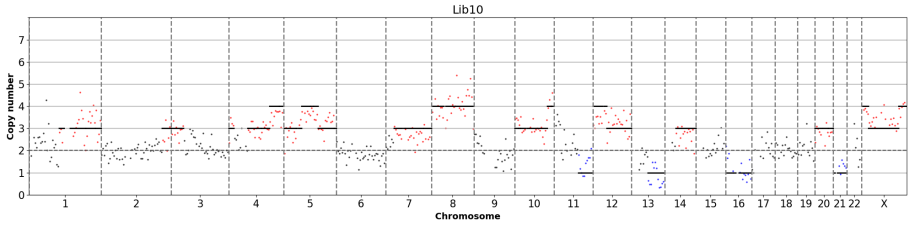
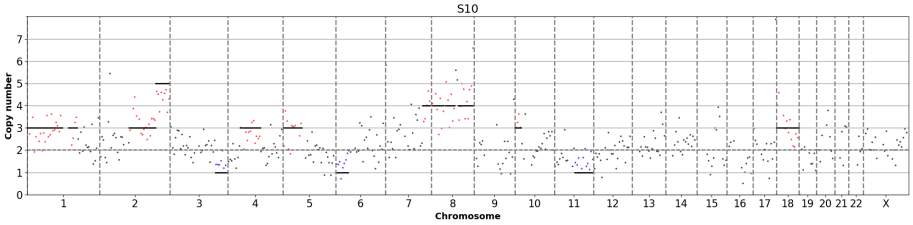
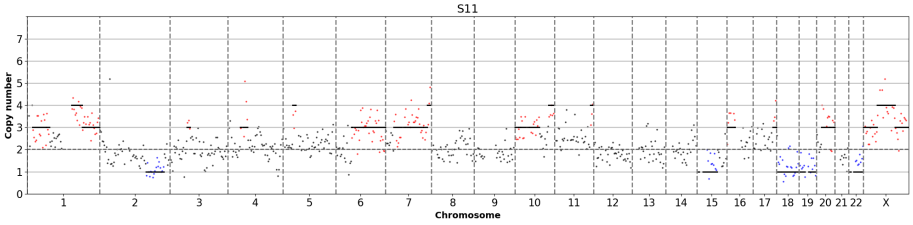
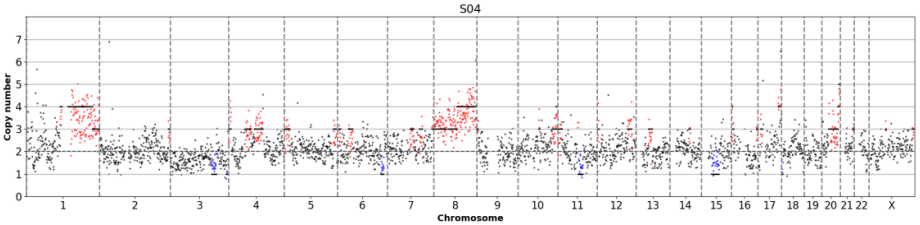
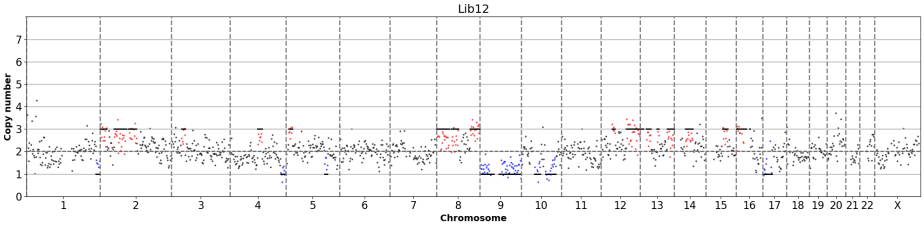
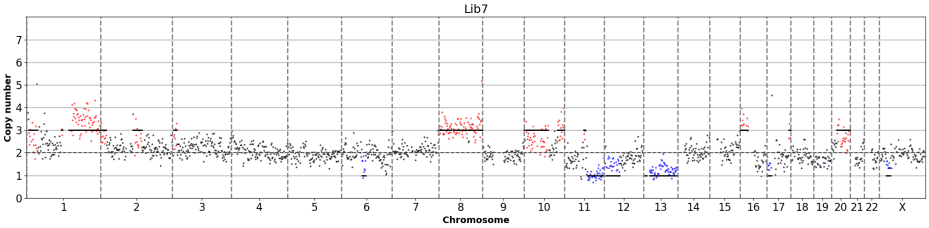
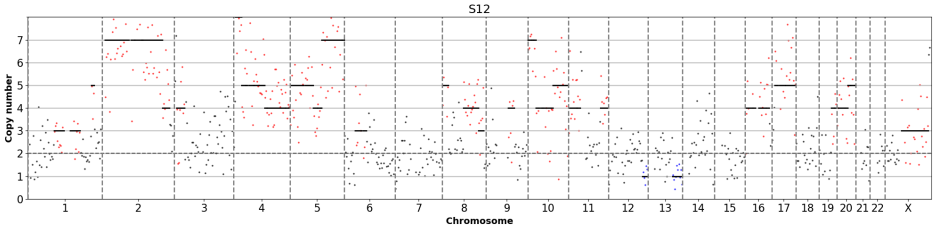
Macrophages: Cell IDs 6164-6188

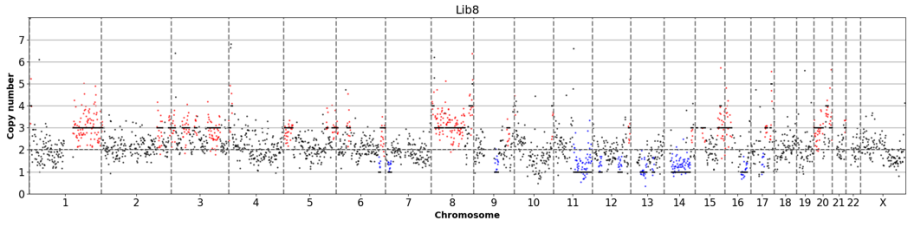
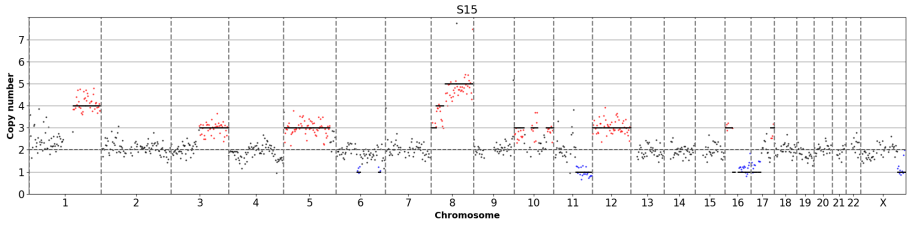
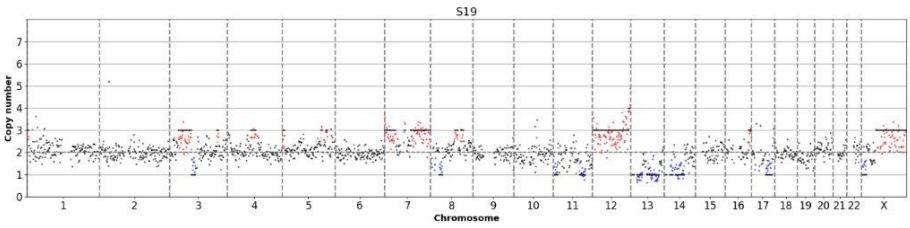
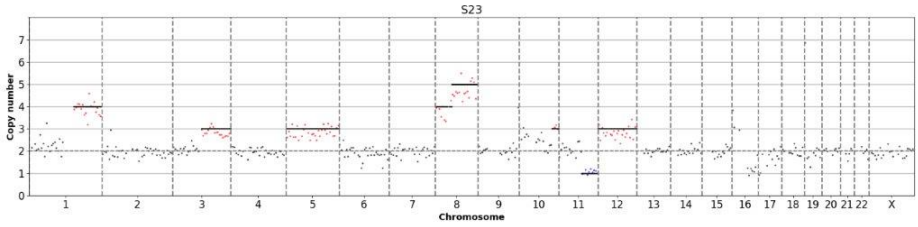
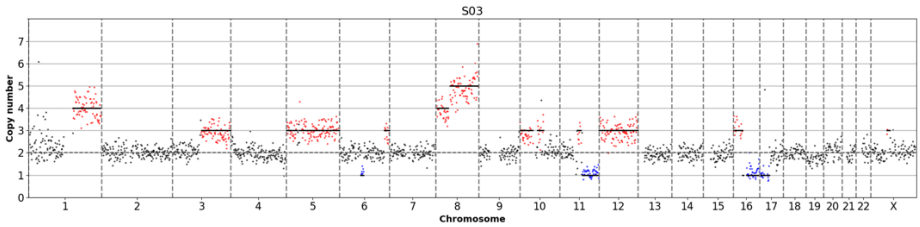
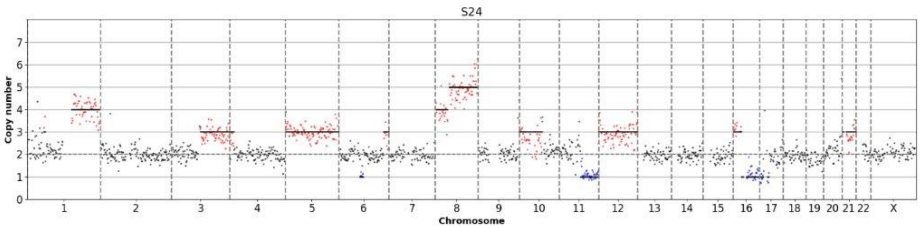
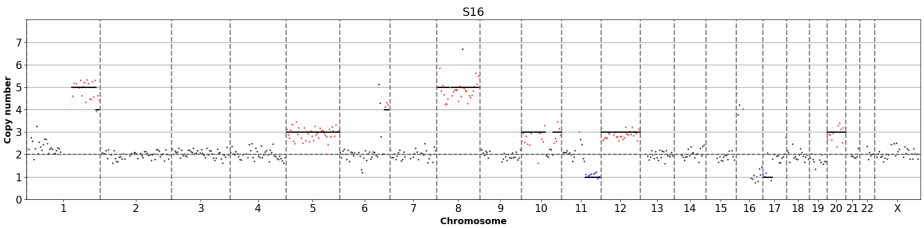
T-cells: Cell IDs 6189-6213

