

Aus der Sektion für Translationale Chirurgische Onkologie und  
Biomaterialbanken

Leiter: Prof. Dr. Dr. med. Jens K. Habermann

Der Klinik für Chirurgie

Direktor: Prof. Dr. med. Tobias Keck

Der Universität zu Lübeck

**Identification of tumor associated DNA-  
mutations in breast cancer for therapy  
guidance and prognostication**

Inauguraldissertation

zur Erlangung der Doktorwürde

der Universität zu Lübeck

- Aus der Sektion Medizin –

vorgelegt von

Johanna Oltmann

aus Wilhelmshaven

Lübeck 2019

**1. Berichterstatter: Prof. Dr. Dr. med. Jens K. Habermann**

**2. Berichterstatterin: Prof. Dr. med. Jutta Kirfel**

**Tag der mündlichen Prüfung: 26.08.2020**

**Zum Druck genehmigt. Lübeck, den 26.08.2020**

**- Promotionskommission der Sektion Medizin -**

## Abkürzungsverzeichnis

2c	Diploid DNA content
ALH	Atypical Lobular Hyperplasia
ASCO	American Society of Clinical Oncology
bp	Base pair
CCVD	Cardio-Cerebro-Vascular Disease
CDK	Cyclin-Dependent Kinase
CEP	Chromosome Enumeration Probe
DNA	Deoxyribonucleic Acid
DAPI	4',6-Diamidino-2-Phenylindole
DCIS	Ductal Carcinoma in Situ
dUTP	Deoxyuridine Triphosphate
EDTA	Ethylenediaminetetraacetic Acid
EGFR	Epithelial Growth Factor Receptor
EMT	Epithelial Mesenchymal Transition
ER	Estrogen Receptor
FFPE	Formalin-Fixed and Paraffin-Embedded
Fig.	Figure
FISH	Fluorescence <i>in situ</i> Hybridization
G1-phase	Gap 1, intervall between mitosis and S-phase
G2/M-phase	Gap 2, interval between S-phase and mitosis/ mitosis
H&E	Hematoxylin and Eosin stain
HCl	Hydrochloric acid
IDC	Invasive Ductal Carcinoma
ILC	Invasive Lobular Carcinoma
LB	Lysogeny Broth
LCIS	Lobular Carcinoma in Situ
LN	Lobular Neoplasia
M	Molar
n	Number
NBF	Formalin, 10%, Neutral phosphate Buffer
NCBI	National Center for Biotechnology Information
PBS	Phosphate Buffered Saline
PG	Prostaglandin
PR	Progesterone Receptor
S-phase	Synthesis
SNB	Sentinel Node Biopsy
SSC	Saline-Sodium Citrate
UICC	Union Internationale Contre le Cancer
VEGF	Vascular Endothelial Growth Factor

# Index

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background	1
1.2	Epidemiology, etiology and prognosis of breast cancer	1
1.3	Clinical and histopathological classification	2
1.4	Treatment	6
1.5	Prognostic limitations of clinically used parameters	8
1.6	Molecular changes in cancer cells	9
1.7	Genes with frequent copy number changes in breast cancer	9
1.8	The impact of tumor cytogenetics on prognosis and current stratified therapies	16
1.9	Aims	17
1.9.1	Specific objective	18
<b>2</b>	<b>Materials</b>	<b>19</b>
2.1	Clinical samples	19
2.2	Laboratory equipment	19
2.3	Consumables	20
2.4	Chemicals and dyes	21
2.5	BAC clones	23
<b>3</b>	<b>Methods</b>	<b>26</b>
3.1	Nuclear DNA content measurement	26
3.2	Fluorescence <i>in situ</i> hybridization	26
3.2.1	Preparation of Cytospins from FFPE Samples	27
3.2.2	DNA extraction of BAC clones	27
3.2.3	NICK Translation	28
3.2.4	DNA Precipitation	30
3.2.5	Hybridization	30
3.2.6	Detection and Imaging	31
3.3	Imaging Data Files	32
3.4	Signal pattern and instability index	33
3.5	Gene gains and losses	33
3.6	FISHtrees	34
3.7	Targeted next generation sequencing	35
3.7.1	DNA extraction from archived FFPE Specimen	35
3.7.2	Sequencing	36
3.8	Statistical analysis	37

<b>4</b>	<b>Results</b>	<b>38</b>
4.1	DNA ploidy	38
4.2	Hybridization Panel I	38
4.3	Hybridization Panel II	38
4.4	Gene gains and losses	39
4.5	Major signal pattern	40
4.6	Case studies	41
4.7	Summarized analysis of clinical parameters	59
4.8	Correlation between clinical parameters and instability index	60
4.9	FISHtrees	60
4.10	Targeted Sequencing Analysis	63
<b>5</b>	<b>Discussion</b>	<b>65</b>
5.1	Hybridization Panel I	65
5.2	Hybridization Panel II	66
5.3	Intra-tumor heterogeneity	67
5.4	Clinical parameters	70
5.5	Gene Mutations	71
5.6	Conclusion	72
5.7	Outlook	72
<b>6</b>	<b>Summary</b>	<b>74</b>
<b>7</b>	<b>Reference list</b>	<b>75</b>
<b>8</b>	<b>Appendix</b>	<b>84</b>
8.1	List of tables and figures	84
8.2	Preparations	90
<b>9</b>	<b>Acknowledgement</b>	<b>92</b>
<b>10</b>	<b>Curriculum Vitae</b>	<b>92</b>

# **1 Introduction**

## **1.1 Background**

The clinical course of breast cancer and disease-free survival times widely vary from one patient to another.<sup>1</sup> Clinical and histological parameters, the aggressiveness of individual tumors, and the response of patients to systemic therapy are more variable in breast cancer than in many other types of cancer.<sup>2</sup> Currently, the choice of therapy relies on clinical parameters as well as histological and molecular tumor features. However, even at an early stage, tumor-intrinsic genetic alterations affect tumor growth, progression and metastatic potential and therefore limit the value of commonly used prognostic markers. In many cases, the current standard of care does not correctly assess a patient's prognosis.<sup>3</sup> Therefore, additional predictors of disease outcome would be valuable for treatment stratification.

Extensive studies have shown that the degree of variation of the nuclear DNA content determines prognosis.<sup>4</sup> The aim of our study is to further elucidate the molecular basis of aneuploidy in breast carcinomas in order to improve existing prognostic tools.

## **1.2 Epidemiology, etiology and prognosis of breast cancer**

Breast cancer is by far the most common cancer among women worldwide. In 2012, 1.67 million women were diagnosed with breast cancer and there were 6.3 million women alive who had been diagnosed with breast cancer in the previous five years. Breast cancer is also the number one cause for neoplastic deaths among women with 521,900 deaths in 2012.<sup>5</sup>

The highest incidence is found in the group of women between 60 and 64 years. With a mean age of 62 at diagnosis, breast cancer has a peak incidence seven years earlier than all other cancers combined.<sup>6</sup> Breast cancer also occurs in men but is about 100 times less common than in women.<sup>7</sup>

A number of factors associated with increased risk of developing breast cancer have been identified in epidemiologic studies and several of these are being used in clinical practice today to assess a woman's risk of developing this type of cancer. They include age and sex, family history of breast cancer, pathogenic mutations in, among others, the *BRCA1* or *BRCA2* gene, reproductive factors such as young age at onset of menses or late age at menopause, as well as lifestyle and environmental risk factors, such as obesity and exposure to radiation.<sup>8,9</sup>

The overall relative five-year-survival rate for breast cancer is about 89%, but ranges from almost 100% at early stage (UICC I; UICC, French: "Union Internationale Contre le Cancer") to as low as 25% for late stage (UICC IV).<sup>1</sup> However, due to intra-tumor and inter-individual heterogeneity of breast cancer, the clinical course and disease-free survival times are extremely variable.<sup>3</sup>

### **1.3 Clinical and histopathological classification**

Histological features separate invasive carcinomas, which have penetrated the basement membrane of the tissue and are able to metastasize, from non-invasive carcinomas, which are considered pre-invasive lesions.<sup>10</sup>

Depending on the site of origin, non-invasive carcinomas are either classified as a ductal carcinoma in situ (DCIS) or a lobular neoplasia (LN), the later including both lobular carcinoma in situ (LCIS) and atypical lobular hyperplasia (ALH).

Invasive carcinomas are also characterized by the site of their origin, the most common one being the invasive ductal carcinoma (IDC), which accounts for about 40-75% of breast cancers. Invasive lobular carcinomas (ILC) comprise about 5-15%. The remaining tumor types are mostly mucinous, tubular, papillary, medullary or mixed carcinomas.

Based on morphological features, namely tubule formation, nuclear polymorphism and mitotic counts, carcinomas are graded G1, meaning well differentiated, G2, moderately differentiated, or G3, indicating poor differentiation.<sup>11</sup>

For classification of tumor stage, the internationally recognized TNM-Classification of the UICC in its 7<sup>th</sup> edition is applied. It includes tumor size at the primary tumor site (T0-4), regional lymph node involvement (N0-3), as well as the presence or absence of distant metastasis (M0-1) (see table 1.1 – 1.3).<sup>11</sup>

<b>TX</b> Primary tumor cannot be assessed	<b>T1</b> Tumor $\leq 20$ mm in the greatest dimension	<b>T3</b> Tumor $> 50$ mm in greatest dimension
<b>T0</b> No evidence of primary tumor	<b>T1mi</b> Tumor $\leq 1$ mm in the greatest dimension	<b>T4</b> Tumor of any size with direct extension to the chest wall and/or to the skin (ulceration or skin nodules).
<b>Tis</b> Carcinoma in situ	<b>T1a</b> Tumor $> 1$ mm but $\leq 5$ mm in greatest dimension	<b>T4a</b> Extension to the chest wall, not including only pectoralis muscle adherence/invasion.
<b>Tis (DCIS)</b> Ductal carcinoma in situ	<b>T1b</b> Tumor $> 5$ mm but $\leq 10$ mm in greatest dimension	<b>T4b</b> Ulceration and/or ipsilateral satellite nodules and/or edema (including peau d'orange) of the skin, which do not meet the criteria for inflammatory carcinoma.
<b>Tis (LCIS)</b> Lobular carcinoma in situ	<b>T1c</b> Tumor $> 10$ mm but $\leq 20$ mm in greatest dimension	<b>T4c</b> Both T4a and T4b
<b>Tis (Paget's)</b> Paget's disease of the nipple NOT associated with invasive carcinoma and/or carcinoma in situ (DCIS and/or LCIS) in the underlying breast parenchyma.	<b>T2</b> Tumor $> 20$ mm but $\leq 50$ mm in greatest dimension	<b>T4d</b> Inflammatory carcinoma

**Table 1.1: UICC (Union Internationale Contre le Cancer), 7<sup>th</sup> edition - Classification of primary tumor (T)**



<b>NX</b>	Regional lymph nodes cannot be assessed.
<b>N0</b>	No regional lymph node metastases.
<b>N1</b>	Metastases to movable ipsilateral level I, II axillary lymph node(s).
<b>N2</b>	Metastases in ipsilateral level I, II axillary lymph nodes that are clinically fixed or matted. OR Metastases in clinically detected ipsilateral internal mammary nodes in the absence of clinically evident axillary lymph node metastases.
<b>N3</b>	Metastases in ipsilateral infraclavicular (level III axillary) lymph node(s) with or without level I, II axillary lymph node involvement. OR Metastases in clinically detected ipsilateral internal mammary lymph node(s) with clinically evident level I, II axillary lymph node metastases. OR Metastases in ipsilateral supraclavicular lymph node(s) with or without axillary or internal mammary lymph node involvement.

**Table 1.2: UICC (Union Internationale Contre le Cancer), 7<sup>th</sup> edition - Regional lymph nodes Clinical (cN)**

<b>pNX</b>	Regional lymph nodes cannot be assessed.
<b>pN0</b>	No regional lymph node metastasis identified histologically.
<b>pN1</b>	Micrometastases. OR Metastases in 1-3 axillary lymph nodes. AND/OR Metastases in internal mammary nodes with metastases detected by sentinel lymph node biopsy but not clinically detected.
<b>pN2</b>	Metastases in 4-9 axillary lymph nodes. OR Metastases in clinically detected internal mammary lymph nodes in the absence of axillary lymph node metastases.
<b>pN3</b>	Metastases in $\geq 10$ axillary lymph nodes. OR Metastases in infraclavicular (level III axillary) lymph nodes. OR Metastases in clinically detected ipsilateral internal mammary lymph nodes in the presence of one or more positive level I, II axillary lymph nodes. OR Metastases in $>3$ axillary lymph nodes and in the internal mammary lymph nodes with micrometastases or macrometastases detected by sentinel lymph node biopsy but not clinically detected. OR Metastases in ipsilateral supraclavicular lymph nodes.

**Table 1.3: UICC (Union Internationale Contre le Cancer), 7<sup>th</sup> edition - Regional lymph nodes Pathologic (pN)**

Depending on these three factors, patients are stratified into four tumor stages (table 1.4), which are associated with the patient's prognosis and are also the basis for the selection of available treatment options.

Stage	T	N	M
<b>0</b>	Tis	N0	M0
<b>IA</b>	T1b	N0	M0
<b>IB</b>	T0	N1mi	M0
	T1b	N1mi	M0
<b>IIA</b>	T0	N1c	M0
	T1b	N1c	M0
	T2	N0	M0
<b>IIB</b>	T2	N1	M0
	T3	N0	M0
<b>IIIA</b>	T0	N2	M0
	T1b	N2	M0
	T2	N2	M0
	T3	N1	M0
	T3	N2	M0
<b>IIIB</b>	T4	N0	M0
	T4	N1	M0
	T4	N2	M0
<b>IIIC</b>	Any T	N3	M0
<b>IV</b>	Any T	Any N	M1

**Table 1.4: UICC (Union Internationale Contre le Cancer), 7<sup>th</sup> edition – tumor stage for breast cancer**

Each sample is also analyzed for its hormone receptor status, namely Estrogen and Progesterone receptor status, its *Her2/neu*-status, as well as optionally others, including e.g. the Ki-67 proliferation index.<sup>11–13</sup>

## 1.4 Treatment

Treatment for breast cancer can consist of several different modalities which can be used as sole treatment or as part of a combination therapy. The five standard treatment options are surgery, radiation therapy, chemotherapy, hormone therapy and targeted therapy. In addition to these, new therapies are constantly tested in clinical trials and may be available for suitable patients.<sup>12</sup>

One part of treatment is surgery, the extent of which depends on tumor size and stage. The type of surgery can vary between local excision and radical mastectomy. By sentinel-node biopsy (SNB), the lymph node or nodes, that are most likely to be reached by metastasizing cancer cells, are examined. Depending on whether the histology shows infiltration or not, the need for and extent of lymph node dissection can be decided upon.

If a tumor is deemed primarily inoperable, neoadjuvant systemic chemotherapy is a treatment option and can lead to histopathologic remission, which in turn makes it possible to surgically remove remaining tumor tissue.<sup>12</sup>

Radiation therapy is generally used postoperatively, if necessary, and decreases the chance of local recurrences. It can also reduce mortality.<sup>12,14</sup>

Another part of the treatment is adjuvant systemic therapy: either as chemotherapy, hormone therapy or targeted therapy. If the tumor cells express hormone receptors, endocrine therapy will be applied. Among the most common substances used for systemic therapy are aromatase inhibitors, such as Anastrozol, and estrogen receptor antagonists, such as Tamoxifen. Some tumor cells also show overexpression of certain other receptors, most importantly the *Her2/neu*-receptor. In this case, a targeted therapy with monoclonal *Her2/neu*-receptor-antibodies, such as Trastuzumab, can be effective.<sup>12,15–17</sup>

The individual treatment regimen for each patient depends on the stage of the disease at diagnosis. Common treatment for DCIS was, until recently, mastectomy but nowadays breast-conserving surgery via lumpectomy is more often the number one choice for this non-invasive lesion. After surgery, the patient undergoes radiation therapy, Tamoxifen can be administered additionally.<sup>12</sup>

For stages I, II, IIIA and operable IIIC there are several options. A loco-regional treatment, which can consist of breast-conserving surgery or radical mastectomy, is combined with evaluation of lymph node infiltration. For axillary node-positive tumors, adjuvant radiation therapy is often recommended. Also depending on lymph node status, adjuvant systemic treatment may be appropriate, meaning chemotherapy, which can be combined with Tamoxifen.<sup>12,17</sup>

Patients with primarily inoperable stage IIIB or IIIC, or inflammatory breast cancer, are usually treated with a multimodality therapy with curative intent, consisting of

neoadjuvant systemic chemotherapy, followed by surgery and postoperative radiation as well as subsequent systemic therapy.<sup>12</sup>

In addition to using neoadjuvant chemotherapy to shrink primarily inoperable tumors, it can also be used to shrink the tumor to allow for less extensive surgery. Another big advantage in comparison with adjuvant therapy is the possibility of clinically observing the tumors' response to the treatment and hence the possibility to switch to more effective options, if one regime does not work.<sup>18</sup>

Recurrent breast cancer treatment is rarely curative, however if the recurrence is only local, patients may still have long-term survival with appropriate treatment.<sup>19,20</sup> Stage IV metastatic disease treatment is of palliative nature, mainly focusing on improvement in quality of life and its prolongation. It generally involves systemic chemotherapy and/or hormone therapy, with or without Trastuzumab. Surgery is mostly performed to reduce pain or lessen other symptoms that are due to metastases, including for example decompression of the spinal cord because of pathologic vertebral fractures or extirpation of lung metastasis in a patient with pleural effusions. Radiation therapy may also be a useful alternative in these cases.<sup>12,21</sup>

## **1.5 Prognostic limitations of clinically used parameters**

Breast cancer is often diagnosed at early stages. In the United States, almost 61% of patients are diagnosed at localized stage, meaning that their cancer is confined to its primary site. 32% of patients are diagnosed with a regional disease, where the cancer has spread to regional lymph nodes. Only 5% of patients are diagnosed after they already developed metastases.<sup>1</sup>

The variety of clinical and histological parameters used to assess a patient's risk of disease progression, its aggressiveness as well as its response to systemic therapy is rather broad compared to other tumor entities and due to effective treatment options, breast cancer is potentially curable.<sup>22</sup>

However, even at an early stage molecular alterations affect tumor growth, progression and metastatic potential and therefore limit the prognostic value of the aforementioned parameters. Thus, in many cases it is impossible to correctly

assess a patient's prognosis and predict the treatment the patient would benefit from most.<sup>3</sup>

## **1.6 Molecular changes in cancer cells**

One important feature of malignantly transformed cells are so-called copy number changes. These changes are structural abnormalities of chromosomes or parts thereof, results of deletions and multiplications of chromosome regions. They lead to certain genes being either lost or gained. Two of the most important changes involve oncogenes, which are frequently gained in cancer, and tumor suppressor genes, which normally inhibit immoderate cell proliferation but are frequently lost or inhibited in tumor cells.<sup>23</sup> Such alterations can cause increased errors during DNA replication in the cell cycle and thus endanger the genetic stability of a cell, eventually causing genomic instability and aneuploidy.

If the acquired mutations confer advantages over non-mutated cells, clones of the mutated cells consequently become dominant within a tumor population, a process called clonal selection. The results are increased proliferation, a higher likelihood of survival and, subsequently, invasion and metastasis.<sup>23</sup>

Fluorescence-in-situ-hybridization allows analysis of these changes by illustrating the gains and losses of specific genes as well as gain and loss patterns, and visualizing the extent of intra-tumor heterogeneity.

## **1.7 Genes with frequent copy number changes in breast cancer**

Eight genes with frequent copy number changes in breast cancer were selected for this study. They include the five oncogenes *COX2*, *MYC*, *HER2*, *CCND1*, and *ZNF217*, as well as the three tumor suppressor genes *DBC2*, *CDH1*, and *TP53*.

All eight genes are already well investigated and were chosen because of their importance in breast cancer according to (i) the literature and (ii) studies conducted by Prof. Ried's laboratory.<sup>24</sup> These genes showed to be promising candidates to visualize copy number changes in breast cancer samples to further

understand tumor heterogeneity, pattern of clonal carcinogenesis and their implications for individual prognostication and treatment guidance.

## **COX2**

Cyclooxygenase 2, also known as Prostaglandin-endoperoxide synthase (PTGS), is the enzyme encoded for by *COX2*. *COX2* is located on chromosome 1q31.1. The enzyme is an inducible prostaglandin synthase, which catalyzes the synthesis of prostaglandins (PGs) G<sub>2</sub> and H<sub>2</sub> from arachidonic acid. PGH<sub>2</sub> is then converted to other prostaglandins, thromboxane and prostacyclin. These have physiological functions, for example in hemostasis or pain reactions. *COX2* is normally expressed in parts of the kidney and the brain but has been found to be overexpressed in multiple human cancers, including the breast. It could be shown by Liu *et al.* that in multiparous mice, the overexpression of *COX2* alone was sufficient to cause breast tumors.<sup>25</sup> In response to growth factors, tumor promoters, cytokines and oncogenes, including the *Her2/neu*-gene, *COX2* expression is up-regulated, hence leading to increased prostaglandin levels in inflamed and neoplastic tissues.<sup>26</sup>

Cyclooxygenase 2 contributes to tumor formation in various ways. Through increased PG levels, it stimulates cell proliferation and mitogenesis of mammary epithelial cells.<sup>27</sup> On the other hand, it inhibits T- and B-cell proliferation, cytokine synthesis and reduces the cytotoxic activity of natural killer cells, causing immune suppression and therefore possibly diminishes the immune systems' ability to detect a growing tumor.<sup>28</sup> PGE<sub>2</sub> has been shown to yield an increase in aromatase activity, which is the enzyme that produces estrogen, a hormone that is known to play a defining role in the development of breast cancer.<sup>29</sup> PGs can also be altered by enzymes, as well as non-enzymatically, causing a production of mutagens such as Malondialdehyde, an agent that induces frame-shifts and base-pair substitutions, therefore causing DNA damage.<sup>30</sup> Additionally, *COX2* also has proangiogenic effects, including increased production of vascular endothelial growth factor (VEGF).<sup>31</sup> Angiogenesis is essential for tumors if they grow beyond a certain size to allow the cells to be sustained. In addition, *COX2* is thought to suppress apoptosis, one of the most important mechanisms in tumorigenesis.

It favors the survival of cells that have been altered by mutations, making them more resistant than those cells that remain unaltered. Thus, suppressing apoptosis supports selection of mutated cells for survival and further proliferation.<sup>26</sup> Lastly, *COX2* also appears to promote metastasis.<sup>32</sup>

The fact that *COX2* is differentially expressed in many malignant tumors, especially in breast cancer, could be verified in various studies.<sup>33–38</sup> Whether its expression is a predictor of poor prognosis remains to be elucidated, since previous data have so far been conflicting. While some studies report no correlation between *COX2* expression and outcome<sup>35,36</sup>, others found that its up-regulation is indeed associated with poor prognosis.<sup>33,37,38</sup> Its expression is, however, frequently associated with unfavorable tumor subtypes, such as triple-negative tumors, as well as grade III tumors.<sup>33,34,38–41</sup>

### ***DBC2***

*DBC2* (deleted in breast cancer), also known as *RhoBTB2*, is located on chromosome 8p21.3. The gene encodes a small Rho GTPase and has been identified as a tumor-suppressor gene. It is homozygously deleted in 3.5% and homozygously mutated in an additional 10% of breast cancers. In about 60%, *DBC2* expression is lost completely.<sup>42,43</sup>

*DBC2* function is involved in the regulation of cell growth by controlling cell-cycle and apoptosis. Correspondingly, when overexpressed in breast tumor cells *in vitro*, an inhibition of growth could be observed. Overexpression also led to decreased ability of forming cell colonies, a quality essential to tumor cells, as well as increased apoptosis.<sup>44</sup> Mao *et al.* could show that patients with *RhoBTB2*-negative breast cancer had poor clinical outcome and that *DBC2* silencing was an independent risk factor for breast cancer.<sup>43</sup>

### ***MYC***

*MYC* (*cMYC*), its official name being v-myc avian myelocytomatosis viral oncogene homolog, is located on chromosome 8q24.21. *MYC* codes for a multifunctional, nuclear phosphoprotein that acts as a transcription factor,



regulating the transcription of various target genes. It is overexpressed and somatically amplified in several human tumor types, both in pre-invasive and invasive lesions, and is one of the most important somatically mutated proto-oncogenes known in human cancers. Reversely, its expression is also driven by several oncogenes.<sup>45,46</sup>

*MYC*'s role in cancer development has been intensely studied for many years. It stimulates cell proliferation through an extenuated need for growth factors, blockade of exit from cell cycle, acceleration of cell division and increase of cell size.<sup>47-51</sup> *MYC* also promotes cell survival, genetic instability and angiogenesis, while inhibiting cell differentiation and cell adhesion, via activation and repression of multiple genes, thus fostering tumorigenesis and metastasis.<sup>47,52</sup> It may also have direct control on cellular invasion and migration, and hence promote metastasis. For example, it regulates the epithelial-mesenchymal transition (EMT), enabling epithelial cells to transform into a mesenchymal state, which is required for invasion and, in succession, metastasis.<sup>46,53</sup> Wolfer *et al.* found *MYC* to be the regulator of a 13 “poor-prognosis” gene expression signature in breast cancer cells.<sup>54</sup>

### **CCND1**

*CCND1* is located on chromosome 11q13.3 and codes for Cyclin D1, the regulatory subunit of a holoenzyme that plays a pivotal role in the regulation of the cell cycle. *CCND1* is amplified in 15% and overexpressed in 30-50% of primary breast cancers. Its overexpression is one of the most common genomic changes in human cancers overall. In most cancer types, including the breast, the increased expression is due to induction by oncogenic signals, rather than a clonal somatic mutation or rearrangement in the gene.<sup>55,56</sup>

Cyclin D1 is the regulatory subunit of a holoenzyme that heterodimerizes with cyclin-dependent kinases (CDKs). The holoenzyme phosphorylates and, together with sequential phosphorylation by cyclin E/CDK2, inactivates the retinoblastoma protein (pRb). pRb can be seen as a gatekeeper of the G<sub>1</sub>-phase of the cell-cycle, inhibiting DNA synthesis.<sup>57,58</sup> Hence, cyclin D1 stimulates progression through the

G<sub>1</sub>-S-phase of the cell cycle and therefore leads to cell proliferation. It integrates extracellular signals by coupling signals from cell surface receptors to transcription factors, thereby regulating diverse gene expression networks. The ability of cyclin D1 to differentially regulate nuclear receptor activity can lead to distinct transcription outputs, depending on the transcription factor. For example, binding to the estrogen receptor  $\alpha$  (ER $\alpha$ ) enhances ligand-independent gene activity, whereas liganded androgen receptor activity is inhibited by cyclin D1.<sup>59</sup> In addition, it regulates cellular metabolism, fat cell differentiation and cellular migration.<sup>55</sup>

The overexpression of *cyclin D1* has been shown to correlate with early onset of cancer as well as risk of tumor progression and metastasis.<sup>60,61</sup> However, the prognostic value is still unclear. While some studies found that *cyclin D1* overexpression correlates with ER-positivity, well-differentiated carcinomas and a favorable clinical outcome, as well as better response to Tamoxifen in those breast cancers that are ER-positive<sup>62–65</sup>, others report conflicting results and see high *cyclin D1* expression levels as a predictor of unfavorable prognosis.<sup>66,67</sup>

### ***CDH1***

*CDH1* is located on chromosome 16q22.1. It encodes cadherin 1, also known as epithelial (E)-cadherin 1, a calcium-dependent cell-cell adhesion glycoprotein.<sup>68,69</sup> Loss of E-cadherin expression has been reported in 85% of ILCs and LCIS, whereas ductal histology often presents with varying levels of expression.<sup>70,71</sup>

E-cadherin is a transmembranic cell adhesion molecule, consisting of a cytoplasmatic and an extracellular domain, of which the latter is released in the presence of calcium-ions. The cytoplasmatic domains are associated with a group of proteins, named catenins, which, in turn, make contact with the microfilament network.<sup>72</sup> Any significant change in expression or structure of one of these components can lead to junctional disassembly and, consequently, can result in more mobile invasive carcinoma cells.<sup>69</sup> Conversely, E-cadherin expression correlates with epithelial differentiation and the prevention of invasiveness of particular carcinomas.<sup>73</sup> Cell lines expressing E-cadherin were shown to be of epitheloid morphology and generally non-invasive, whereas cell lines that did not

express E-cadherin were often of a fibroblastoid morphology and invasive.<sup>74</sup>

Reduced expression of E-cadherin correlates with loss of tumor differentiation and progression to lymph node and distant metastases, and heterogeneity of E-cadherin expression is associated with poor prognosis.<sup>69,75</sup>

## ***HER2***

*HER2*, human epidermal growth factor receptor 2, also known as *ERBB2* or *NEU*, is located on chromosome 17q12 and encodes a tyrosine kinase receptor.<sup>76</sup> Due to gene amplification, *HER2* is overexpressed in about 30% of breast carcinomas.<sup>77</sup>

The gene encodes a transmembrane tyrosine kinase receptor with extensive homology to the epidermal growth factor receptor (EGFR). It does not have a ligand binding domain of its own but forms heterodimers with other ligand-bound EGFR family members, which stabilizes ligand binding and enhances kinase-mediated activation of various downstream signaling pathways, including cell proliferation signals.<sup>76</sup> Some of these pathways are involved in cell proliferation, enabling the tumor cells to fix oncogenic mutations by clonal expansion.<sup>78</sup>

Overexpression of *HER2* correlates with tumor size, spread of the tumor to lymph nodes, high grade, high percentage of S-phase cells, aneuploidy and lack of steroid receptors, as well as reduced time to relapse and overall survival.<sup>77,79</sup> The amplification of *HER2* is already tested in clinical practice today for every breast cancer patient, using FISH or IHC, in order to detect those patients at high risk for recurrence and disease-related death. *HER2* positive patients often benefit from Trastuzumab therapy.<sup>12,78</sup>

## ***TP53***

Tumor protein p53 is located on chromosome 17p13.1.<sup>80</sup> *TP53* has been described as the “guardian of the genome”.<sup>81</sup> This tumor suppressor is often mutated and its activity frequently lost in a variety of human cancers.<sup>82</sup>

The gene encodes a tumor suppressor protein containing transcriptional activation, DNA binding, and oligomerization domains. It responds to diverse cellular stresses by regulating the expression of target genes, inducing DNA repair, cell cycle arrest, senescence, or apoptosis.<sup>80</sup> Together, these functions prevent errors in the duplication process of a cell that is under stress, and as such the p53 pathway increases the fidelity of cell division and prevents cancers from arising.<sup>83</sup> Recently, p53 has also been found to play a role in invasion and motility,<sup>84</sup> as well as angiogenesis.<sup>85</sup>

### ***ZNF217***

Zinc finger protein 217 is a strong candidate oncogene located on chromosome 20q13.2, a region highly amplified in about 7-18% of breast tumors.<sup>86,87</sup>

*ZNF217* encodes a transcription factor with eight C2H2 Krüppel-like DNA-binding motifs and a proline-rich transcription activator domain, that associates with proteins involved in transcriptional repression.<sup>86,88</sup> Overexpression of the gene stimulates cell survival and proliferation, cell migration and invasion. It also promotes EMT, which plays a role in invasiveness of cancer cells.<sup>87</sup> Nonet *et al.*<sup>89</sup> even managed to immortalize human mammary epithelial cells by transducing *ZNF217*.

High levels of *ZNF217* expression are associated with resistance to chemotherapy and deregulated apoptotic signals in breast cancer cells.<sup>88,90</sup> A study by Tanner *et al.*<sup>91</sup> showed that gene amplification was associated with high histological grade, DNA aneuploidy, high S-phase fraction and shorter disease-free survival of node-negative breast cancer patients. In summary, high *ZNF217* amplification in breast tumors indicates poor prognosis.

## 1.8 The impact of tumor cytogenetics on prognosis and current stratified therapies

The fact that increased expression of specific genes plays an important role in the pathogenesis of solid tumors has been known for many years.<sup>92–95</sup> To date, a variety of genes that are frequently gained in human cancers, including cancer of the breast, have been identified. The same holds true for genes that are frequently lost. Heselmeyer-Haddad *et al.* recently published a study using multicolor FISH, in which the probe panels' consisted of the same eight genes that were used in this thesis, to determine the evolution of copy number changes in the transition from DCIS to IDC.<sup>24</sup> In both entities, the oncogenes (*COX2*, *MYC*, *HER2*, *CCND1*, *ZNF217*) were frequently gained, but rarely lost, and the tumor suppressor genes (*DBC2*, *CDH1*, *TP53*) were frequently lost and only rarely gained.<sup>24</sup>

In the past years, several groups have made intensive efforts to attain gene expression profiles that can predict clinical outcome in breast cancer patients, independent of clinically used parameters such as tumor size or node status.<sup>11,96–100</sup> Several multigene expression tests have since been devised and gene expression profiling has already been implemented into routine clinical management. Commonly used tests are, e.g., OncotypeDX® and MammaPrint®<sup>101,102</sup>, both of which have been implemented in breast cancer treatment guidelines (e.g. ASCO; S3 Guideline in Germany).<sup>12,99,103,104</sup> Measured in these tests are the expression of 21 and 70 genes, respectively, resulting in a score which can predict the risk of disease recurrence as well as benefit from chemotherapy. The two tests are usually applied to stratify early stage breast cancer.<sup>12,97,103–105</sup> Other gene expression tests such as Prosigna® or EndoPredict® are currently under evaluation in prospective clinical trials.<sup>100,106</sup>

In addition to specific gene expression profiles, extensive studies show that the degree of genomic instability, i.e., aneuploidy, determines disease outcome. In general, patients with genomically stable, diploid tumors have a significantly better prognosis compared to patients whose tumors are genomically unstable and have aneuploid DNA content.<sup>4,107–109</sup> The study groups of Habermann, Ried, and Auer have previously determined a gene expression signature of chromosomal

instability, which recapitulated the expression signatures of both OncotypeDX® and Mammaprint®.<sup>110</sup> They were able to show that nuclear aneuploidy is reflected in these expression profiles and lays the genetic basis for poor prognosis. Yet, knowledge about how crude aneuploidy affects chromosomal gains and losses in individual cells across the tumor population, and to which extent aneuploidy drives intra-tumor heterogeneity is still vague. It is also unclear whether aneuploidy correlates with the gene mutation burden.