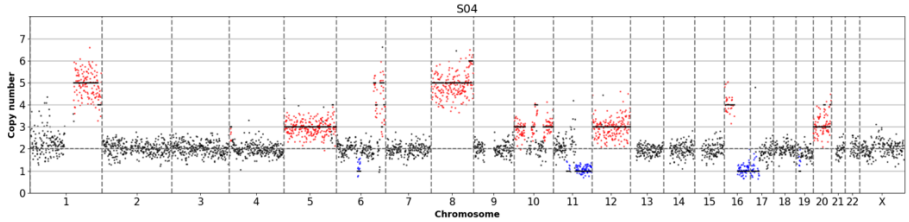
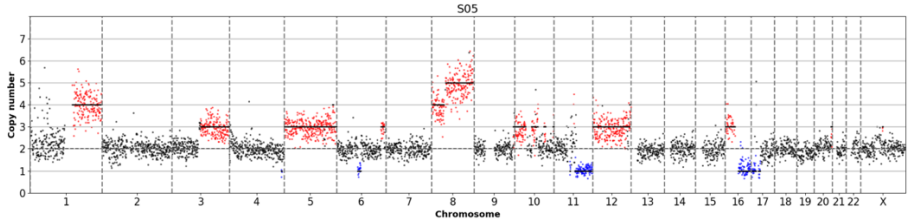
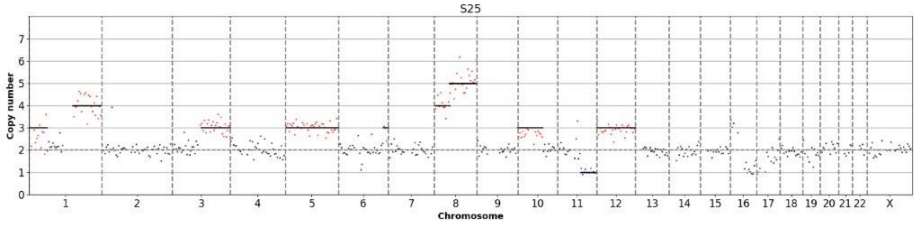
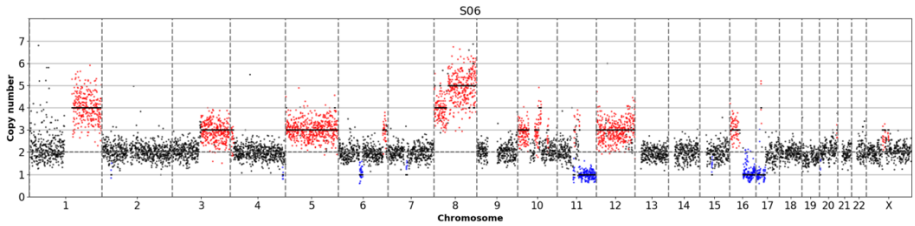
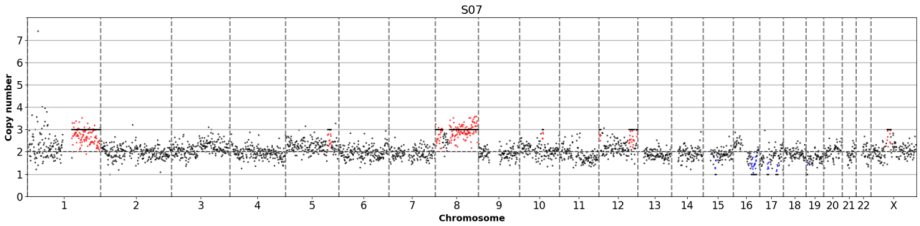
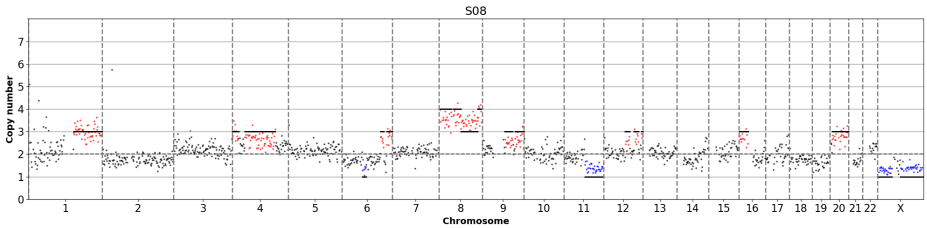
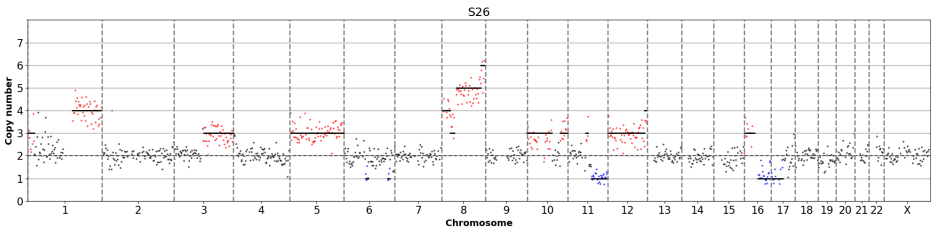
OECs: Cell IDs 6725-6735