## 1.9 Aims

The clinical and histopathological features evaluated by clinicians today do not suffice to adequately estimate the individual risk of breast cancer patients and, sequentially, administer the optimal treatment. Therefore, there are patients with low risk, who receive chemotherapy and radiation they do not need. And there are patients with a high risk, who do not receive a treatment that might stop disease progression or even cure them from their cancer. Hence, the identification of patients at a very low risk of relapse, meaning potentially curable disease with only loco-regional therapy and those at high risk, needing aggressive treatment, is extremely important. Currently, the choice of therapy for breast cancer is based on prognostic and predictive factors, including disease-independent, e.g. age, and disease-related patient characteristics, e.g. tumor size, axillary lymph node status and histological grade, as well as molecular tumor features, such as hormone receptor-, *Her2/neu*, or Ki67-status.<sup>12,111</sup>

Although different guidelines have been developed for selecting breast cancer patients who should receive adjuvant systemic therapy, it still remains a challenge to distinguish those patients who could be spared such a treatment.<sup>16,112,113</sup> Additional specific and sensitive prognostic biomarkers might be helpful to facilitate the challenging situation to accurately define individual risk profiles for patients at the initial diagnosis and to set up an appropriate individualized treatment. Established methods like multicolor interphase fluorescence *in situ* hybridization (miFISH) are suitable to analyze specific promising biomarkers. On the other hand, molecular high-throughput technologies such as Next-Generation-

Sequencing (NGS) also enable both untargeted and targeted approaches to detect or analyze a growing range of biomarkers.<sup>114</sup>

### **1.9.1 Specific objective**

This thesis addresses in particular the following questions:

- Which gains and losses of the described target genes can be found in breast tumors and are there differences between diploid and aneuploid cancers?
- What is the degree of intra-tumor heterogeneity in both groups?
- Is there a correlation between ploidy status and clinical parameters?
- Is the ploidy status correlated with overall survival?
- Is there a difference in the mutation load between the diploid and the aneuploid group?

In order to address these questions, five diploid and six aneuploid breast cancers with extensive clinical annotation, including a 20-year follow-up, were analyzed by use of miFISH to study eight well-characterized genes for their potential as biomarkers to differentiate between breast cancer patients with short-term survival and patients with long-term survival and to analyze the degree of intra-tumor heterogeneity. On the same breast cancer specimen, Next-Generation-Sequencing was applied to analyze the mutations of 563 genes most commonly found to be differentially expressed in human solid carcinomas.

## 2 Materials

### 2.1 Clinical samples

Samples were chosen from a cohort of 5618 breast cancer patients with a minimum follow up of 20 years, who were treated between 1986 and 2001 in Stockholm, Sweden and enrolled in the Stockholm county epidemiological registry. All tumors were histologically confirmed after surgery. Five diploid and six aneuploid samples were selected based on DNA content measurements. No additional selection criteria were applied.

Clinical data were collected during treatment and follow-up. Table 2 in the appendix includes the detailed collected clinical data.

Use of samples and data for this study was approved by the regional ethics committee in Stockholm, Sweden Dnr 2013/707-31/3 as well as by the Office of Human Subjects Research Protection (OHSRP) at the National Institutes of Health exempt #12758.

### 2.2 Laboratory equipment

<b>Equipment</b>	<b>Manufacturer</b>
Automated Fluorescence Microscope, BX63	Olympus, Tokyo, Japan
BioView® Software	BioView, Rehovot, Israel
Centrifuge VWR Scientific Model V	VWR International, West Chester, PA, USA
Centrifuge Eppendorf 5415 D	Eppendorf AG, Hamburg, Germany
Freezer (-86°C)	Sanyo Electric Co., Ltd., Japan
Fridge (4°C)	Sanyo Electric Co., Ltd., Japan



Illumina NextSeq 500 Sequencers	Illumina, San Diego, CA, USA
Incubator (37°C)	Fisher Scientific, Pittsburgh, PA, USA
Hoefer HE33 Mini Horizontal Agarose Electrophoresis Unit	Hoefer, Inc., Holliston, MA, USA
Mettler Toledo PM 300 Balance	Mettler Scientific, Brentwood, NH, USA
Microwave	KitchenAid, Benton Harbor, MI, USA
NanoDrop®	Thermo Fisher Scientific, Pittsburgh, PA, USA
Optical filters	Chroma, Bellow Falls, VT, USA
Pipetman® Classic Pipettes	Gilson, Inc., Middleton, WI, USA
Pipette filler Cellmate II	Fisher Scientific, Pittsburgh, PA, USA
Phase Contrast Microscope	Carl Zeiss, Oberkochen, Germany
Thermo Shandon Cytospin ® 3	Thermo Fisher Scientific, Grand Island, NY, USA
Thermomixer Eppendorf 5436	Eppendorf AG, Hamburg, Germany
ThermoMixer D Eppendorf	Eppendorf AG, Hamburg, Germany
Thermomixer R Eppendorf	Eppendorf AG, Hamburg, Germany
Water baths (37°C, 48°C, 80°C)	Thermo Fisher Scientific, Grand Island, NY, USA

### 2.3 Consumables

<b>Material</b>	<b>Manufacturer</b>
Bard-Parker® Protected Disposable Scalpels	Fisher Scientific, Pittsburgh, PA, USA
Corning® Centrifuge tubes 15mL, 50mL	Corning, Inc., Corning, NY, USA
Eppendorf Safe-Lock Tubes	Eppendorf AG, Hamburg, Germany

Marabu Fixogum Rubber Cement	Marabu GmbH & Co. KG, Tamm, Germany
Parafilm M	Bemis Company, Inc., Oshkosh, WI, USA
Pipette tips 10ul, 20ul, 100ul, 200ul, 1000ul	Neptune Scientific, San Diego, CA, USA
Thermo Microscope Slides Superfrost® Plus	Thermo Fisher Scientific, Grand Island, NY, USA
VWR Micro Coverglass	VWR International, West Chester, PA, USA

## 2.4 Chemicals and dyes

Reagent	Manufacturer
Absolute Ethyl Alcohol, 200 proof	The Warner Graham Co., Cockeysville, MD, USA
Agarose	Invitrogen, Carlsbad, CA, USA
β-Mercaptoethanol, 99%	Sigma-Aldrich, St. Louis, MS, USA
Bovine Serum Albumine (BSA)	Roche, Indianapolis, IN, USA
CEP© 4	Abbott, Abbott Park, IL, USA
CEP© 10	Abbott, Abbott Park, IL, USA
Chloramphenicol (20mg/ml)	Sigma-Aldrich, St. Louis, MS, USA
DAPI-sulforhodamine Solution	Sigma-Aldrich, St. Louis, MS, USA
dATP, dCTP, dGTP, dTTP	Boehringer, Mannheim, Germany
DNA Marker II	Invitrogen, Carlsbad, CA, USA
DNase I (bovine pancreas) 100mg	Boehringer, Mannheim, Germany
DNeasy® Blood & Tissue kit	Qiagen, Germantown, MD, USA
dUTPs	Boehringer, Mannheim, Germany

EDTA 0.5M	Quality Biologicals, Gaithersburg, MD, USA
Ethydiumbromide	Thermo Fisher, Waltham, MA, USA
Deionized Formamide	Thermo Fisher, Waltham, MA, USA
Dextran Sulfate (50%)	Sigma-Aldrich, St. Louis, MS, USA
Formalin, 10%, Neutral Phosphate Buffer (NBF)	Polysciences, Inc., Warrington, PA, USA
Glycerol	Invitrogen, Carlsbad, CA, USA
HCl 1M	Thermo Fisher, Waltham, MA, USA
Human Cot-1© DNA (1mg/dl)	Invitrogen, Carlsbad, CA, USA
Isopropanol (2-propanol)	Sigma-Aldrich, St. Louis, MS, USA
Kanamycin	Sigma-Aldrich, St. Louis, MS, USA
Loading dye solution	Thermo Fisher, Waltham, MA, USA
Lysogeny Broth (LB)	KD Medical, Columbia, MD, USA
Magnesium Chloride 0.5M	Quality Biologicals, Gaithersburg, MD, USA
100bp Marker	Invitrogen, Carlsbad, CA, USA
Mineral Oil	Northeast Laboratory, Waterville, ME, USA
Natrium-Acetate 3M (pH=5.2)	Quality Biologicals, Gaithersburg, MD, USA
Natrium-Chloride 1M	Quality Biologicals, Gaithersburg, MD, USA
Pepsin	Sigma-Aldrich, St. Louis, MS, USA
PBS 1X	KD Medical, Columbia, MD, USA
Plasmid Maxi Kit	Qiagen, Germantown, MD, USA
Polymerase (Kornberg)	Boehringer, Mannheim, Germany
Protease Type XXIV, Bacterial	Sigma-Aldrich, St. Louis, MS, USA

RNAse A	Qiagen, Germantown, MD, USA
Salmon Testes DNA	Sigma-Aldrich, St. Louis, MS, USA
20X SSC	Roche, Basel, Switzerland
Sulphorodamine Solution	Sigma-Aldrich, St. Louis, MS, USA
Tris-HCl 1M (pH 8.0)	Quality Biologicals, Gaithersburg, MD, USA
Vectashield Fluorescence Mounting Medium	Vector Laboratories, Inc., Youngstown, OH, USA
Xylene	Sigma-Aldrich, St. Louis, MS, USA

## 2.5 BAC clones

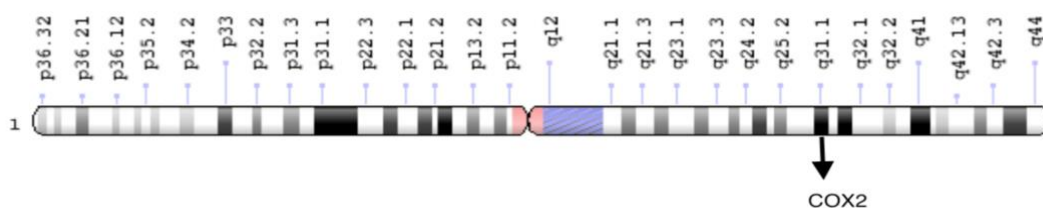
The following BAC (Bacterial Artificial Chromosome) clones were chosen for DNA extractions. All BAC clones were ordered from BACPAC Resources, Oakland, CA, USA.

Probe panel	Centromere probe	Gene probe		
		Gene	Location	BAC clones
I	CCP 10	ZNF217	20q13.2	RP5-823G15 RP4-724E16 CTD-2573N1
		HER2	17q12	RP11-94L15 CTD-3211L18 CTD-2248E4
		TP53	17p13.1	RP11-404G1 RP11-199F11 RP11-186B7
		CDH1	16q22.1	RP11-615I2 RP11-354M1 RP11-354N7
II	CCP 4	MYC	8q24.21	RP11-1136L8 CTD-3056O22 RP11-55J15 RP11-709E21
		COX2	1q31.1	RP11-1149C23 CTD-2509N15 RP11-809N5
		DBC2	8p21.3	RP11-244I22 RP11-875O11 RP11-109B10
		CCND1	11q13.3	CTD-2507F7 RP11-300I6 RP11-186D19

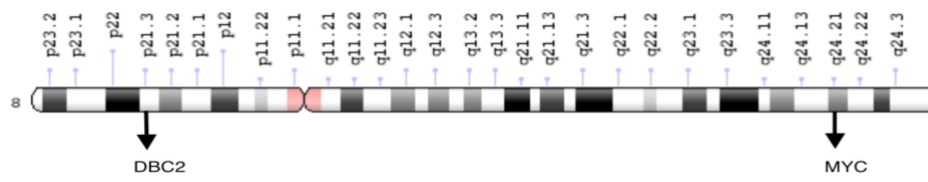
**Table 3: Detailed list of gene probes and BAC (Bacterial Artificial Chromosome) clones**

The following images are modified from the Genome Decorations Page/ NCBI <sup>115</sup> and visualize the locations of each of the eight genes on the chromosome.

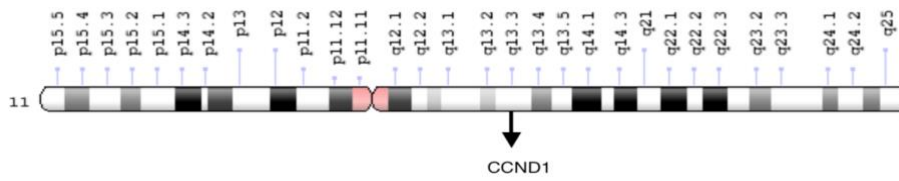
Chromosome 1:



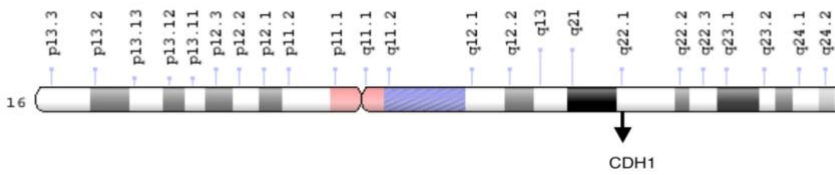
Chromosome 8:



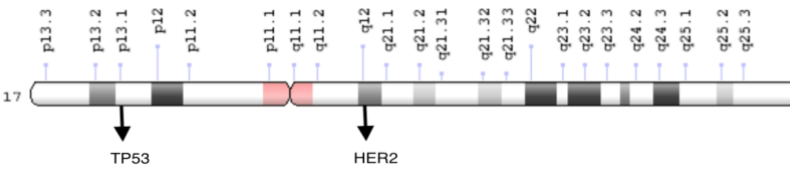
Chromosome 11:



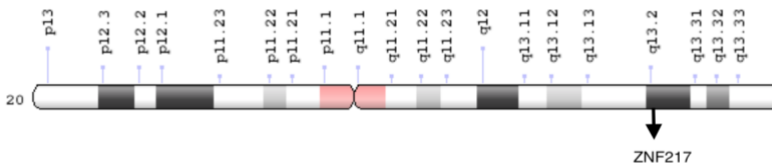
Chromosome 16:



Chromosome 17:



Chromosome 20:



**Fig. 1: Gene models.** Simplified gene models with indicated localization of genes analyzed for this study, modified from the Genome Decorations Page/ NCBI <sup>115</sup>

### **3 Methods**

The laboratory work for this study was in major parts carried out during a research semester in the laboratory of Prof. Dr. Thomas Ried, M.D at the National Cancer Institute (NCI), National Institutes of Health (NIH) in Bethesda, MD, USA.

#### **3.1 Nuclear DNA content measurement**

To measure the DNA content, each tumor sample was analyzed at the Karolinska Institute by DNA cytometry, which converts the extinction coefficient of stained cells into a ploidy value. Imprints were stained using Feulgen-technique (Feulgen und Rosenberg, 1924).

DNA content of the nuclei on the slides was measured using static image cytometry analysis, which converted the computer-aided extinction coefficient of the stained cells into a ploidy value.<sup>109</sup> Lymphocytes with a ploidy value of 2c, meaning diploid, were used as reference. Generated DNA-profiles were then classified according to Auer *et al.*<sup>107</sup>

#### **3.2 Fluorescence *in situ* hybridization**

All analysis was performed on Formalin-Fixed and Paraffin-Embedded (FFPE) specimen. Every FFPE sample was cut into six sections of the following thicknesses: 4  $\mu\text{m}$ , 6  $\mu\text{m}$ , 6  $\mu\text{m}$ , 50  $\mu\text{m}$ , 50  $\mu\text{m}$ , 4  $\mu\text{m}$ . The 4  $\mu\text{m}$  sections were stained with H&E and histomorphologic features were assessed by a pathologist (A. Höög, Karolinska Institute, Stockholm, Sweden) to identify tumorous regions. The 6  $\mu\text{m}$  sections were stored for potential FISH and immunohistochemical studies. For the FISH and sequencing analysis for this study, both 50  $\mu\text{m}$  sections were used.

### 3.2.1 Preparation of Cytospins from FFPE Samples

Using the 4 µm H&E-stained samples, the pathologist outlined the tumor area in these sections. The tissue in the marked area was removed with a scalpel. It was placed in an Eppendorf tube with xylene for 20 minutes and then centrifuged at 750 x g for 3 minutes. The xylene was removed, and the xylene wash repeated twice. Then the tissue was rehydrated in an ethanol series of 100% (2 x), 90%, 70% and 50%. The sample was then centrifuged at 16,000 x g at 4°C for 15 minutes. After removal of the ethanol, we added 1 mL of sterile water and let the sample sit for 20 minutes. It was again centrifuged at 16,000 x g at 4°C for 15 minutes. The water was removed and 500 µL of 0.1 % protease in 1 X PBS was added. Then, the sample was placed in a Thermomixer at 45°C for 60 minutes. Tissue disintegration was frequently monitored under microscopic control. To evaluate the level of disintegration, 10 µL of the protease digest was placed on a slide and stained with 10 µL of DAPI-sulforhodamine solution. Around 30 nuclei per 16x objective field view, lack of cytoplasm, and a certain intensity of the DAPI-stain were deemed as optimal disintegration. Once this assessment was made, 500 µL of 1 X PBS were added to stop the reaction. The tube was then centrifuged at 530 x g for 5 minutes. The concentration of the nuclei was adjusted to 80 µL. The tube was then centrifuged at 180 x g in a Cytospin centrifuge for 5 minutes to obtain a medium-dense, monolayered cytospin. The cytospins were dehydrated in 70% and 90% ethanol for 5 minutes each and 100% ethanol for 10 minutes. After air-drying, they were stored at 4°C for further work.

### 3.2.2 DNA extraction of BAC clones

Bacterial artificial chromosome contigs centering around the eight genes *COX2* (1q31.1), *DBC2* (8p21.3), *MYC* (8q24.21), *CCND1* (11q13.3), *CDH1* (16q22.1), *HER2* (17q12), *TP53* (17p13.1), and *ZNF217* (20q31.2) were assembled.

To extract the DNA from these BAC clones, they were streaked on an agar plate and incubated upside down at 37°C overnight. The next day a single colony was picked and swirled in a 1:1000 solution of Lysogeny Broth (LB) with an antibiotic to



create a starter culture. For the BAC clones RP5-823G15 and RP4-724E16 of chromosome 20q, the antibiotic was Kanamycin, for all other clones we used Chloramphenicol. Starter cultures were grown for 4-5 hours, shaking at 37°C in a thermomixer. Then, 1ml of the starter culture was added to a 1:1000 solution of LB with the same antibiotic and was grown for 16-18 hours, shaking at 37°C. The overnight culture was spun down at 6,000 x g for 15 minutes at 4°C, the supernatant discarded, and the pellet frozen for a minimum of one night at -80°C.

All following buffers and tips were taken from the Qiagen Plasmid Maxi kit. The pellet was resuspended in 10 mL chilled Buffer P1 and transferred to an Oak Ridge PPCO tube. 10 mL of each Buffer P2 and P3 were added consecutively and mixed with the suspension. After 30 minutes incubation on ice, the tube was centrifuged at 20,000 x g at 4°C for 30 minutes. Qiagen 500 tips were equilibrated by applying 10 mL Buffer QBT each and allowing it to empty by gravity flow. The resulting supernatant from the centrifugation was then applied to a Qiagen 500 tip and allowed to enter the resin by gravity flow. The Qiagen tip was then washed with 30 mL Buffer QC twice. DNA was eluted with 15 mL 65°C-warm Buffer QF and precipitated with 10.5 mL of 2-propanol. After the tube was spun down at 15,000 rpm at 4°C for 30 minutes, the supernatant was removed, and the DNA pellet washed with 5 mL cold 70% ethanol. After centrifugation at 15,000 x g at 4°C for 10 minutes and removal of the supernatant, the pellet was air-dried. The DNA was then re-dissolved in 200 µL dH<sub>2</sub>O and each sample measured with Nanodrop®.

### **3.2.3 NICK Translation**

To obtain color-labeled BAC clone DNA, the DNA had to be treated with DNase I, which cut the double-stranded DNA, resulting in single-stranded so-called “nicks”. Polymerase I then elongated these nicks by inserting color-labeled dUTPs.

For this project, *ZNF217* and *MYC* were labeled in aqua with DY-415-dUTP, *HER2* and *COX2* were labeled in red with DY-590-dUTP, *TP53* and *DBC2* were labeled in gold with DY-547-dUTP, and *CDH1* and *CCND1* were labeled in green with

DY-505-dUTP. Two centromere probes, CEP4 and CEP10, were labeled in far red with DY-651.

Panel	Gene probe red	Gene probe green	Gene probe gold	Gene probe aqua	Centromere probe Cy5
I	<i>HER2</i>	<i>CDH1</i>	<i>TP53</i>	<i>ZNF217</i>	CEP 10
II	<i>COX2</i>	<i>CCND1</i>	<i>DBC2</i>	<i>MYC</i>	CEP 4

**Table 4: Hybridization panels I and II**

For 100  $\mu$ L of nicked DNA, an amount of DNA equivalent to 2  $\mu$ g was added to a 1.5 mL Eppendorf tube. We then added MasterMix and the labeled dUTP Mix, as well as Polymerase and a 1:1000 dilution of DNase I (see Preparations in appendix for exact measurements and ingredients). The tubes were then incubated at 15°C for 2 hours.

To control the length of the resulting DNA strands, each sample was run on a 1% agarose gel with a 100 bp (base pair) DNA ladder as well as molecular weight marker II for reference. The ideal length of gene specific probes was 300-600 bp and 500-900 bp for chromosome probes respectively. If the resulting DNA was too long, additional DNase I was added to the probe and incubated at 15°C for 15-30 minutes, then the probe was run on a gel once again. Once the desirable length was acquired, the nick translation was stopped by adding 1  $\mu$ l of 0.5 M EDTA (pH 5.2) to each Eppendorf tube, which was then incubated at 65°C for 10 minutes.

### 3.2.4 DNA Precipitation

To concentrate the DNA, it was next precipitated in ethanol. For DNA nicked in aqua, 40  $\mu$ l of each NICK probe was aliquoted. For red and green probes, we used 20  $\mu$ l of each, for gold 30  $\mu$ l of each probe. Two-thirds of the total amount of probe DNA was added of Human Cot-1 DNA, as well as 1.5  $\mu$ l of Salmon Sperm. 1/10 of the resulting total volume was added of Sodium Acetate (pH 5.2). Lastly, 100% ethanol, three times the amount of the total volume of the mixture, was added. Everything was vortexed for 5 seconds and stored in a freezer at -20°C overnight.

The next day the probe was centrifuged at 4°C at 14,000 rpm for 30 minutes. The supernatant was removed, and the tube speed vacuumed for 10 minutes on medium heat until there was a dry pellet visible. 5  $\mu$ l of 37°C-warm deionized formamide was added and the tubes incubated, shaking at 700 rpm at 37°C for 30 minutes. MasterMix was added and tubes put in the thermomixer at 37°C, shaking at 700 rpm for 30 minutes, once again. For new clones, the precipitations were done individually at first, once the clones were tested and established, they were precipitated as a contig.

The FISH probes were combined into two panels, the first one containing *ZNF217*, *HER2*, *TP53*, *CDH1*, and *CEP10*, the second one consisting of *MYC*, *COX2*, *DBC2*, *CCND1*, and *CEP4* (see Table 4).

### 3.2.5 Hybridization

Proper hybridization was controlled by hybridizing the precipitated probes on lymphocyte slides and imaging them after detection washes and a counterstain with DAPI-antifade. Once quality was confirmed on the lymphocyte slides, the two probe panels were consecutively hybridized onto the same cytopsin, so that counts of all 10 probes could be detected in the same nuclei.

Before hybridization each case slide had to be pretreated in a jar with 0.05% pepsin and 0.01 M HCl. The sample was then fixed with NBF and dehydrated in 70%, 90%, and 100% ethanol for 5 minutes each and then left to air-dry. The slides were denatured in 70% deionized formamide/ 2X SSC (standard saline concentrate) for 2 minutes at 73°C on a ThermoBrite StatSpin system. Then they were dehydrated in an ice-cold series of 70%, 90% and 100% ethanol for 3 minutes each and allowed to air-dry. The FISH gene probes were denatured at 73°C for 5 minutes and pre-annealed for 1 hour, shaking in a Thermomixer at 37°C. The centromere probes were denatured at 73°C for 5 minutes and then mixed with the pre-annealed gene probes. The probe panel was then hybridized onto the case slide, covered with coverslips and sealed with rubber cement. The slides were placed in a humid, light tight hybridization chamber at 37°C for 12-48 hours.

Panel I and II were consecutively hybridized onto the same cytospins, so that all ten probes could be visible within the same nucleus for each cell. For the re-hybridization, once the slides had been imaged with the first probe panel, the coverslip was removed, and the panel was washed off in 2X SSC three times for 2 minutes, and then in 70% formamide/ 2X SSC at 80°C for 1 minute. Slides were dehydrated in a cold ethanol series of 70%, 90%, and 100% ethanol for 2 minutes each and allowed to air-dry. The mix of denatured, pre-annealed probe and denatured centromere probe were then added onto the slide, a coverslip was added and sealed off with rubber cement, and the slide was hybridized at 37°C overnight.

### **3.2.6 Detection and Imaging**

After removal of the coverslip, the slides were washed in 48°C warm 2X SSC/ 0.3% NP<sub>4</sub>O solution for 2 minutes with gentle agitation, 2X SSC/0.1% NP<sub>4</sub>O for 1 minute at room temperature, and 2X SSC for 10 seconds at room temperature. Then they were dehydrated in an ethanol series of 70%, 90%, and 100% for 2 minutes each and allowed to air-dry. DAPI-counterstain was applied onto the slide, which was then covered with a coverslip.

For each of the 11 samples, 12,000 nuclei were automatically imaged with a fluorescence microscope and a 20x immersion oil objective, equipped with custom optical filters with a motorized stage and custom scanning and analysis software. The software allowed for automatic overlay of each scan of the same target nuclei. Each of the 12,000 nuclei with its ten signal counts was presented in a custom gallery overview. Enumeration was performed in a consecutive manner starting with target one. All signal counts were performed automatically using the BioView® software but manually checked for accuracy and corrected where necessary. Nuclei were only accepted for further analysis if all ten signal counts were clearly visible. If nuclei overlapped with one another or if a nucleus was damaged or incomplete, they were excluded from analysis. Corresponding with this procedure, 400 nuclei were counted for all cases, except case A3 and A4, for which only 300 nuclei were applicable for analysis. Signal counts were automatically recorded in Excel spreadsheets that were exported and used for subsequent analysis.

### **3.3 Imaging Data Files**

Automatically recorded excel spreadsheets contained signal counts for the two centromere probes (CEP10 and CEP4) and the eight gene probes (*COX2*, *DBC2*, *MYC*, *CCND1*, *CDH1*, *TP53*, *HER2*, *ZNF217*), the later ones sorted by chromosome order. Every row of the 10 probe signals represented a signal pattern. Any patterns that existed more than once were grouped together and the count delineated in an extra row. Any signal pattern consisting of two counts for every probe was classified as a diploid cell, most likely reflecting a stromal or immune cell, and was not included into the final count. The most common signal pattern in every cell population was determined as the major signal pattern clone.

### 3.4 Signal pattern and instability index

For every breast cancer sample, we evaluated the major signal pattern and the percentage of cells with the major signal pattern clone. A signal pattern of 2-2-2-2-2-2-2-2-2-2 indicated a cell with two signals for both centromeric probes and each of the eight genes, most likely a stromal or immune cell.

The number of different signal patterns per sample was divided by the number of counted cell nuclei to generate the instability index. Cells with the signal pattern described above were not included in the evaluation. Thus, the instability index indicated the aberrance from genetic stability of the malignant cells.

### 3.5 Gene gains and losses

To evaluate gains and losses of the analyzed genes, each nucleus analyzed by FISH was assigned an integer ploidy value in collaboration with Dr. K. Heselmeyer-Haddad at the National Cancer Institute, National Institutes of Health, MD, USA. For cases D2-D5, A1, A3, and A6, the assignment was based on the average of all ten FISH probes. For the remaining cases, ploidy was assigned by using only a subset of probes, while excluding those probes with amplifications or losses that biased the average. For case D1, *DBC2*, *CDH1*, *TP53* and *MYC* were not included. For case A2, *MYC* and *CCND1* were not included. For A5, *DBC2*, *CDH1* and *TP53* were not included. Each signal pattern was assigned a ploidy value based on average values of gene counts. If any disagreement arose in assigning ploidy, knowledge of which genes were oncogenes and therefore frequently gained and which genes were tumor suppressor genes and hence frequently lost, was taken into consideration. The resulting ploidy assignments matched measurements from the DNA histograms. All diploid cases showed predominantly diploid nuclei, cases A2 and A5 exhibited mainly triploid nuclei. Aneuploid cases A1 and A3 presented with populations of both diploid and triploid nuclei. Due to vastly varying signal patterns, case A4 could not be assigned a ploidy value using signal count average. Hence, only the DNA histogram was

used. As it showed an average ploidy value of 4.9, case A4 was assigned as pentaploid.

To get a clear overview of common gains and losses within the cell population of each sample, and not only see each individual signal pattern, we used a signal pattern comparator. Instead of each signal count, the comparator used three symbols: G (gained), N (normal), and L (lost).

Depending on the assigned ploidy value, every signal count in a pattern was defined as either G, if the value was greater than the nominal ploidy value; L, if the count was less than the ploidy value; or N, if the signal count equaled the ploidy value. Again, gene probes were in order of their chromosome location. For example, a gain of *MYC* and loss of *TP53*, with all other gene probes normal, was described as NNGNNLNN. Equal patterns were then totaled and sorted by frequency. Based on these files we then designed color displays for every case (see figures 3.1 - 3.11). These illustrate numbers and percentages of clone populations, and also depict gain and loss percentages for each gene.

In addition to the signal pattern evaluations, gains and losses were also analyzed for every cell of the tumor population. A threshold was then set at 15% to ensure an overview of common gains and losses. Thus, only if at least 15% of cells of a tumor cell population exhibited a copy number change, they were included in the comparison. We then divided the total number of changes by the number of cases in the diploid and the aneuploid group, respectively, to receive the average number of lesions per ploidy group. All changes are depicted in table 5 in the Results section.

### **3.6 FISHtrees**

M. Gertz of the Computational Biology Branch, National Center for Biotechnology Information at the National Institutes of Health used the FISHtrees 3.1 software<sup>116</sup> to compute tumor progression tree models, to facilitate the reconstruction of clonal relationships. Utilized was the weighted, ploidyless mode,<sup>117</sup> in which gains and

losses of single genes, gains and losses of single chromosomes, and genome doubling were modeled as distinct events with different probabilities.

### **3.7 Targeted next generation sequencing**

#### **3.7.1 DNA extraction from archived FFPE Specimen**

To prepare the samples for sequencing, DNA was extracted from the FFPE specimen. We used the DNeasy® Blood & Tissue kit from Qiagen for all buffers and chemicals. For each sample, tumor tissue was scraped off a 50 µm slide using a scalpel and placed into a 2.0 mL screwcap tube. 1 mL of mineral oil, 200 µL buffer ATL and 40 µL of Proteinase K were added. The tube was then placed in a thermomixer at 65°C for overnight incubation. The following morning the tube was quickly centrifuged. If the samples were not completely digested, 20-40 µL of Proteinase K were added, and the tube was again placed in the thermomixer for a few more hours or overnight, until the tissue was completely digested. The tubes were centrifuged for one minute at 13,200 rpm. 100 µL of the lysate were moved to a 1.5 mL Eppendorf tube. 1.5 µL of RNase A was added, the tube vortexed for 5 seconds and centrifuged for 10 seconds at 8,000 rpm and left to incubate for 5 minutes at room temperature. 500 µL of buffer PM and 10 µL of 3 M sodium acetate were added, and the tube vortexed for 5 seconds. The mixture was pipetted onto Quick spin columns and centrifuged at 5,000 rpm for 30 seconds to collect the eluate. To ensure all DNA was bound to the column, the flow-through was re-applied to the same column and spun again at 5,000 rpm for 30 seconds. Collection tubes were changed, 700 µL of buffer PE added and centrifuged at 8,000 rpm for 30 seconds. Again, collection tubes were changed, 700 µL of 80% Ethanol was added and centrifuged at 13,200 rpm for 1 minute. Collection tubes were once again changed, the filter columns uncapped and centrifuged at 13,200 rpm for 5 minutes and then placed into the thermomixer at 65°C for 10 minutes to ensure that any traces of ethanol were dried off. The filter columns were then placed in 1.5 mL Eppendorf tubes. 40-75 µL of 10% AE were added to the columns and incubated in the thermomixer at 65°C for 5 minutes, then centrifuged



for 1 minute at 13,200 rpm. To ensure all DNA was bound to the column, the first elution was re-applied to the same column, left to incubate for 5 minutes in the thermomixer at 65°C and centrifuged for 1 minute at 13,200 rpm. DNA quality and yield were assessed using NanoDrop®.

### **3.7.2 Sequencing**

The sequencing for this project was performed by the laboratory of Dr. P. Meltzer, Genetics Branch, Center for Cancer Research, National Cancer Institute at the National Institutes of Health, in Bethesda, MD.

Targeted next generation sequencing was performed with a capture assay called OncoVar®, which sequences coding exons of 563 cancer related genes.<sup>118</sup> The resulting paired-end libraries were sequenced on Illumina NextSeq 500 sequencers.

Sequencing data processing and analysis was performed by M. Gertz and A. Schäffer of the Computational Biology Branch, National Center for Biotechnology Information/ National Institutes of Health, Bethesda, MD. Data processing and variant calling procedure followed the Best Practices workflow recommended by the Broad Institute.<sup>119</sup> Briefly, the raw sequencing reads were mapped to the human genome build 19 by Burrows-Wheeler Aligner <sup>120</sup> followed by local realignment. Duplicated reads were marked by Picard tools. The UnifiedGenotyper from the Broad Institute was used for variant calling.<sup>121</sup> Multiple annotation databases, including dbNSFP <sup>122</sup>, dbSNP <sup>147</sup> (NCBI) <sup>123</sup>, ESP6500 (NHLBI Exome Sequencing Project) , and the COSMIC database <sup>124</sup> were used to annotate and predict the effects of variants.