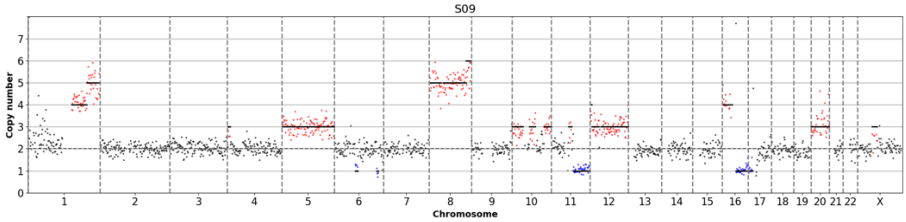
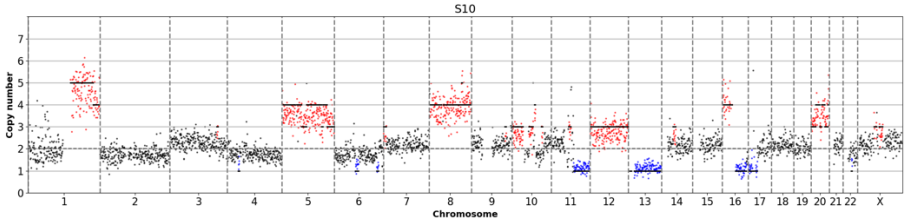
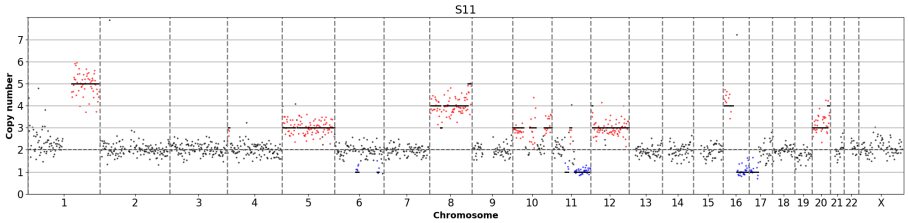
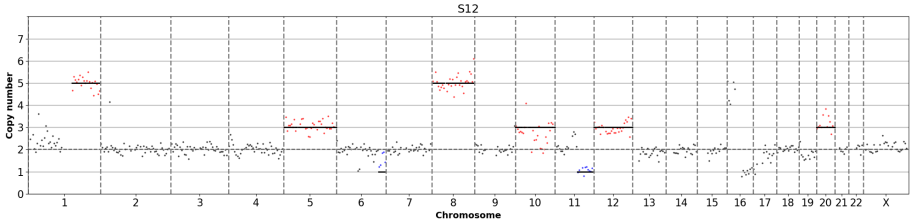
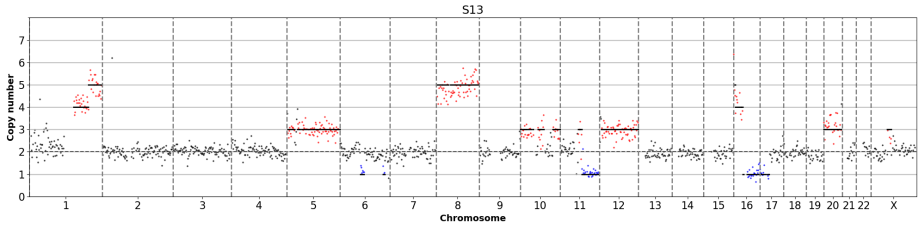
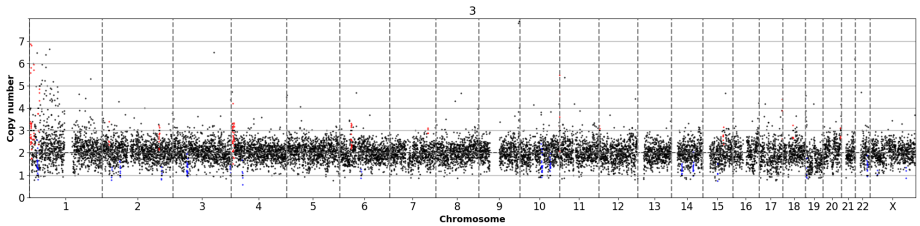
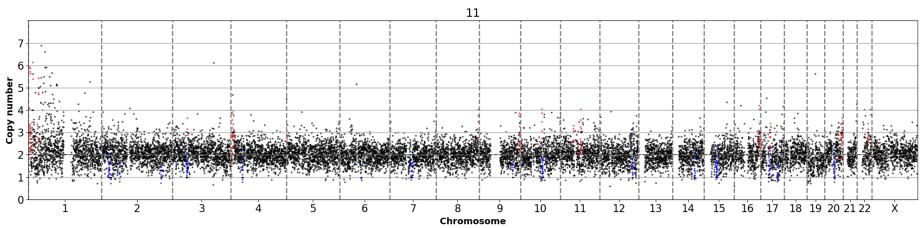
Tree-ID, Cell-ID	CNA Profile
40216, 6119	
40217, 6120	
40218, 6121	
40219, 6122	
40220, 6123	
40221, 6125	
40222, 6126	

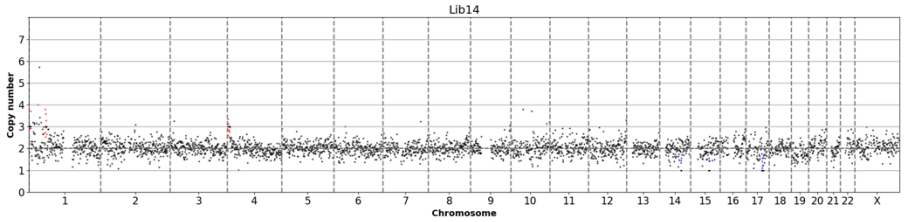
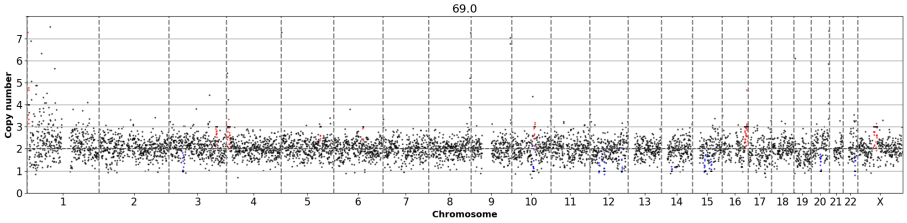
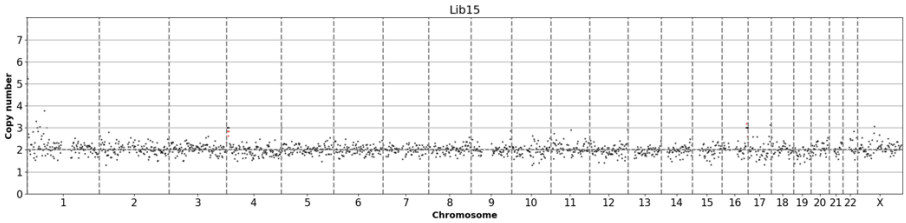
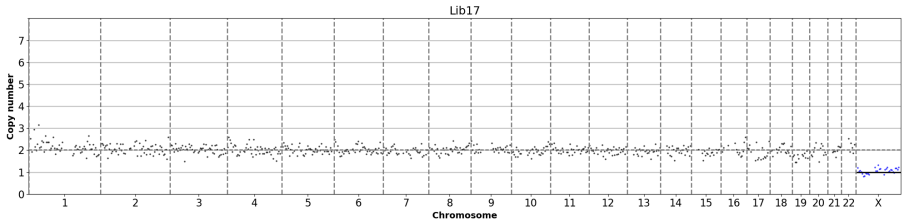
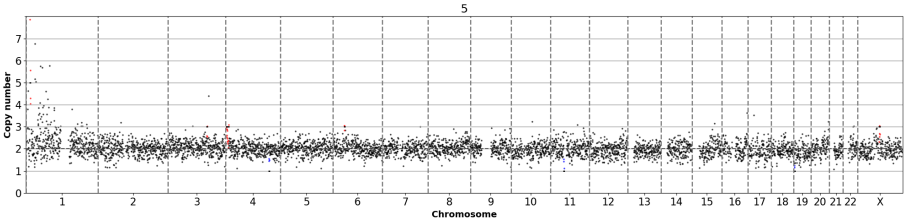
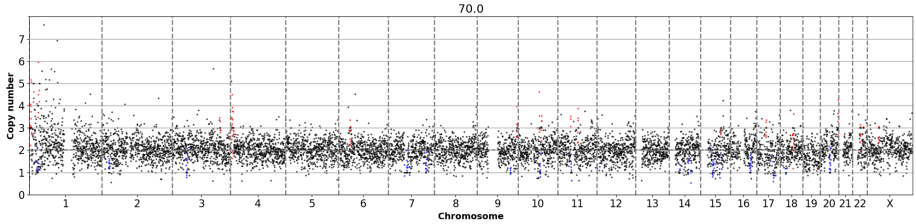
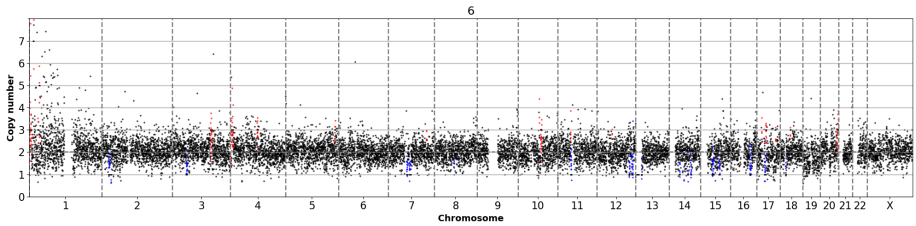
Tree-ID, Cell- ID	CNA Profile
40223, 6127	
40224, 6128	
40225, 6129	
40226, 6130	
40227, 6132	
40228, 6133	
40229, 6134	