Several filtering criteria were used to drop a fraction of the variant calls. Then, allele frequency filtering was applied and validated visually using the Integrative Genomics Viewer (IGV, Broad Institute).

### **3.8 Statistical analysis**

Statistical analyses were performed along with Mrs. L. S. Hernandez of the National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.

For all comparisons between diploid and aneuploid tumors, two-tailed t-tests with Welch's correction were performed. A p-value of  $< 0.05$  was deemed statistically significant.

## 4 Results

In this study, we conducted a multi-FISH analysis of diploid and aneuploid tumors to analyze the distinctions between both groups regarding individual gene gains and losses as well as the degree of intra-tumor heterogeneity. We investigated existing correlations between ploidy and clinical parameters, and we analyzed the tumor cell populations for aneuploidy associated gene mutation spectra.

### 4.1 DNA ploidy

Five of the tumors in the study showed a diploid DNA content and six were aneuploid. All cases were classified according to Auer et al. <sup>107</sup> as described in the Methods section.

### 4.2 Hybridization Panel I

Of the 12,000 automatically counted nuclei, 400 nuclei (300 for case A3 and A4, respectively) were included in further analysis for each case. Fluorescence signals were counted for *CDH1* (16q), *TP53* (17p), *HER2* (17q), and *ZNF217* (20q) and analyzed in comparison to the centromeric probe 10. Centromeric probe 10 hybridizes specifically to chromosome ten and acts as a ploidy control to deduce conclusions about gene amplifications and deletions of the four genes analyzed in this panel.

### 4.3 Hybridization Panel II

For panel II, we analyzed signal counts for *COX2* (1q), *DBC2* (8p), *MYC* (8q) and *CCND1* (11q). Centromeric probe 4 was used as ploidy control. All cell nuclei which had been counted for panel I were once again counted for panel II.

For each of the eleven cases, the combined data for both panel I and II was then analyzed.

#### 4.4 Gene gains and losses

Marker	No. of lesions with >15% of cells with specific marker gain or loss	
	Diploid Sample (n=5)	Aneuploid Sample (n=6)
<i>COX2</i> gain	1/5	5/6
<i>COX2</i> loss	0/5	3/6
<i>DBC2</i> gain	0/5	1/6
<i>DBC2</i> loss	1/5	5/6
<i>MYC</i> gain	1/5	6/6
<i>MYC</i> loss	0/5	0/6
<i>CCND1</i> gain	1/5	3/6
<i>CCND1</i> loss	0/5	2/6
<i>CDH1</i> gain	0/5	2/6
<i>CDH1</i> loss	3/5	5/6
<i>TP53</i> loss	4/5	6/6
<i>HER2</i> gain	2/5	2/6
<i>HER2</i> loss	1/5	3/6
<i>ZNF217</i> gain	0/5	3/6
<i>ZNF217</i> loss	0/5	2/6
<b>Total no. of changes</b>	14	48
<b>Average no. of changes/lesion</b>	2.8	8.0

**Table 5: Comparison of gain and loss frequency of analyzed genes for diploid and aneuploid tumors. E.g. a *COX2* gain was observed in at least 15% of cells in one out of the five diploid tumor samples and in five out of six aneuploid tumor samples.**

Gene gain and loss frequencies are depicted in table 5. We set the threshold for the comparison of gain and loss frequency for diploid and aneuploid tumors at >15%. Thus, only if a gain or loss occurred in at least 15% of the population, it was included in the analysis. This ensured a comparison of only the common changes. For the average count of copy number changes (total number of changes divided

by number of cases per group), we found 2.8 for the diploid tumors and 8.0 for the aneuploid ones ( $p=0.0001$ ). The diploid tumors displayed slightly more losses than gains. Particularly, four out of the five diploid cases showed a loss of *TP53* and three out of five diploid cases had a loss of *CDH1*. Among the aneuploid samples, gains and losses were almost balanced. Notably, all six aneuploid cases had a loss of *TP53* and a gain of *MYC*. Five out of six aneuploid cases showed a gain of *COX2*. *DBC2* as well as *CDH1* were lost in five out of the six aneuploid cases, respectively.

#### 4.5 Major signal pattern

Ploidy	Case No.	Major signal pattern in following order: Ploidy-COX-DBC-MYC-CCND- CDH-TP53-HER-ZNF	Gains	Losses
Diploid	D1	2-2-1-4-2-1-1-2-2	MYC	DBC2, CDH1, TP53
	D2	2-2-2-2-3-2-1-3-2	CCND1, HER2	TP53
	D3	2-4-2-2-2-1-1-1-2	COX2	CDH1, TP53, HER2
	D4	2-2-2-2-2-2-1-3-2	HER2	TP53
	D5	2-2-2-2-2-1-2-2-2		CDH1
Aneuploid	A1	(2-3-2-4-2-2-1-1-2) 3-2-2-10-3-4-2-2-5	COX2, MYC, CDH1, ZNF217	COX2, DBC2, TP53, HER2
	A2	3-3-2-12-6-2-2-4-4	MYC, CCND1, HER2, ZNF217	DBC2, CDH1, TP53
	A3	(2-2-2-2-2-1-2-2-2) 3-4-4-4-2-2-1-2-3	COX2, DBC2, MYC	CCND1, CDH1, TP53, HER2
	A4	5-6-2-8-8-5-3-3-8	COX2, MYC, CCND1, ZNF217	DBC2, TP53, HER2
	A5	3-4-1-4-3-2-2-4-2	COX2, MYC, HER2	DBC2, CDH1, TP53, ZNF217
	A6	3-6-2-4-3-2-2-3-3	COX2, MYC	DBC2, CDH1, TP53

**Table 6: Major signal pattern (genes in the following order: Ploidy-COX-DBC-MYC-CCND-CDH-TP53-HER-ZNF) and gene gains and losses for each case.**

The major signal pattern for each case as well as the most frequent gains and losses in these cell clones were found as presented in table 6. All gains and losses are in reference to the assigned ploidy of the case. Both case A1 and A3 consisted of a diploid as well as an aneuploid population and are therefore listed with two major signal patterns. Because both cases were assigned as aneuploid, the major signal pattern of the diploid population is shown in parenthesis.

For the major signal pattern, we found on average one gene gained and 1.8 genes lost for the diploid cases, and 3.3 genes gained and 3.5 lost for the aneuploid samples.

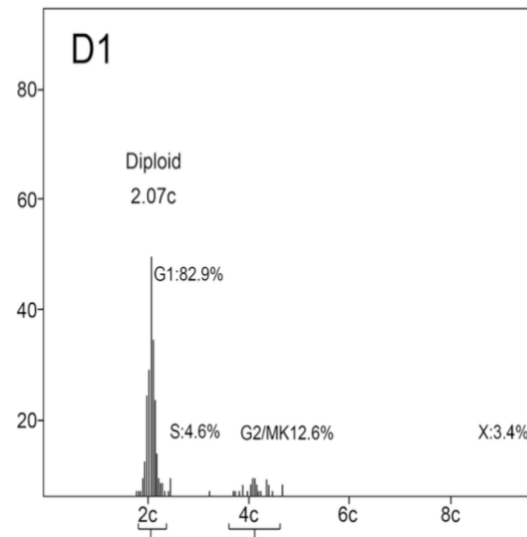
For the five diploid cases the mean percentage of cells with the major signal pattern clone was on average 69.7 % (49.3 – 84.5) while the aneuploid samples only showed an average of 11.0 % for the major clone (4.8 – 15.5) ( $p=0.0002$ ) (see table 8 in the appendix for complete data).

#### **4.6 Case studies**

##### **Case D1:**

Tumor sample D1 came from a 66-year-old postmenopausal patient. The tumor was clinically and pathologically classified as T1, was 7 mm<sup>2</sup> in size and ER- and PR-positive. None of the ten surgically resected lymph nodes were infiltrated by cancer cells. The patient received adjuvant radiochemotherapy. The patient showed neither local relapse nor distant metastases and was still alive at the end of the follow-up period with an overall observed survival time of 22.4 years.

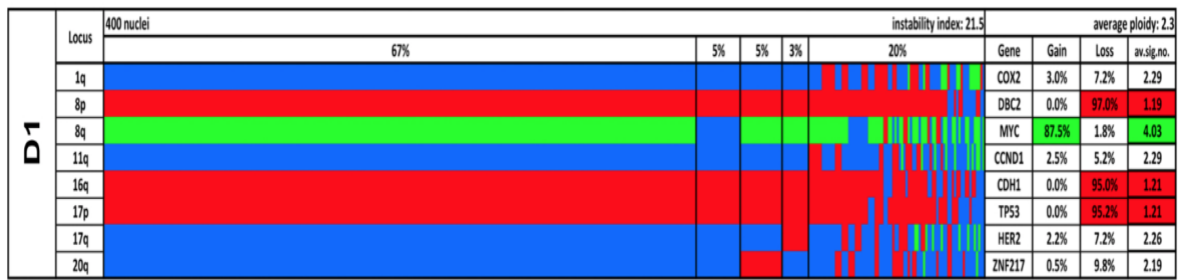
Figure 2.1 shows the histogram of case D1 which was classified as diploid according to Auer <sup>107</sup> with a nuclear DNA content of 2.07c. The histogram shows the percentage of cells in G1, S and G2/M phases of the cell cycle. In case D1, 82.9% of cells were in G1 phase, 4.6% of cells were in the S-phase and 12.6% of cells were in G2/MK-phase of the cell cycle.



**Fig. 2.1: Quantitative DNA content measurements for case D1. The DNA histogram shows the quantitative measurements of the nuclear DNA content, which was 2.07c for case D1, and the percentage of cells in G1, S, and G2/M phases of the cell cycle.**

As displayed in fig. 3.1, in case D1, 67% of the cells of the population had the same imbalanced clone, which showed a loss of *DBC2*, *CDH1* and *TP53*, as well as a gain of *MYC*. The gain of *MYC* (8q) and loss of *DBC2* (8p) seen here were indicative of the formation of an isochromosome 8q. A minor clone, which consisted of 5% of the tumor cell population, showed the same losses as the major clone, but did not show a gain of *MYC*. Another clone of 5% of the cells showed the same gain and loss pattern as the major clone, with an additional loss of *ZNF217* (20q). 20% of the cells of the population had more diverse gain and loss patterns, although the imbalances seen in the major clone were still present in most of these cells.

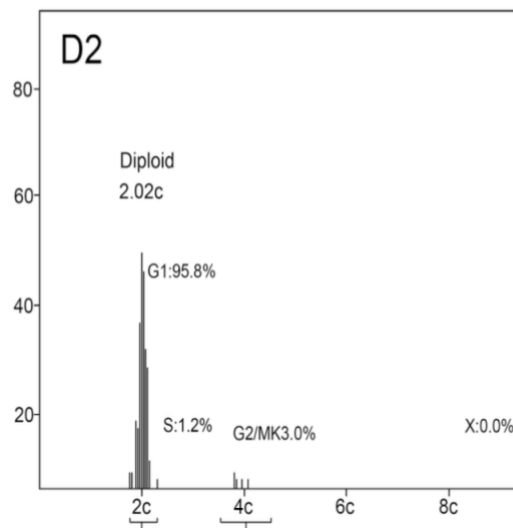
The instability index (number of different signal patterns per sample divided by number of counted cell nuclei) for this case was 21.5.



**Fig. 3.1:** Color display of miFISH analysis with eight gene-specific probes for case D1. Copy number counts for each nucleus are displayed as gains (green), losses (red) and unchanged (blue). Markers are plotted vertically with the “Locus” column depicting the specific chromosome arm for each probe. Nuclei are plotted horizontally by pattern frequency. Each vertical line discerns specific gain and loss patterns and how prevalent these clones are in the population. The instability index and the average ploidy were calculated as described in Material and Methods.

**Case D2:**

Sample D2 originated from a 72-year-old postmenopausal patient. The tumor measured 7 mm<sup>2</sup>, was classified as stage T1, and showed both ER- and PR-negativity. All nine analyzed lymph nodes were negative for tumor infiltration. Neither local relapse nor distant metastases were found. The patient received endocrine treatment after surgery and had an overall survival of 15 years, at which time she died of chronic cardiovascular disease.

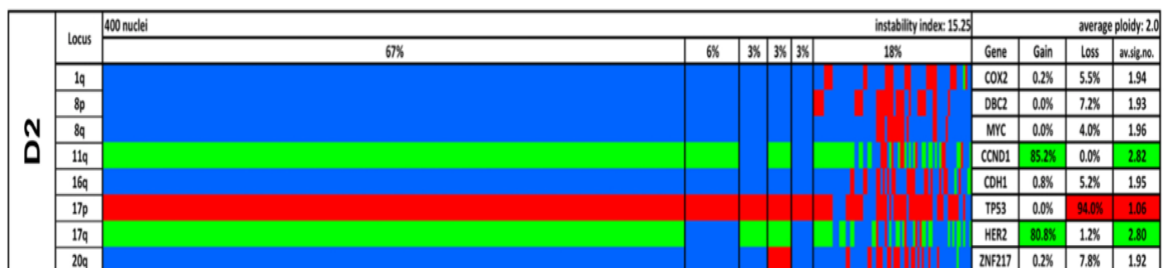


**Fig. 2.2:** Quantitative DNA content measurements for case D2. The DNA histogram shows the quantitative measurements of the nuclear DNA content, and the percentage of cells in G1, S, and G2/M phases of the cell cycle.



Figure 2.2 shows the nuclear DNA content measurements for case D2. With a DNA content of 2.02c it was classified as diploid according to Auer <sup>107</sup>; 95.8% of cells were in G1 phase.

Analysis of the color display (Fig. 3.2) revealed the following information: 67% of cells had a gain of *CCND1* and *HER2* and a loss of *TP53*, and thus showed the same gains and losses which were observed in the major signal pattern. Loss of 17p and gain of 17q was indicative of the formation of an isochromosome 17. 6% of the tumor cell population showed a gain of *CCND1* and a loss of *TP53* without a gain of *MYC*. Three clones consisting of 3% of the tumor population showed patterns similar to the major clone: One displayed an isochromosome 17 with no further changes, another showed a loss of *ZNF217* in addition to the changes of the major clone, and the third clone exhibited solely the loss of *TP53*. 18% of cells exhibited higher degrees of genomic instability. The instability index for this case was 15.3.



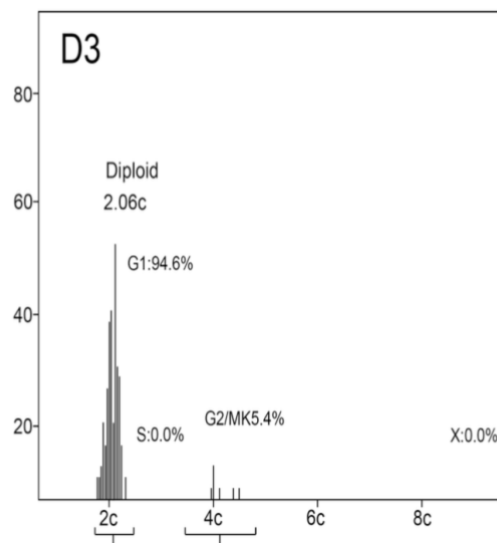
**Fig. 3.2:** Color display of miFISH analysis with eight gene-specific probes for case D2. Copy number counts for each nucleus are displayed as gains (green), losses (red) and unchanged (blue). Markers are plotted vertically with the “Locus” column depicting the specific chromosome arm for each probe. Nuclei are plotted horizontally by pattern frequency. Each vertical line discerns specific gain and loss patterns and how prevalent these clones are in the population. The instability index and the average ploidy were calculated as described in Material and Methods.

### **Case D3:**

The patient from which tumor sample D3 was taken was 45 years of age and premenopausal. The tumor was classified as T1, was 20 mm<sup>2</sup> in size and ER- as well as PR-positive. No lymph nodes were affected. The patient received a combination of radiotherapy and endocrine treatment post-surgery. A local relapse

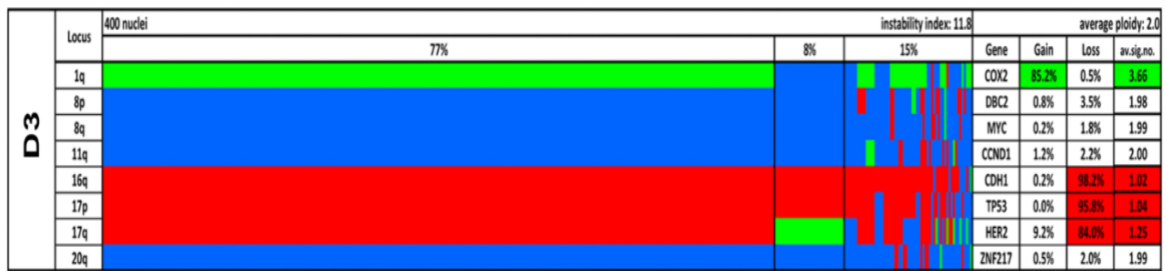
was diagnosed in the patient's ipsilateral breast. No distant metastases were found during the follow-up interval. The patient was alive at the end of the follow-up period, with an overall observed survival of 22 years.