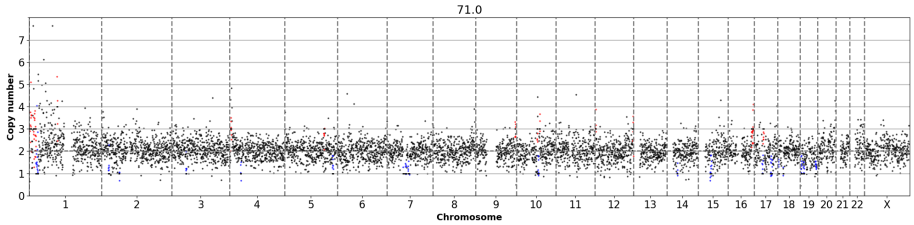
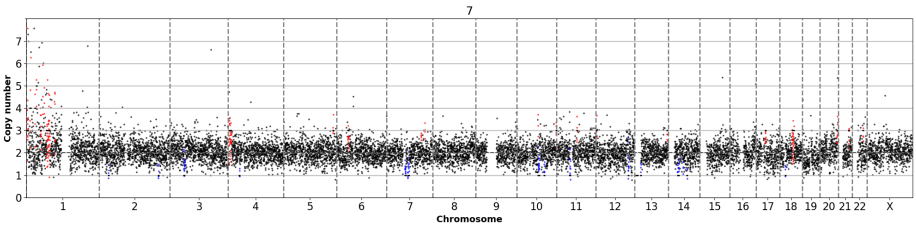
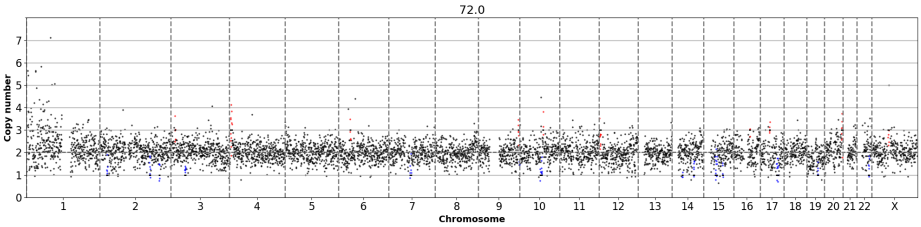
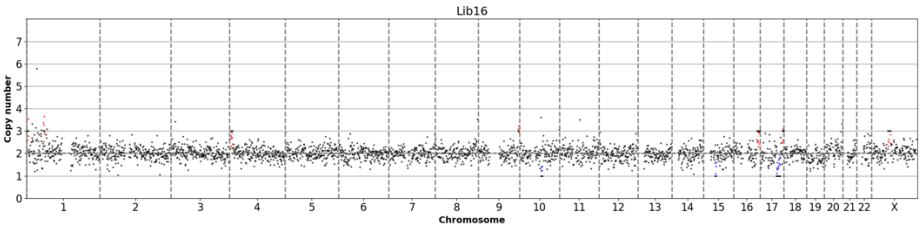
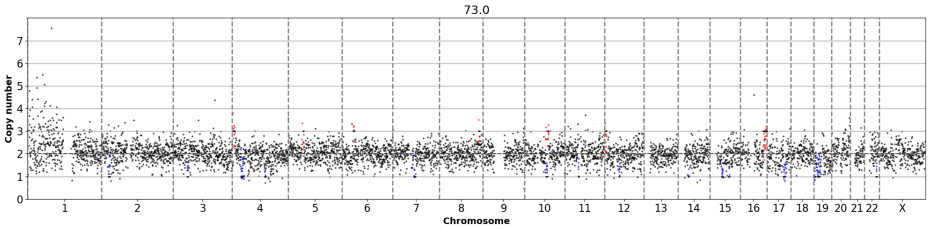
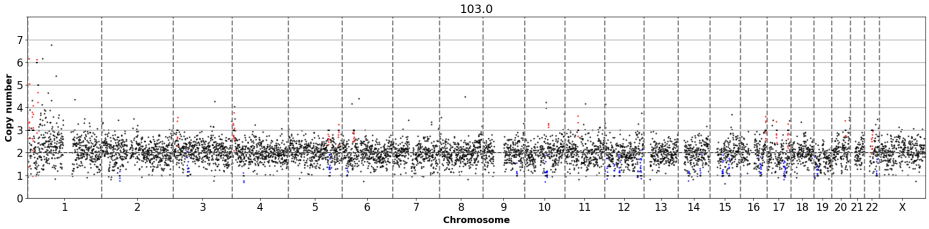
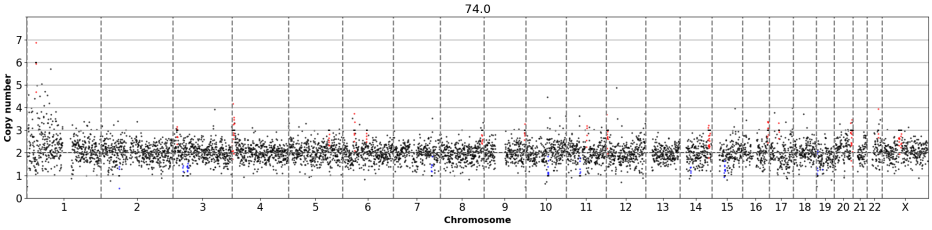
Tree-ID,Cell-ID	CNA Profile
40230, 6135	
40231, 6136	
40232, 6137	
40234, 6139	
40235, 6140	
40236, 6141	
40237, 6142	

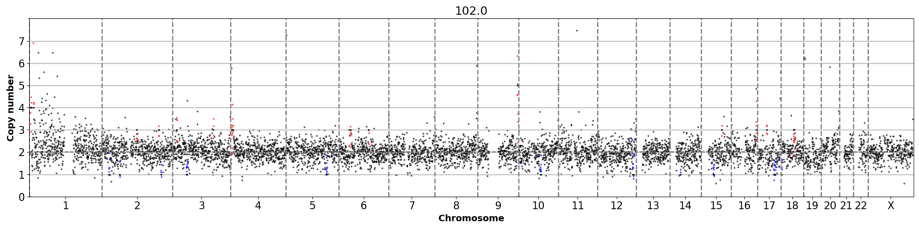
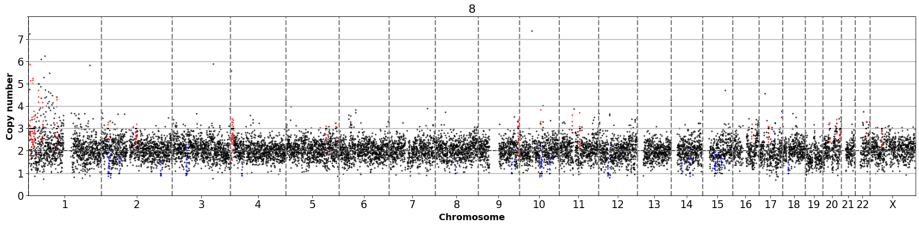
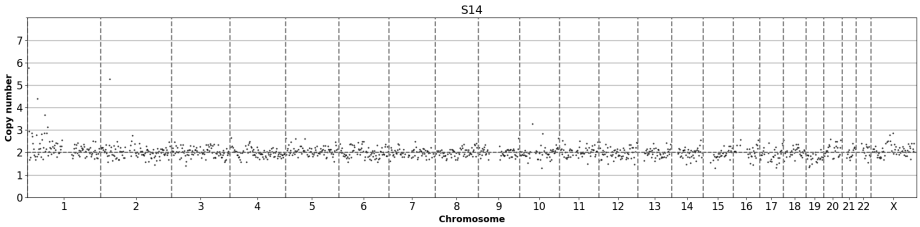
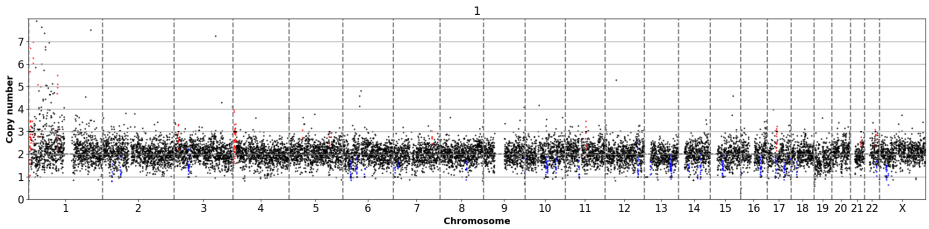
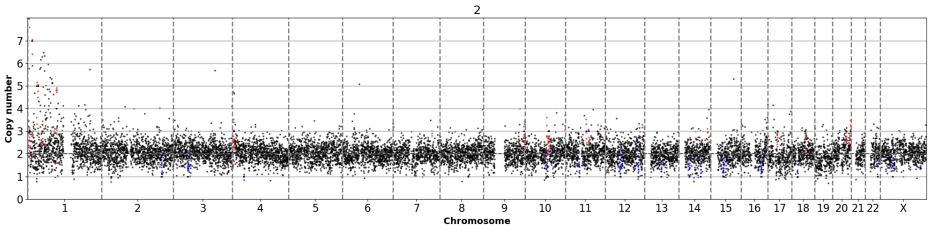
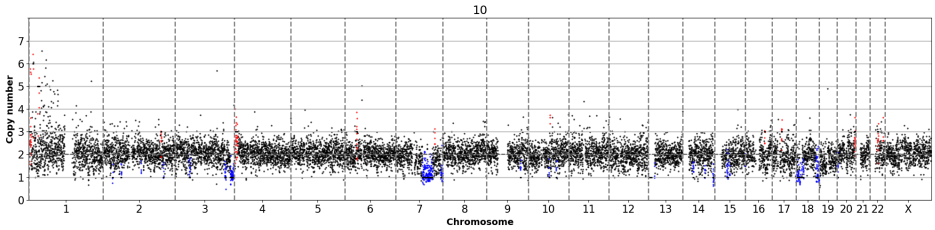
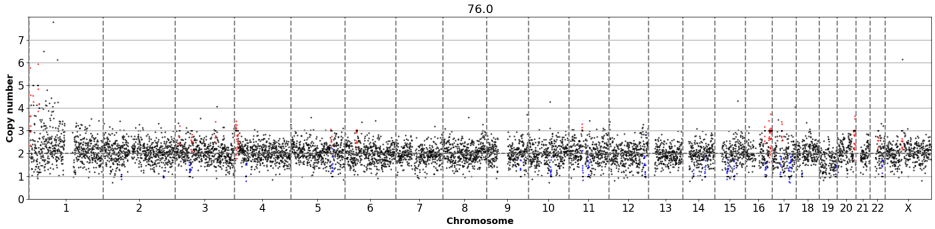
Tree-ID,Cell-ID	CNA Profile
40238, 6143	
40239, 6144	
40240, 6145	
28823, 6146	
40241, 6147	
40242, 6148	
40243, 6149	