The case was classified as diploid according to Auer <sup>107</sup> with a nuclear DNA content of 2.06c (see Fig. 2.3 below). 94.6% of cells were in G1 phase of the cell cycle and 5.4% were in G2/MK-phase.



**Fig. 2.3: Quantitative DNA content measurements for case D3. The DNA histogram shows the quantitative measurements of the nuclear DNA content, and the percentage of cells in G1, S, and G2/M phases of the cell cycle.**

The color display for case D3 is shown in Figure 3.3. For this case, we found that 77% of cells had a gain of *COX2*, as well as a loss of *CDH1* and the entire chromosome 17. One clone consisting of 8% of the population displayed an isochromosome 17 and a loss of *CDH1*. 15% of cells showed various imbalance clones. Notably, most of these maintained the changes seen in the major clone, but showed additional imbalances, mostly losses, of other genes. The instability index for case D3 was 11.8.

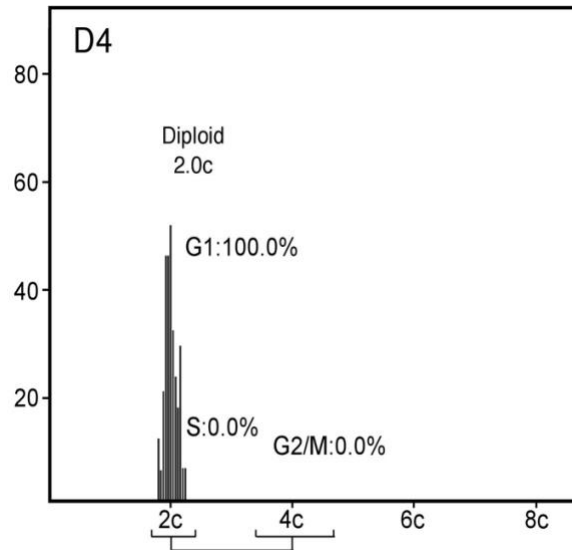


**Fig. 3.3: Color display of miFISH analysis with eight gene-specific probes for case D3. Copy number counts for each nucleus are displayed as gains (green), losses (red) and unchanged (blue). Markers are plotted vertically with the “Locus” column depicting the specific chromosome arm for each probe. Nuclei are plotted horizontally by pattern frequency. Each vertical line discerns specific gain and loss patterns and how prevalent these clones are in the population. The instability index and the average ploidy were calculated as described in Material and Methods.**

#### **Case D4:**

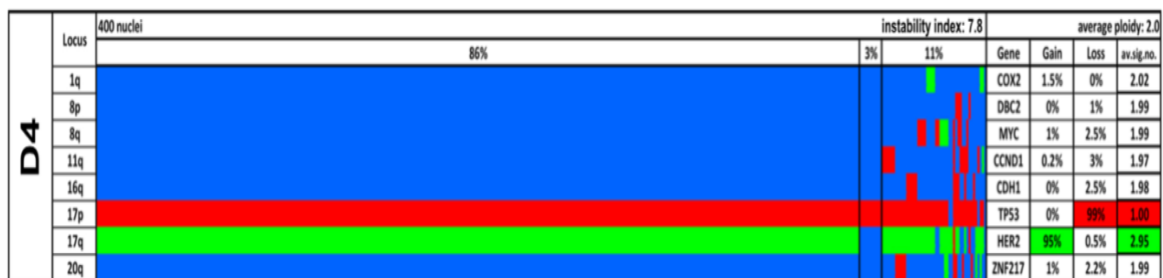
Sample D4 was taken from a 62-year-old, postmenopausal patient. Her breast tumor was clinically classified as T2, however surgery revealed a malignancy of 18 mm<sup>2</sup> in size, classifying it as pT1. Both ER- and PR- status were negative. No lymph nodes were affected. The patient received adjuvant endocrine treatment. Neither a local relapse nor distant metastases were recorded. The patient was alive at the end of the follow-up period with a recorded overall survival of 21.6 years.

Case D4 was classified as diploid according to Auer <sup>107</sup> with a nuclear DNA content of 2.0c (as displayed in Fig. 2.4). 100% of cells were in G1 phase of the cell cycle.



**Fig. 2.4: Quantitative DNA content measurements for case D4.** The DNA histogram shows the quantitative measurements of the nuclear DNA content, and the percentage of cells in G1, S, and G2/M phases of the cell cycle.

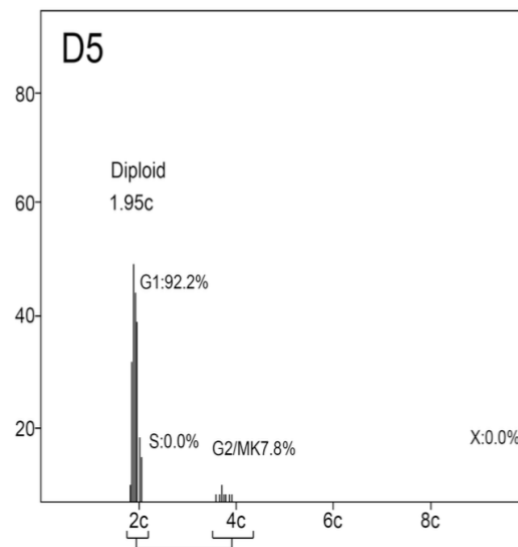
86% of the tumor cell population displayed the same imbalances: a loss of *TP53* and a gain of *HER2* (see Fig. 3.4). Again, an isochromosome 17q had likely been formed. A minor clone of 3% showed a loss solely of *TP53*. 11% exhibited more diverse gain and loss patterns with additional gains and losses. However, in most of these the loss of *TP53* and the gain of *HER2* were maintained. In the overall population, *TP53* was in fact lost in 99% of cells, while a gain of *HER2* was observed in 95% of all cells. The instability index for case D4 was 7.8.



**Fig. 3.4: Color display of miFISH analysis with eight gene-specific probes for case D4.** Copy number counts for each nucleus are displayed as gains (green), losses (red) and unchanged (blue). Markers are plotted vertically with the “Locus” column depicting the specific chromosome arm for each probe. Nuclei are plotted horizontally by pattern frequency. Each vertical line discerns specific gain and loss patterns and how prevalent these clones are in the population. The instability index and the average ploidy were calculated as described in Material and Methods.

**Case D5:**

Sample D5 came from a 69-year-old, postmenopausal patient. The tumor was classified as T1, was 11 mm<sup>2</sup> in size and both ER- and PR-positive. There was no lymph node infiltration. The patient received adjuvant radiochemotherapy. Neither a local relapse nor distant metastases were recorded during the 21.6 years of follow-up and the patient was still alive at the end of the follow-up period.



**Fig. 2.5:** Quantitative DNA content measurements for case D5. The DNA histogram shows the quantitative measurements of the nuclear DNA content, and the percentage of cells in G1, S, and G2/M phases of the cell cycle.



**Fig. 3.5:** Color display of miFISH analysis with eight gene-specific probes for case D5. Copy number counts for each nucleus are displayed as gains (green), losses (red) and unchanged (blue). Markers are plotted vertically with the “Locus” column depicting the specific chromosome arm for each probe. Nuclei are plotted horizontally by pattern frequency. Each vertical line discerns specific gain and loss patterns and how prevalent these clones are in the population. The instability index and the average ploidy were calculated as described in Material and Methods.

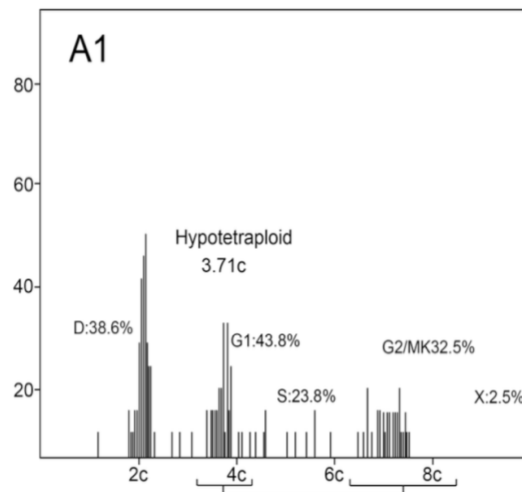
As depicted in Fig. 2.5, case D5 was classified as diploid according to Auer <sup>107</sup> with a nuclear DNA content of 1.95c. 92.2% of cells were in G1 phase and 7.8% of cells were in G2/MK-phase of the cell cycle.

Case D5 was characterized by the loss of *CDH1* in 97% of all cells of the population (see Fig. 3.5). 73% of the tumor population showed an exclusive loss of this gene. A minor clone of 2% had a loss of *TP53* in addition to the loss of *CDH1*. Another clone of 2% of the population showed a loss of *CDH1*, as well as a gain of *CCND1*. 23% of cells of the population had more diverse imbalances, with additional gene gains and losses. Interestingly, in addition to the loss of *CDH1*, one part of this group almost exclusively displayed copy number losses, while another part of the group showed only copy number gains. The instability index for case D5 was 14.3.

#### **Case A1:**

Tumor sample A1 came from a 40-year-old premenopausal patient. The tumor was classified as T2, was 30 mm<sup>2</sup> in size and both ER- and PR-negative. None of the eleven surgically resected lymph nodes were infiltrated by cancer cells. The patient received adjuvant radiochemotherapy. She showed neither local relapse nor any distant metastases and was still alive at the last follow-up with an overall observed survival time of 22.1 years.

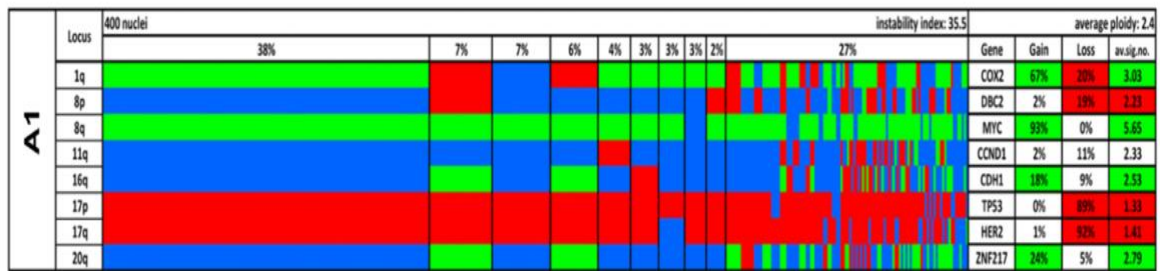
Case A1 was hypotetraploid and therefore classified as aneuploid according to Auer <sup>107</sup> with a nuclear DNA content of 3.71c and high percentages of cells in G1 (43.8 %), S (23.8%), and G2/M phases (32.5%) as can be seen in Fig. 2.6 below.



**Fig. 2.6: Quantitative DNA content measurements for case A1. The DNA histogram shows the quantitative measurements of the nuclear DNA content, and the percentage of cells in G1, S, and G2/M phases of the cell cycle.**

Case A1 exhibited a diploid and a triploid/ hypotetraploid population. The diploid clone made up the major clone with 38% of the cells. A diploid minor clone (7%) displayed the same losses and a gain of *MYC*, but no increase of *COX2*.

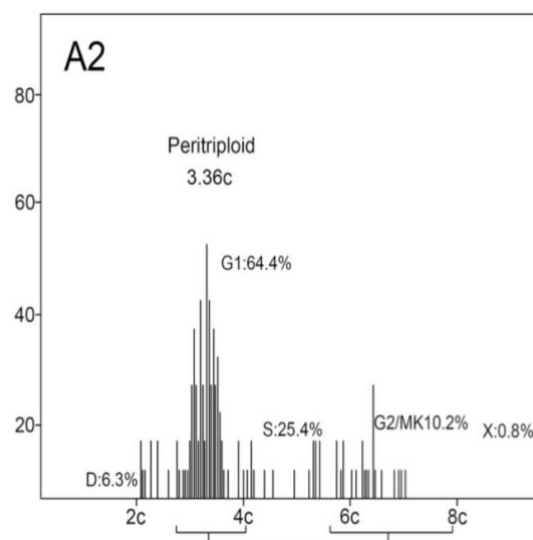
There were also two triploid/ hypotetraploid clones, one with 7% and one with 6% of all cells. Both of which had a gain of *MYC* and loss of chromosome 17, just like the diploid clones. In addition to these changes, both showed a loss of *COX2*, as well as gains of *CDH1* and *ZNF217*. The 7% clone additionally had a loss of *DBC2*. Five minor clones, consisting of at least 2% of the population, displayed several similarities to the major clone. All five clones had a gain of *COX2* and a loss of *TP53*. Four out of five clones also showed a gain of *MYC* and the loss of *HER2*, respectively. In summary, each of the five clones had just the same gain and loss pattern as the major clone, except for one gene each. 27% of the population showed very diverse patterns and a relatively high heterogeneity with numerous copy number gains and losses affecting almost every gene. Notably, the loss of chromosome 17 and the gain of *MYC* were still maintained in most of these clones. Overall, 93% of cells had a gain of *MYC*, 89% a loss of *TP53* and 92% a loss of *HER2*. A gain of *COX2*, as observed in the major clone, was seen in two thirds of the population. Case A1 displayed an instability index of 35.5. All information is depicted in Fig.3.6.



**Fig. 3.6:** Color display of miFISH analysis with eight gene-specific probes for case A1. Copy number counts for each nucleus are displayed as gains (green), losses (red) and unchanged (blue). Markers are plotted vertically with the “Locus” column depicting the specific chromosome arm for each probe. Nuclei are plotted horizontally by pattern frequency. Each vertical line discerns specific gain and loss patterns and how prevalent these clones are in the population. The instability index and the average ploidy were calculated as described in Material and Methods.

**Case A2:**

Sample A2 came from a 60-year-old, postmenopausal patient. The tumor was classified as T2, was 30 mm<sup>2</sup> in size and both ER- and PR-positive. One of three resected lymph nodes showed an infiltration with cancer cells. The patient received adjuvant endocrine treatment; she was diagnosed with a local relapse in the lymph nodes 81 months after initial diagnosis. Distant metastases were found in the bones and later on in the liver. The patient died due to breast cancer 8.5 years after diagnosis of the primary carcinoma.

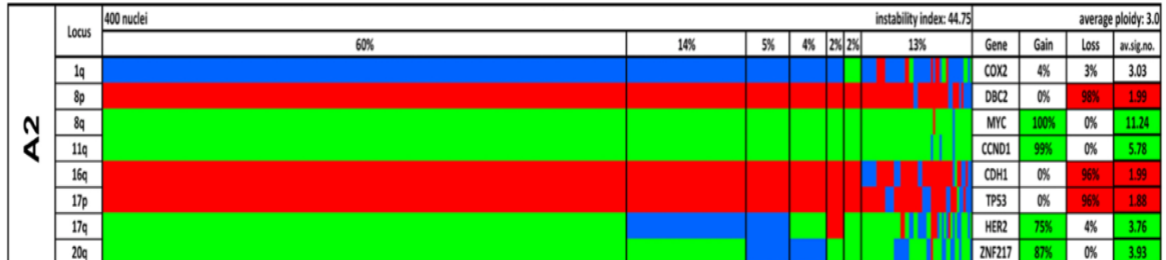


**Fig. 2.7:** Quantitative DNA content measurements for case A2. The DNA histogram shows the quantitative measurements of the nuclear DNA content, and the percentage of cells in G1, S, and G2/M phases of the cell cycle.



As can be seen in Fig. 2.7, case A2 was peritriploid and hence classified as aneuploid according to Auer <sup>107</sup> with a nuclear DNA content of 3.36c. About one third of the cells of the population were proliferating at the time the biopsy was taken, with 25.4% of cells in S-phase and 10.2% of cells in G2/MK-phase.

Consistent with the DNA measurements, the FISH analysis of case A2 (displayed in Fig. 3.7 below) revealed a triploid clone. 60% of the population constituted the major clone, which showed, in accordance with the major signal pattern, losses of *DBC2*, *CDH1*, and *TP53*. *MYC*, *CCND1*, *HER2* and *ZNF217* were gained. Only *COX2* showed no copy number changes. Generally, this gain and loss pattern was maintained in a high percentage of the tumor population. Strikingly, 100% of cells had a gain of *MYC*, 99% a gain of *CCND1* and 98% a loss of *DBC2*. 96% of the cells showed a loss of *CDH1* and *TP53*, respectively. A minor clone of 14% had exactly the same changes as the major clone, except for a gain of *HER2*. The instability index for case A2 was calculated as 44.8.