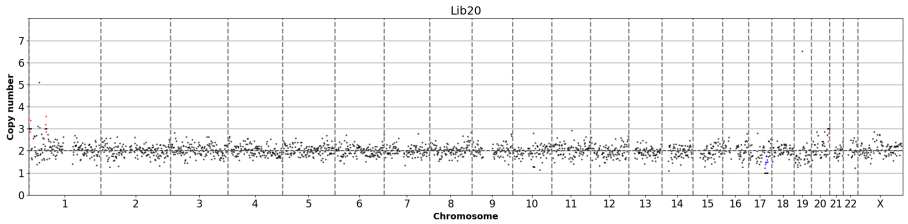
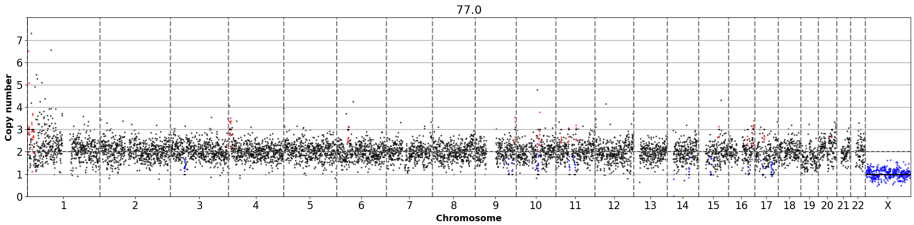
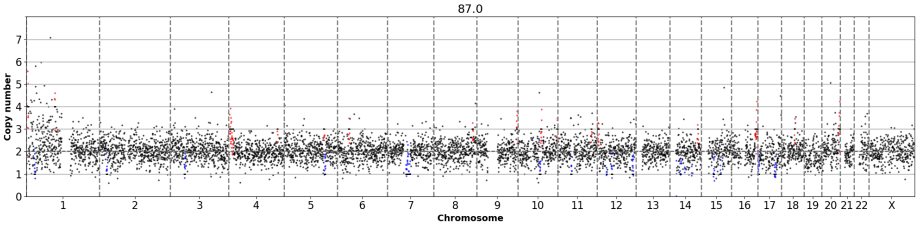
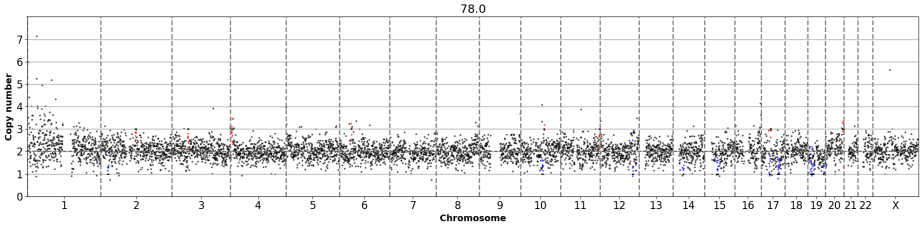
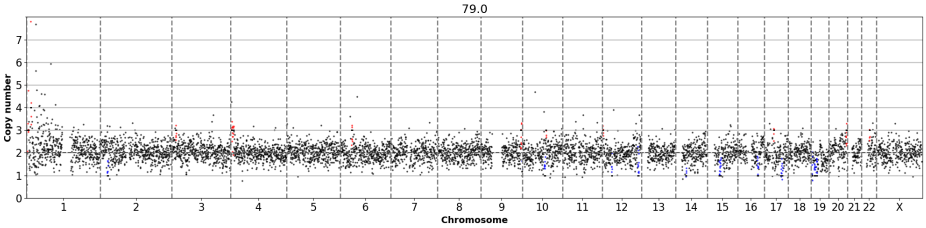
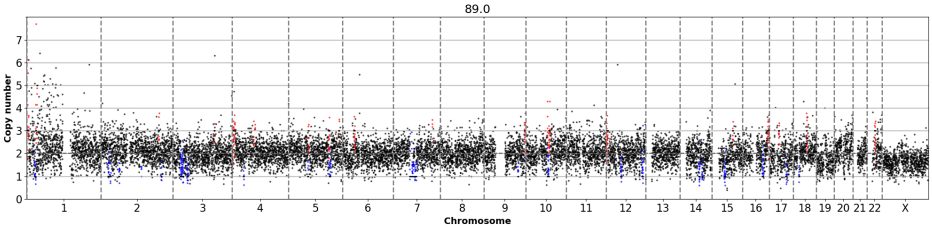
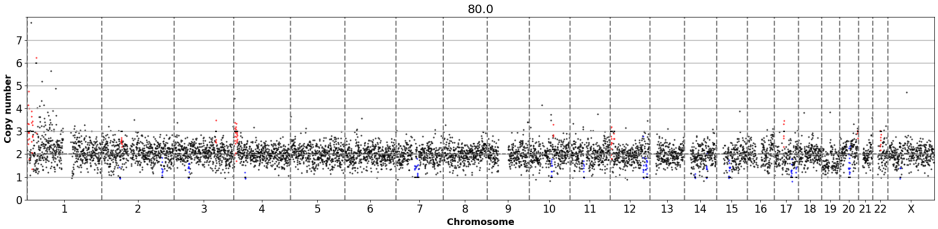
Tree-ID,Cell-ID	CNA Profile
40244, 6150	
40245, 6151	
40246, 6152	
40247, 6153	
40248, 6154	
40249, 6155	
40250, 6156	