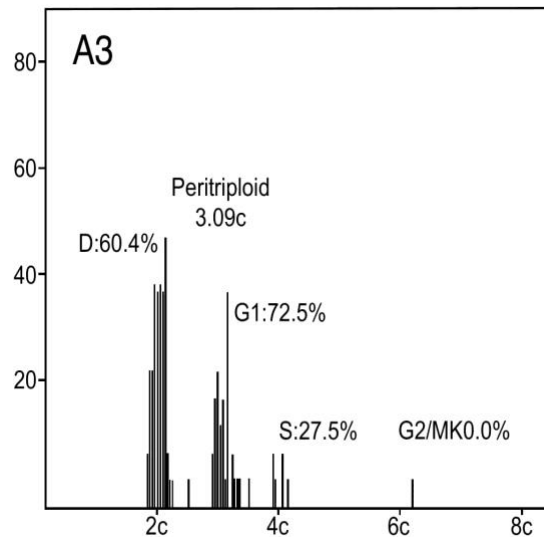


**Fig. 3.7:** Color display of miFISH analysis with eight gene-specific probes for case A2. Copy number counts for each nucleus are displayed as gains (green), losses (red) and unchanged (blue). Markers are plotted vertically with the “Locus” column depicting the specific chromosome arm for each probe. Nuclei are plotted horizontally by pattern frequency. Each vertical line discerns specific gain and loss patterns and how prevalent these clones are in the population. The instability index and the average ploidy were calculated as described in Material and Methods.

### **Case A3:**

Sample A3 was taken from a 53-year-old, postmenopausal patient. Her breast tumor was classified as stage T2 and was 27 mm<sup>2</sup> in size. The sample was ER-positive and PR-negative. 14 out of 18 resected lymph nodes showed infiltration by cancer cells. The patient was administered neoadjuvant radiochemotherapy.

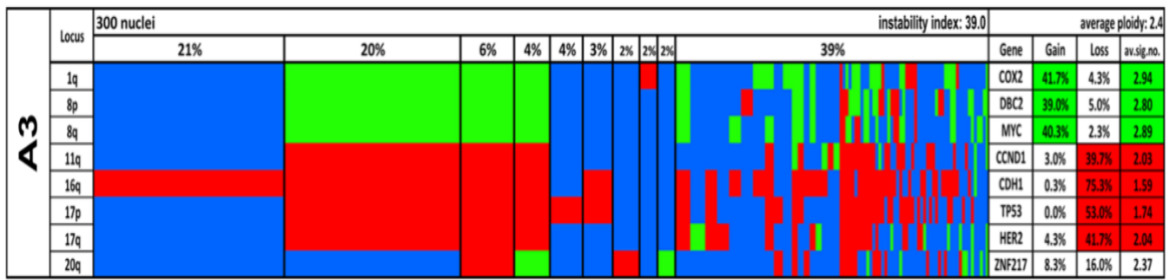
There was neither a local relapse nor distant metastases. However, the patient died from breast cancer 1.6 years after her initial diagnosis.



**Fig. 2.8: Quantitative DNA content measurements for case D3. The DNA histogram shows the quantitative measurements of the nuclear DNA content, and the percentage of cells in G1, S, and G2/M phases of the cell cycle.**

The case was peritriploid and therefore classified as aneuploid according to Auer<sup>107</sup> with a nuclear DNA content of 3.09c (see Fig. 2.8). 72.5% of cells were in G1 and 27.5% in S phase. No cells were found to be in G2/M phase.

Fig. 3.8 depicts the color display for case A3. We saw two major clones, which were almost equal in size. One consisted of 21% of the cells of the population, the other of 20% of cells. The former clone was diploid and showed a sole loss of *CDH1*. The later clone was triploid and showed a gain of *COX2* and the entire chromosome 8; *CCND1*, *CDH1*, and the entire chromosome 17 were lost. *ZNF217* was not subject to any copy number changes, although we found two minor clones with this additional change. One (6%) showed a loss of this gene, while the other (4%) showed a gain. 39% of cells of the tumor population exhibited more diverse gain and loss patterns. However, in the majority of these the gain and loss pattern of the triploid major clone was once again maintained. The instability index for case A3 was found to be 39.0.

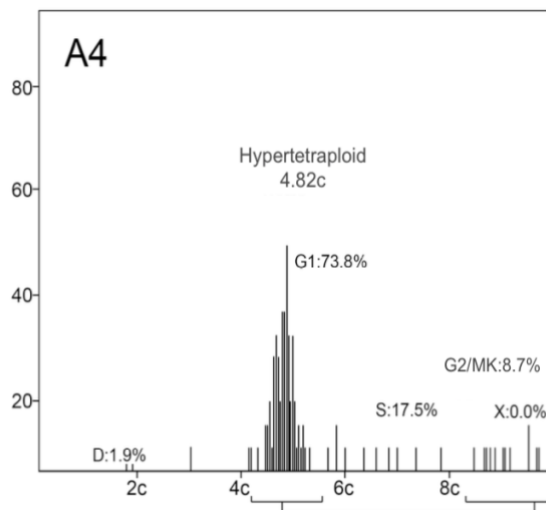


**Fig. 3.8:** Color display of miFISH analysis with eight gene-specific probes for case A3. Copy number counts for each nucleus are displayed as gains (green), losses (red) and unchanged (blue). Markers are plotted vertically with the “Locus” column depicting the specific chromosome arm for each probe. Nuclei are plotted horizontally by pattern frequency. Each vertical line discerns specific gain and loss patterns and how prevalent these clones are in the population. The instability index and the average ploidy were calculated as described in **Material and Methods**

#### **Case A4:**

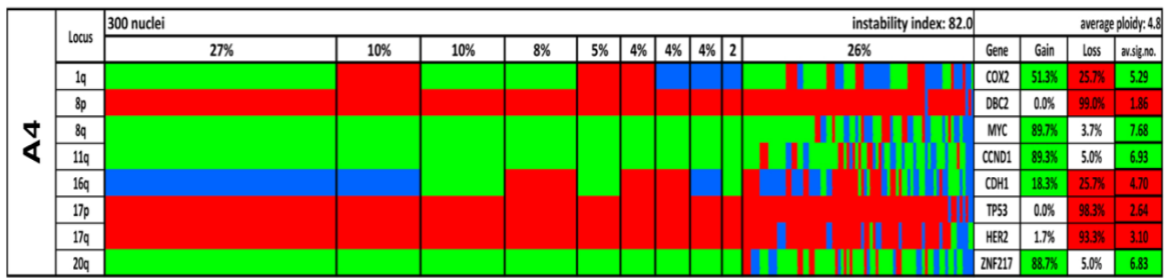
Sample A4 originated from a 49-year-old premenopausal patient. The tumor measured 20 mm<sup>2</sup>, was classified as stage T2, and showed both ER and PR positivity. All analyzed lymph nodes were negative for tumor infiltration. The patient received adjuvant endocrine treatment. She experienced a local relapse in the ipsilateral breast, as well as distant metastases in the mediastinum. The patient had an overall survival of 4.8 years, at which time she died of breast cancer.

Figure 2.9 shows the nuclear DNA content measurements for case A4. With a DNA content of 4.82c it was as hypertetraploid and thus classified as aneuploid according to Auer<sup>107</sup>. 73.8% of cells were in G<sub>1</sub>, 17.5% in S, and 8.4% of cells were in G<sub>2</sub>/M phase.



**Fig. 2.9: Quantitative DNA content measurements for case A4. The DNA histogram shows the quantitative measurements of the nuclear DNA content, and the percentage of cells in G1, S, and G2/M phases of the cell cycle.**

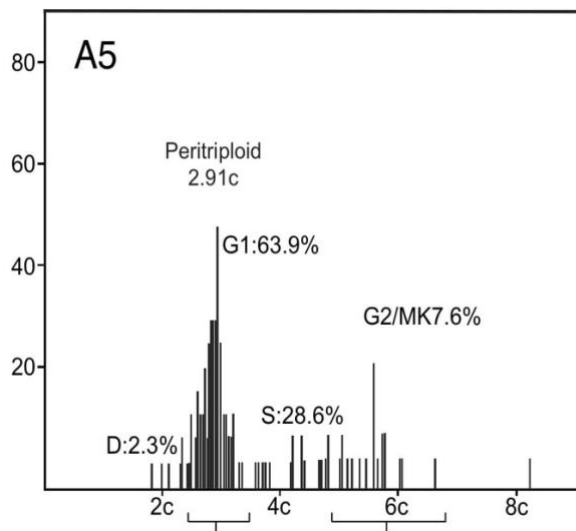
The major clone for case A4 consisted of only 27% of the cells of the population. It showed gains of *COX2*, *MYC*, *CCND1*, and *ZNF217*. *DBC2*, *TP53*, and *HER2* were lost. Only *CDH1* remained unaffected by copy number changes in this pentaploid clone. Case A4 included the only pentaploid major clone and hence displayed the major clone with the highest ploidy of all tumor samples included in this thesis. This correlates well with the fact that the instability index of 82% was the highest of all cases. In around 30% of cells of the population, all eight genes were affected by copy number changes. Interestingly, in this case we frequently observed both copy number gains and losses for both tumor suppressor genes as well as oncogenes. This was true especially for *COX2*, which was gained in about 50% of cells, but was also found lost in around 25% of cells. *CDH1* was lost in about 25% of cells but also showed gains in 18.3% of cells. All information is depicted in Fig. 3.9.



**Fig. 3.9:** Color display of miFISH analysis with eight gene-specific probes for case A4. Copy number counts for each nucleus are displayed as gains (green), losses (red) and unchanged (blue). Markers are plotted vertically with the “Locus” column depicting the specific chromosome arm for each probe. Nuclei are plotted horizontally by pattern frequency. Each vertical line discerns specific gain and loss patterns and how prevalent these clones are in the population. The instability index and the average ploidy were calculated as described in Material and Methods.

**Case A5:**

Sample D5 came from a 64-year-old, postmenopausal patient. The tumor was classified as T2, was 25 mm<sup>2</sup> in size, ER-positive and PR-negative. One of ten analyzed lymph nodes showed cancer cell infiltration. She received a combination of adjuvant radiation and endocrine therapy. The patient showed neither local relapse nor distant metastases during the 21.6 years of follow-up and was still alive at the last follow-up.



**Fig. 3.10:** Quantitative DNA content measurements for case A4. The DNA histogram shows the quantitative measurements of the nuclear DNA content, and the percentage of cells in G1, S, and G2/M phases of the cell cycle.

Case A5 was peritriploid and therefore classified as aneuploid according to Auer 107 with a nuclear DNA content of 2.91c (see Fig. 3.10). 63.9% of cells were in G1 phase, 28.6% in S phase, and 7.6% in G2/M phase of the cell cycle.

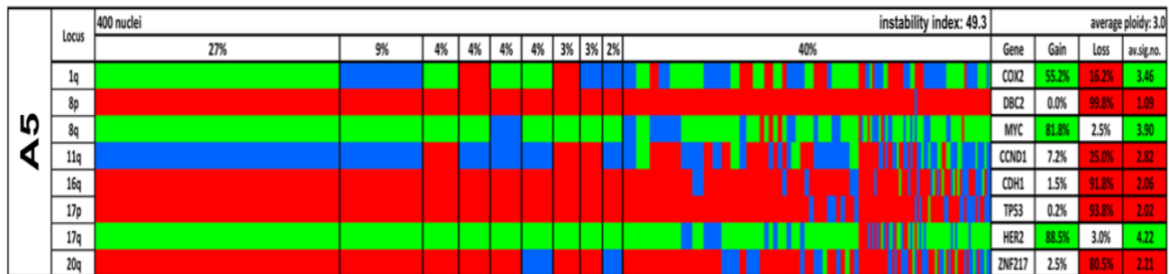


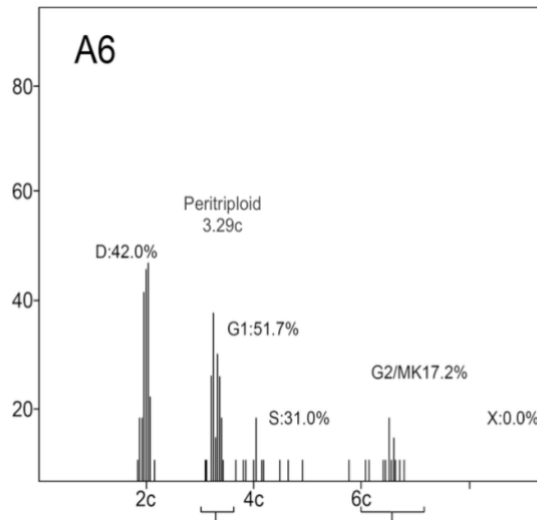
Fig. 3.10: Color display of miFISH analysis with eight gene-specific probes for case A5. Copy number counts for each nucleus are displayed as gains (green), losses (red) and unchanged (blue). Markers are plotted vertically with the “Locus” column depicting the specific chromosome arm for each probe. Nuclei are plotted horizontally by pattern frequency. Each vertical line discerns specific gain and loss patterns and how prevalent these clones are in the population. The instability index and the average ploidy were calculated as described in Material and Methods.

Fig. 3.10 shows the color display for case A5. The major clone for this case consisted of 27% of the cells of the tumor cell population. It showed a gain of *COX2*, a gain of *MYC* and a loss of *DBC2*, indicating an isochromosome 8q, a gain of *HER2* and a loss of *TP53*, indicating an isochromosome 17q, and losses of *CDH1* and *ZNF217*. Only *CCND1* remained unchanged. Several smaller clones, consisting of between 2% and 9% of the population, showed great similarity to the major clone. All of these clones displayed a loss of *DBC2*, *CDH1*, and *TP53*, as well as a gain of *HER2*. Distinctions were particularly based on changes in *COX2*, which was unchanged in three, gained in three, and lost in two of these small clones, respectively. In the 40% of cells which displayed more diverse signal patterns, the gain and loss pattern of the major clone was once again still apparent, however these clones showed fewer copy number changes overall. The instability index for case A5 was 49.3.

### Case A6:

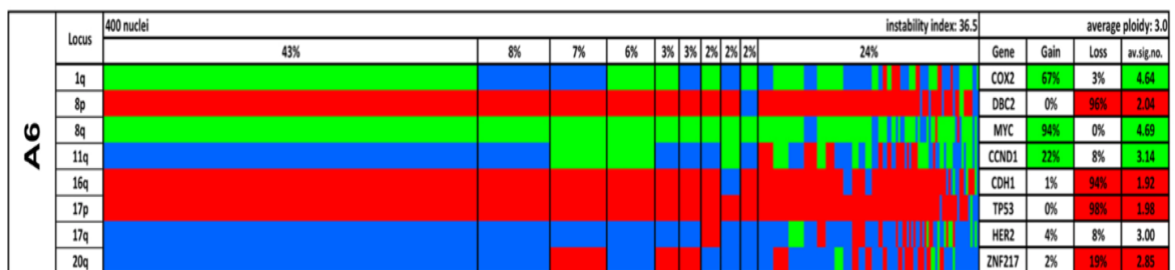
Sample D2 originated from a 46-year-old postmenopausal patient. The tumor was classified clinically as stage T2. It measured 15 mm<sup>2</sup>, categorizing it a pT1. Further analysis revealed both ER- and PR-negativity. None of the analyzed lymph nodes

showed tumor infiltration. The patient received adjuvant endocrine treatment and suffered a local relapse of the lymph nodes, but developed no distant metastases. The patient had an overall survival of 13.9 years, before she died of breast cancer.



**Fig. 2.11: Quantitative DNA content measurements for case A6.** The DNA histogram shows the quantitative measurements of the nuclear DNA content, and the percentage of cells in G1, S, and G2/M phases of the cell cycle.

Figure 2.11 shows the nuclear DNA content measurements for case A6. With a DNA content of 3.29c it was peritriploid and thus also classified as aneuploid according to Auer <sup>107</sup>. 51.7% of cells were in G1 phase, 31% in S phase, and 17.2% in G2/M phase.



**Fig. 3.11: Color display of miFISH analysis with eight gene-specific probes for case A5.** Copy number counts for each nucleus are displayed as gains (green), losses (red) and unchanged (blue). Markers are plotted vertically with the “Locus” column depicting the specific chromosome arm for each probe. Nuclei are plotted horizontally by pattern frequency. Each vertical line discerns specific gain and loss patterns and how prevalent these clones are in the population. The instability index and the average ploidy were calculated as described in Material and Methods.

Case A6 showed a gain of *COX2*, the indication of an isochromosome 8q, and loss of *CDH1* and *TP53* as its major clone in 43% of the population (see Fig. 3.11 for the color display of case A6). A clone of 8% of the population displayed equal changes except for the gain of *COX2*. Two small clones (7% and 6%) had both a gain of *MYC* and *CCND1* and losses of *DBC2*, *CDH1*, and *TP53*. One of these clones (7%) showed an additional loss of *ZNF217*, the other one (6%) an additional gain of *COX2*. Several minor clones also showed similar imbalances as the major clone, with an additional one or two gains or losses. The general gain and loss pattern was mostly still visible in the 24% of cells which exhibited more infrequent gain and loss patterns. The instability index for case A6 was 36.5.

#### **4.7 Summarized analysis of clinical parameters**

An overview of the clinical parameters for the eleven breast cancer samples can be seen in Table 2 in the Methods section.

Intriguing is the strong correlation between ploidy status and clinical features. Tumor samples from the aneuploid group were almost twice as large as samples from the diploid group (24.5 mm<sup>2</sup> and 12.6 mm<sup>2</sup>) ( $p=0.0033$ ). All diploid samples were classified as pT1. Of the aneuploid samples five out of six were classified as pT2, while one (A6) was classified as pT1.