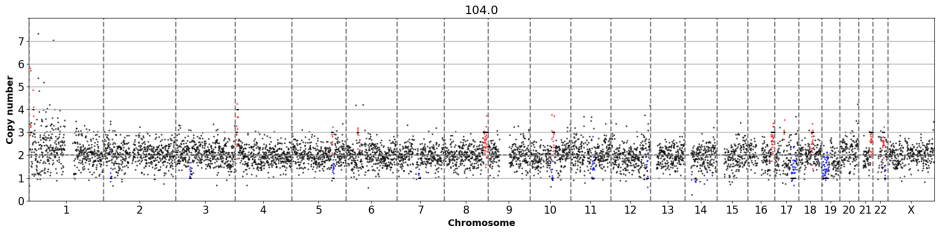
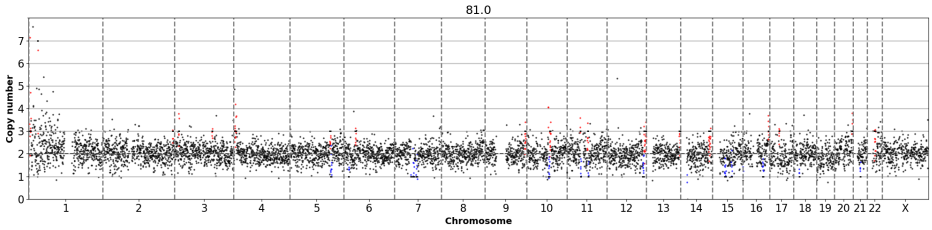
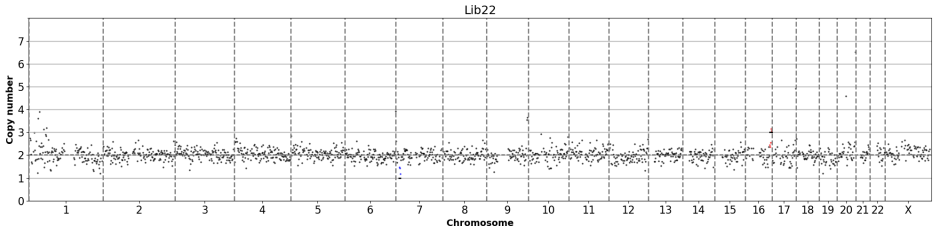
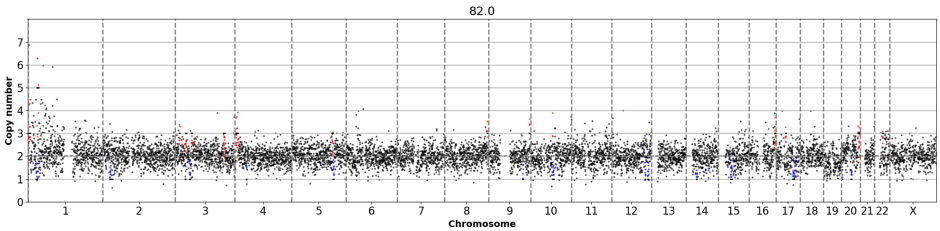
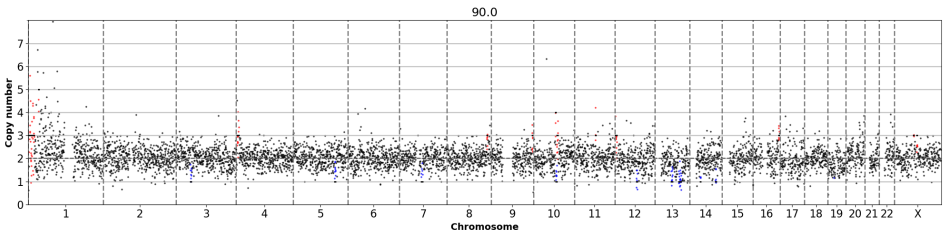
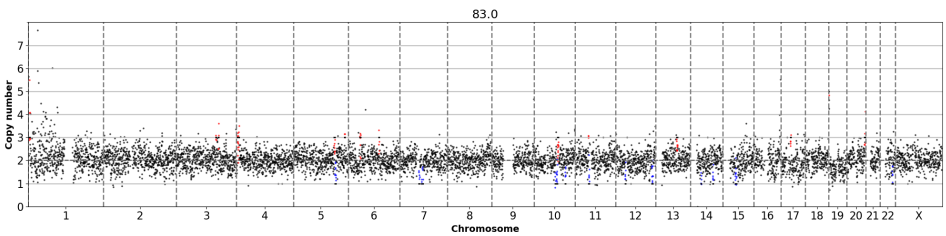
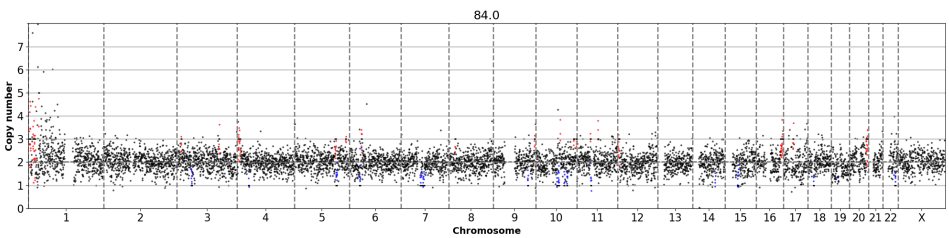
Tree-ID,Cell-ID	CNA Profile
40251, 6157	
28879, 6158	
40252, 6159	
40253, 6160	
40254, 6163	
40255, 6164	
40256, 6165	

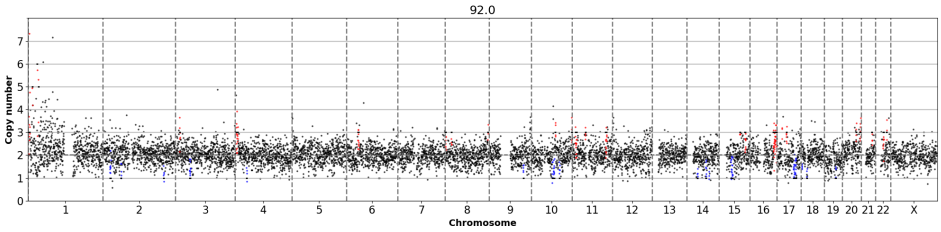
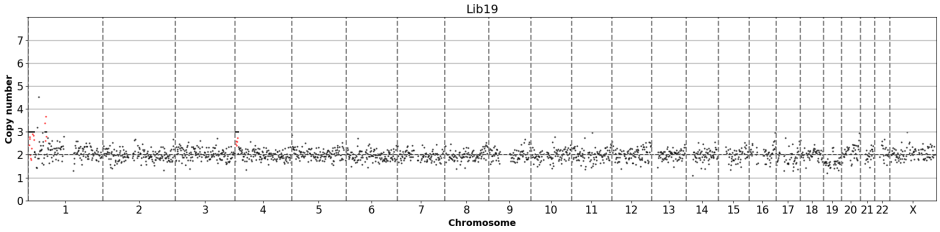
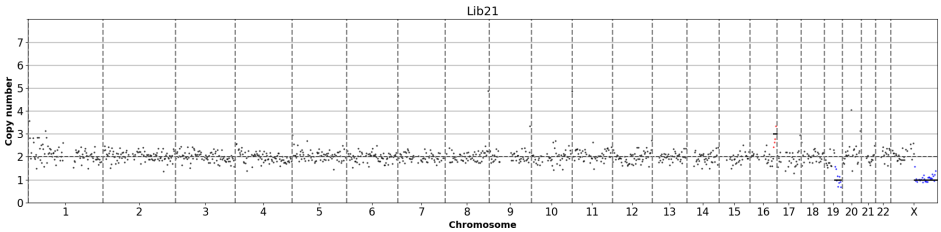
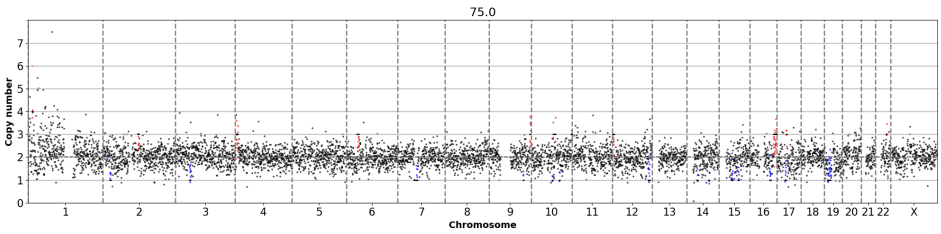
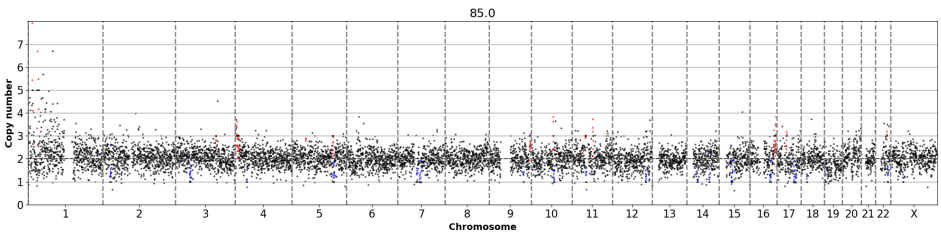
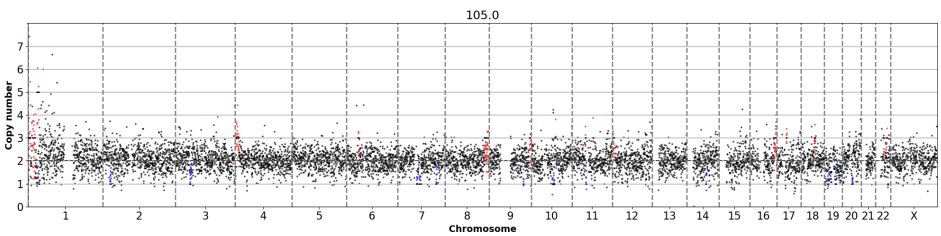
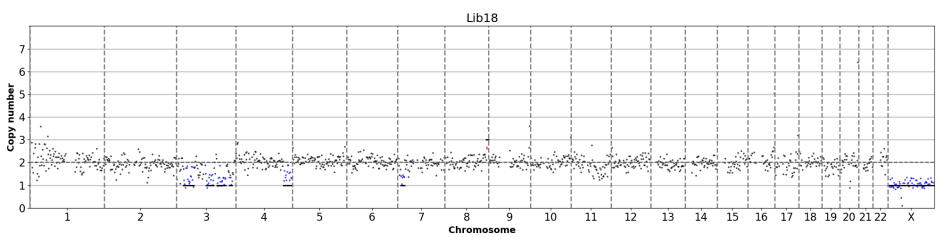
Tree-ID,Cell-ID	CNA Profile
40257, 6168	
40258, 6169	
40259, 6170	
40260, 6171	
40261, 6172	
40262, 6173	
40263, 6174	

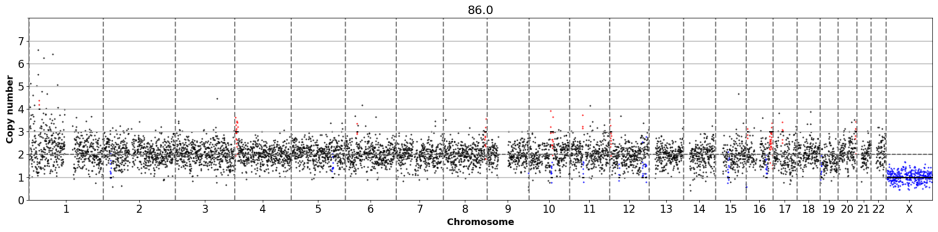
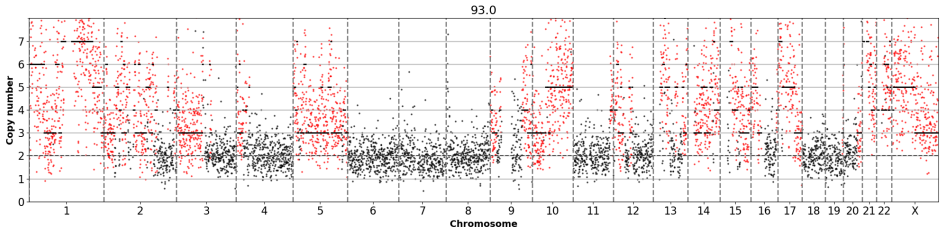
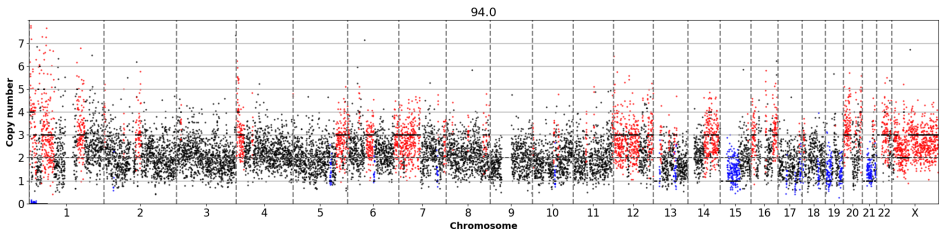
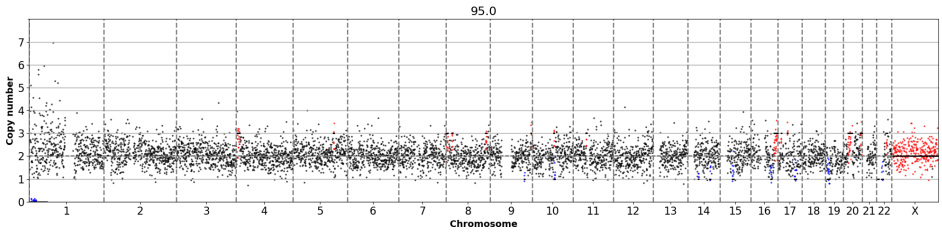
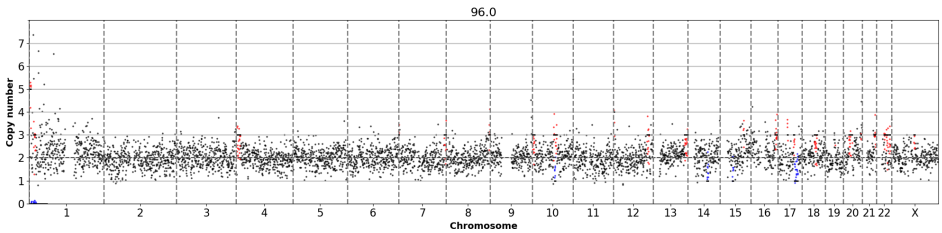
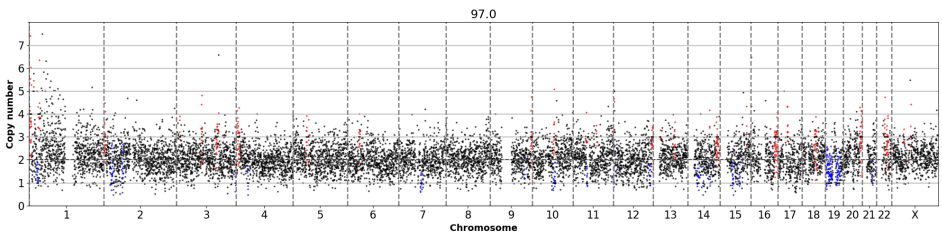
Tree-ID,Cell-ID	CNA Profile
40264, 6175	
40265, 6176	
40266, 6177	
40267, 6178	
40268, 6179	
40269, 6180	
40270, 6181	

Tree-ID,Cell-ID	CNA Profile
40271, 6182	
40272, 6183	
40273, 6184	
40274, 6185	
40275, 6187	
40276, 6188	
40277, 6189	

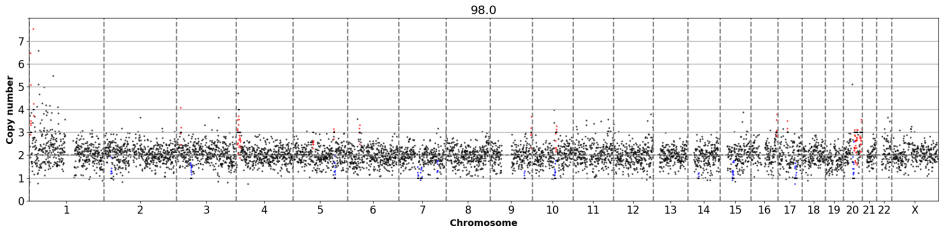
Tree-ID,Cell-ID	CNA Profile
40278, 6190	
40279, 6191	
40280, 6192	
40281, 6193	
40282, 6195	
40283, 6196	
40284, 6197	

Tree-ID,Cell-ID	CNA Profile
40285, 6198	
28844, 6199	
40286, 6200	
40287, 6201	
40288, 6202	
40289, 6203	
40290, 6205	

Tree-ID,Cell-ID	CNA Profile
40291, 6206	
40292, 6207	
40293, 6208	
40294, 6209	
40295, 6210	
40296, 6211	
28895, 6212	

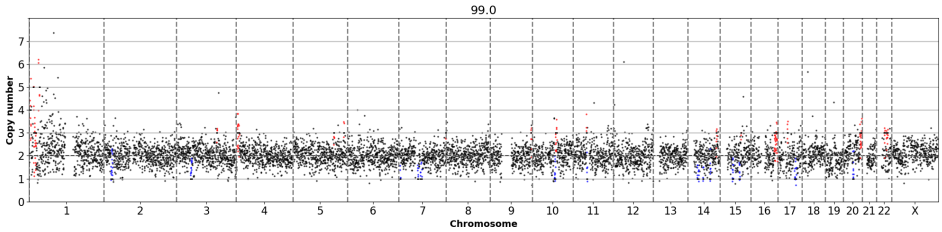
Tree-ID,Cell-ID	CNA Profile
40297, 6213	
40298, 6725	
40299, 6726	
40300, 6727	
40301, 6728	
40302, 6729	

40303, 6730

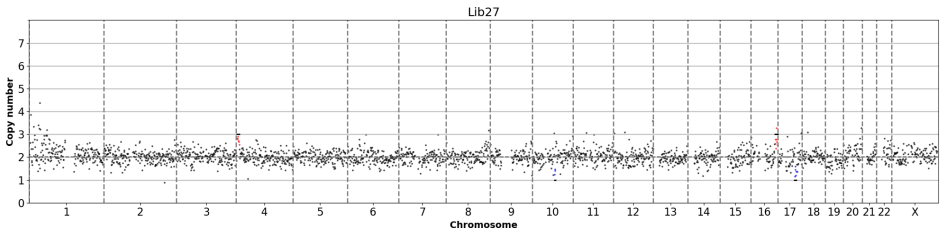


Tree-ID,Cell-ID	CNA Profile

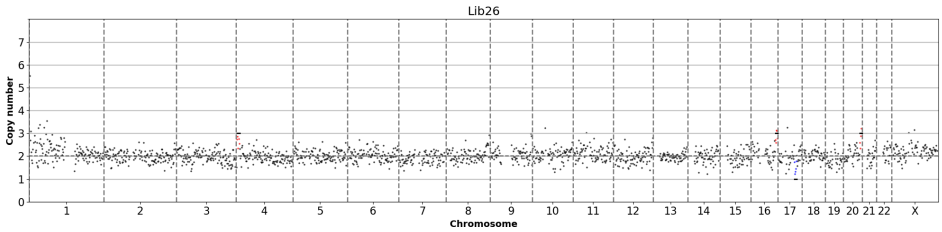
40304, 6731



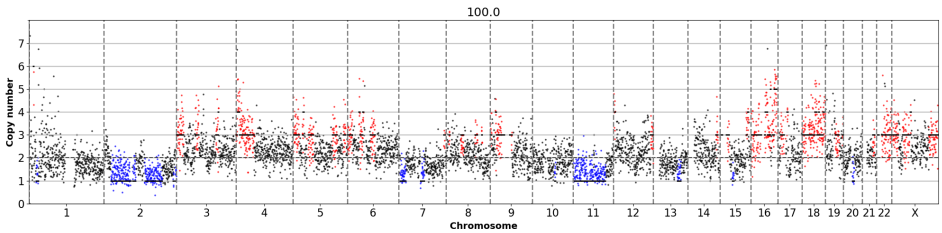
40305, 6732



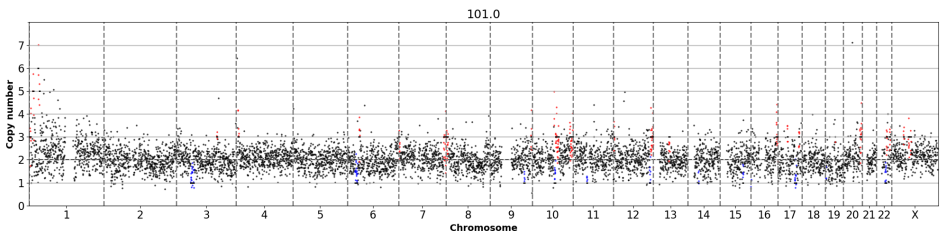
40306, 6733



40307, 6734



40308, 6735



4. Hierarchical clustering of cells with CNA data

CNA data obtained by *Ampli 1*^{TM} Lowpass-Sequencing of all the cancer cells of patient 01 were used by Dr Martin Hoffman at ITEM-R to obtain a heatmap of genome wide aberrations (Figure 36A). The heatmap and denodogram is based on segmentation files from LowPass sequencing (Silicon Biosystems pipeline) which provide integer copy numbers. The dendogram showing hierarchical clustering of the cancer cells on the basis of CNAs can be seen in figure below. It can be noted that majority of the metastatic M1-CTCs are clustering separately from the primary tumor cells. From the heatmap (Figure 36A) it can be seen that two M1-CTCs and one M0-DCC (marked by arrows) have fewer aberrations compared to the rest of M1-CTCs and are clustering separately from the rest of the M1-CTCs (Figure 36B). Interestingly, the same cells are also clustered away from the M1-CTC cluster in the STR based cell lineage tree (ID40248, ID40249, ID40240 in Figure 34).

Cell lineage tree reconstruction using CNV data is an ongoing project and is under development, however this preliminary hierarchical clustering gives us a qualitative and quantitative overview of the heterogeneity in the distribution of CNAs across all the tumor cells from various time points.

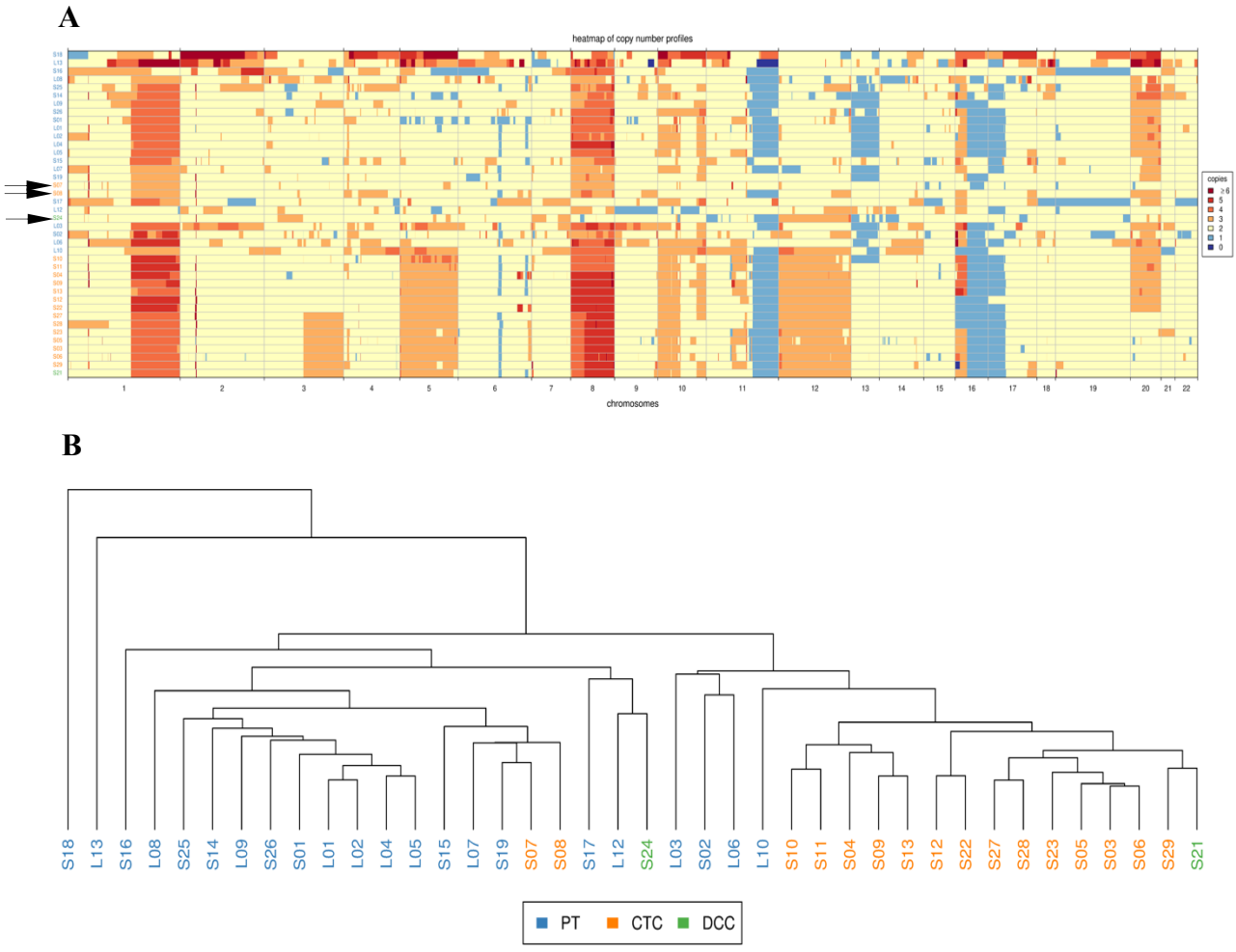


Figure 36: Hierarchical clustering of cells with CNA data. A: Heatmap of CNAs, X axis shows the chromosome numbers, samples are denoted on the Y axis, CNAs are color coded for various intensities according to the legend, where deep blue indicates loss of 2 copies of a chromosomal locus and deep red indicates gain of >6 copies**.** B. Dendogram from the heatmap in A; blue, PT cell; orange, CTC; green; DCC. Sample ordering and dendrogram are derived from complete linkage hierarchical clustering using Euclidean distance