Among the patients with diploid tumors, three received a combination of adjuvant radiation and endocrine therapy. Two were treated with adjuvant endocrine therapy alone. In the aneuploid group, three patients were treated with the combination therapy and three patients received endocrine monotherapy.

Of the five diploid cases, all were lymph node negative. In only one of these cases local metastases were found. None had distant metastases. None of these patients died from breast cancer. At the last follow-up for this study, four out of five patients were still alive, while one had passed away due to cardio-cerebrovascular disease (CCVD).

Of the six aneuploid cases, three out of six were lymph node positive. In three patients, a local recurrence occurred. Two out of six developed distant



metastases. One had both local as well as distant metastases. Four of the six patients died from breast cancer during follow-up time.

Overall survival time of patients with a diploid tumor was significantly longer compared to the aneuploid tumor group (20.6 years opposed to 12.1 years) ( $p=0.0048$ ).

#### **4.8 Correlation between clinical parameters and instability index**

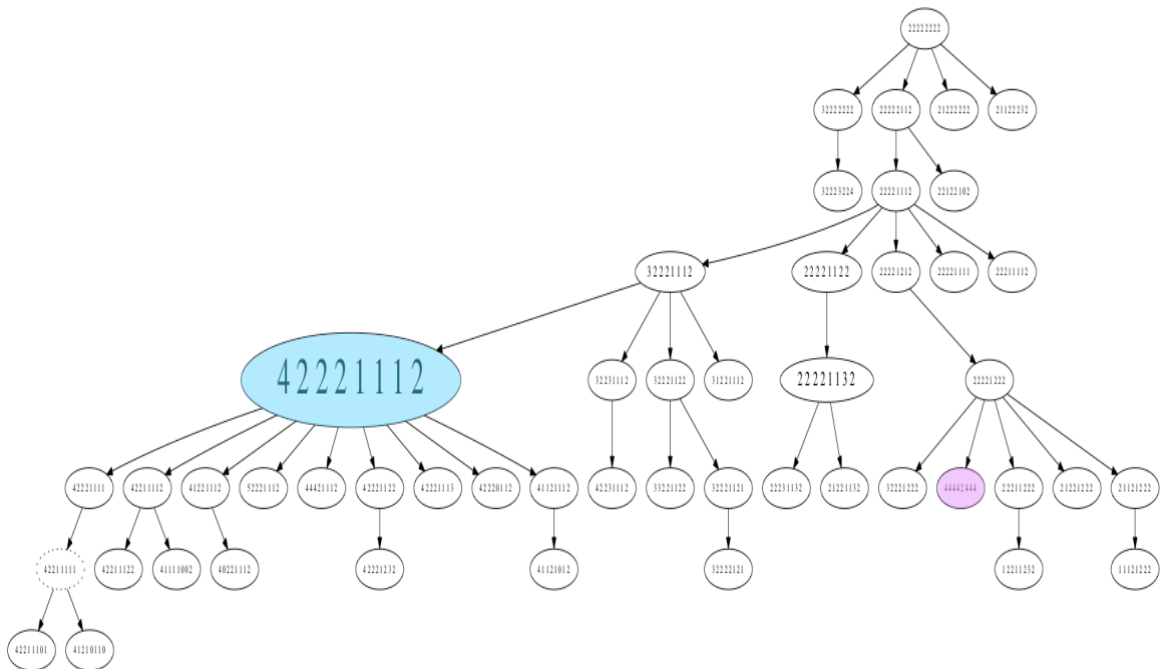
The significance of the correlation between ploidy status and clinical parameters is also illustrated when looking at the level of genomic instability as represented by the instability index (number of different signal patterns per sample divided by number of counted cell nuclei). All five diploid samples showed significantly lower instability indices compared to the aneuploid samples (mean instability index in diploid group 14.1 [7.8 – 21.5], in aneuploid group 47.9 [35.5 – 82.0],  $p=0.004$ ; see table 8 for complete data), emphasizing the increased level of genetic instability in the six aneuploid cases compared to the diploid tumor samples. All five patients with tumors with low instability index ( $< 30$ ) were still alive or had died from causes other than breast cancer during the follow up period. Of the six patients with tumors with a high instability index ( $> 30$ ), four died of breast cancer.

#### **4.9 FISHTrees**

We used the FISHTrees software version 3.1<sup>116</sup> to reconstruct clonal relationships. Each node reflects a signal pattern and its diagramed size equates to the percentage of cells in the tumor population with precisely this signal pattern. Where there was little branching in the FISHTree, it signified relative stability, while extensive branching reflects a high degree of copy number changes and intra-tumor heterogeneity. In all instances, several copy number changes were required to attain the cell with the most common signal pattern from which the less abundant and minor clones then emerged. In general, for both groups the individual cases tended to cluster together, i.e., the initiating events gave rise to tumor cell populations that are related in terms of copy number changes.

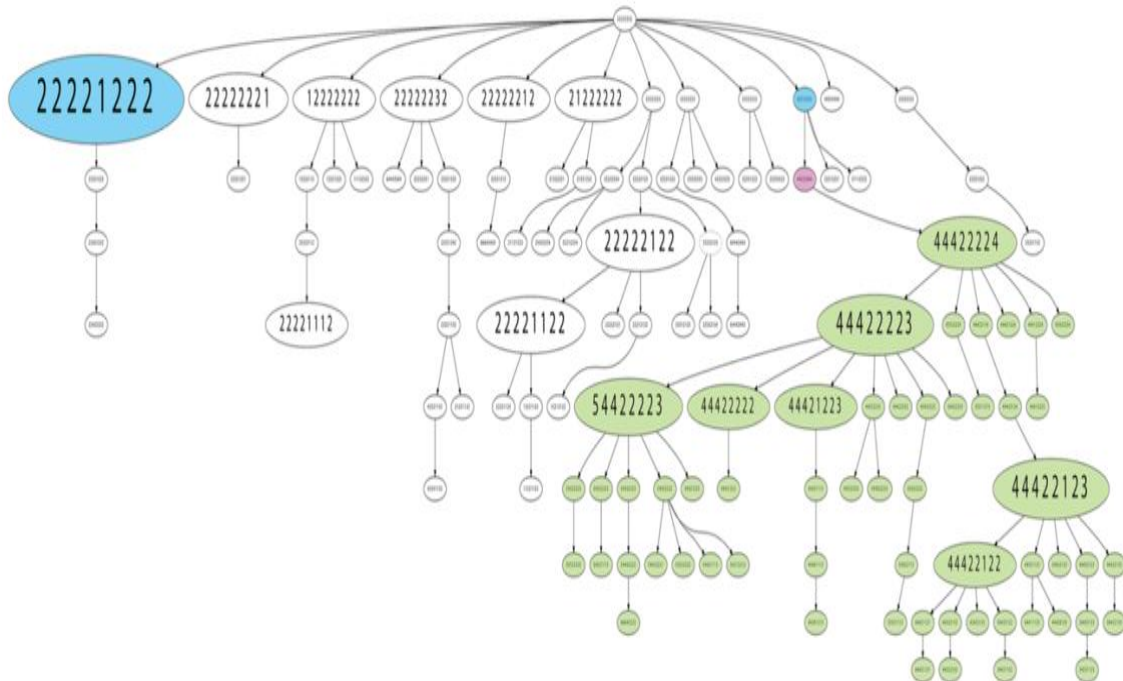
The diploid tumors displayed fewer nodes and less branching as a reflection of the lower instability indices compared to the aneuploid cases, which presented a multitude of nodes in different tree levels and substantial branching. Figures 4.1 and 4.2 exemplify the clonal relationships for one diploid and one aneuploid case each. Details of the analysis are described in Materials & Methods.

Figure 4.1 shows the FIShtrees analysis for case D3. Case D3 consisted almost exclusively of diploid cells. The major signal pattern, highlighted in blue, was prevalent in 75.3% of cells; only one single clone was tetraploid (pink node).



**Fig. 4.1: FIShtrees analysis of case D3.** The FIShtrees analysis shows the clonal evolution of tumor D3. FISH patterns are depicted in the following gene order: *COX2*, *DBC2*, *MYC*, *CCND1*, *CDH1*, *TP53*, *HER2*, *ZNF217*. The size of the nodes reflects the frequency of the patterns in the cell population. The blue labeled node indicates the presence of the major diploid clone (pattern 42221112). All other clones had a ploidy of two except for one small tetraploid clone (pattern 44442444) labeled in pink.

Figure 4.2 shows the FISHTrees analysis of sample A3, as an example of a FISHTrees output for an aneuploid case.



**Fig. 4.2: FISHTrees analysis of case A3.** The FISHTrees analysis shows the clonal evolution of tumor A3. FISH patterns are depicted in the following gene order: *COX2*, *DBC2*, *MYC*, *CCND1*, *CDH1*, *TP53*, *HER2*, *ZNF217*. The size of the nodes reflects the frequency of the patterns in the cell population. The blue labeled nodes indicate the presence of the major diploid clone (pattern 22221222) and a smaller diploid clone (pattern 22212222) from which a genome duplication (labeled in pink, pattern 44422444) might have possibly originated leading in turn to the emergence of a triploid clone (pattern 44422224) and its progeny, labeled in green.

Case A3 consisted of a mixture of diploid and triploid cell populations. In the FISHTrees analysis, we could identify one near-tetraploid cell that most likely emerged after genome duplication and potentially gave rise to the triploid clone observed in around 40% of cells of the tumor cell population. The FISH tree highlights the major diploid clone (large blue node) and the diploid cell of origin (small blue node) of the near-tetraploid intermediate (pink) that progressed via additional losses to a triploid cell clone. Further diversification then resulted in a complex branch for the triploid cell population of this case (green nodes).

#### 4.10 Targeted Sequencing Analysis

Pictured below (Table 7) are the gene mutations which were detected for each breast cancer case using targeted sequencing of 563 cancer related genes. Across the diploid cases we found on average of 2.8 mutations, while for the aneuploid samples we detected on average 3.3 mutations per case; thus, the mutation load was not significantly different in the diploid and the aneuploid cases. Overall, the most frequently mutated gene was *TP53*. Strikingly, this mutation was observed in all aneuploid cases, but in none of the diploid cases ( $p=0.002$ ). Four diploid cases (D1, D2, D3, D4), as well as one aneuploid case (A4) had *PIK3CA* mutations. Several other sporadic mutations, for example of *AKT1*, *BRCA1*, *CDH1*, *ERBB1*, and *NF1*, were noted.

	Case No.	<i>TP53</i> mutation	Mutated genes and their mutation sites	Tumor cell frequency (%)
Diploid	D1	No	<i>ERBB2</i> (V777L), <i>FDF6</i> (Y148*), <i>PIK3CA</i> (H1047R)	70
	D2	No	<i>BCOR</i> (R342*), <i>PIK3CA</i> (R108del)	70
	D3	No	<i>FLCN</i> (R194W), <i>PIK3CA</i> (H1047R), <i>PTPRD</i> (R995H)	50
	D4	No	<i>ATP7B</i> (Q1210frameshift), <i>FGFR4</i> (V510M)	70
	D5	No	<i>ABL1</i> (P918L), <i>CSF1R</i> (W159*), <i>MYB</i> (T758M), <i>PIK3CA</i> (E545K)	50
	<b>Averages and Totals</b>	<b>0/5</b>	<b>Average no. of mutations: 2.8</b>	
Aneuploid	A1	Yes	<i>EPH86</i> (S172_173dup), <i>SCN5A</i> (R1195C), <i>TP53</i> (G154frameshift)	40
	A2	Yes	<i>EPHA6</i> (R268C), <i>FGFR4</i> (R411frameshift), <i>TP53</i> (R209frameshift)	60
	A3	Yes	<i>AKT1</i> (E17K), <i>CDH1</i> (Q23*), <i>PDE4DIP</i> (D1912N), <i>TP53</i> (Q192*)	25
	A4	Yes	<i>BRCA1</i> (Q12frameshift), <i>LRP1B</i> (G4525V), <i>PIK3CA</i> (E542K), <i>TP53</i> (Q16frameshift)	50
	A5	Yes	<i>AFF3</i> (T619F), <i>NAT2</i> (D251N), <i>NF1</i> (L2671frameshift), <i>TP53</i> (R213L)	40
	A6	Yes	<i>ORM2</i> (F144frameshift), <i>TP53</i> (R196*)	25
	<b>Averages and Totals</b>	<b>6/6</b>	<b>Average no. of mutations: 3.3</b>	

Table 7: Targeted sequencing analysis results. Each case is displayed with noted gene mutations and according mutation site in parenthesis, presence of *TP53* mutation as well as overall tumor cell frequency for the tumor sample.

## 5 Discussion

This thesis aimed to analyze tumor cell populations for their degree of genomic instability and to gain insight into the causes and consequences of nuclear gross aneuploidy for clonal tumorigenesis on an individual patient level. We therefore investigated copy number changes of eight genes commonly affected in breast cancer and analyzed these first on a single cell level and second in the whole tumor cell population. These changes were portrayed in color displays to visualize the degree of intra-tumor heterogeneity. Furthermore, we determined the spectrum of gene mutations. We also explored whether there is a correlation between ploidy and clinical parameters and overall survival, respectively. All analyses were performed on five diploid and six aneuploid breast cancer samples.

### 5.1 Hybridization Panel I

*CDH1* encodes cadherin 1. Significant changes in expression or structure of one of the components of this protein could lead to junctional disassembly and can, consequently, result in more mobile invasive carcinoma cells.<sup>69</sup> Among the diploid samples, *CDH1* was lost in three out of five samples. Within the aneuploid group, we found it to be gained in two out of six cases and lost in five, supporting the concept of clinically more aggressive, aneuploid tumor cells and, subsequently, a worse clinical outcome in this group.

*TP53*, commonly known as the “guardian of the genome”<sup>81</sup>, encodes a tumor suppressor protein, which pathway increases the fidelity of cell division by preventing errors during the duplication process of the cell and thus preventing carcinogenesis.<sup>83</sup> We found a loss of *TP53* in almost all diploid and all aneuploid tumor samples (4/5 and 6/6). Through the loss of this gene, cells are no longer sufficiently protected from DNA damage. By lack of an adequate DNA-damage-repair process, cells are subject to devolution. This study thereby confirms the key role of *TP53* in breast malignancies.

*HER2* protein overexpression, as a result of an amplified gene, correlates with poor prognosis. Among other factors this is due to increased tumor size, lymph

node metastasis, and high grade in these tumors.<sup>77,79</sup> We found a gain of *HER2* in two samples of each ploidy group (2/5 and 2/6). These numbers are in accordance with the literature, which states a *HER2* protein overexpression in 25-30% of breast cancers.<sup>125</sup> However, it delineates no significant difference between the diploid and the aneuploid group in our study.

*ZNF217* is an oncogene which, when overexpressed, stimulates cell survival and proliferation, cell migration and invasion, and is also involved in invasiveness of cancer cells.<sup>87</sup> It has been shown that gene amplification is associated with DNA aneuploidy <sup>126</sup>, which is supported by our finding of copy number gains in three aneuploid samples, whereas none of the diploid samples exhibited such changes. However, two out six aneuploid samples also showed a loss of *ZNF217*.

## 5.2 Hybridization Panel II

*CCND1* encodes Cyclin D1, which plays a pivotal role in the regulation of the cell cycle and induces cell proliferation.<sup>55,127</sup> We found *CCND1* gained in one of five diploid and three out of six aneuploid breast cancer samples, suggesting a beneficial role in the evolution of genetic instability and thereby fostering a more aggressive tumor cell population.

*DBC2* is a tumor-suppressor gene which inhibits cellular proliferation through the regulation of cell-cycle and apoptosis.<sup>44</sup> In this study, *DBC2* was found to be lost in one of five diploid and five of six aneuploid cases, which emphasizes its suitability as a FISH marker for genomic instability.

*COX2* enhances cell proliferation and mitogenesis of mammary epithelial cells, has proangiogenic effects and is thought to suppress apoptosis.<sup>26,27,31</sup> Here, *COX2* was gained in one of five and five of six samples, respectively, making it another likely candidate as a genomic instability marker.

*MYC* is one of the most important somatically mutated proto-oncogenes known in human cancers.<sup>45,46</sup> Strikingly, *MYC* showed copy number gains in all six aneuploid samples, but in only one diploid case. This makes *MYC* pivotal for a reliable discrimination between diploid and aneuploid breast carcinomas in this thesis.

### 5.3 Intra-tumor heterogeneity

The percentage of cells of a tumor cell population with the same major signal pattern averaged 69.7% for the diploid cases, but a mere 9.5% for the aneuploid cases. Thus, among the diploid cases almost 70% of cells are clones of the same cell. The aneuploid samples display a significantly higher heterogeneity. Nearly 90% of the tumor population show signal patterns that are aberrant from one another, which acquired multiple genetic imbalances over the course of the tumor evolution. These numbers reflect an intra-tumor heterogeneity which is 7.3 times higher in aneuploid tumors than in diploid ones. Leading to the conclusion that the potential for genomic instability fosters the tendency to acquire a genotype which is beneficial for proliferation, metastatic invasiveness, and treatment resistance. This high degree of intra-tumor heterogeneity is also reflected by the instability index, which equals the number of different FISH signal patterns per 100 nuclei and was found to be 3.4 times higher in the aneuploid tumors.

The color displays (Figures 3.1 - 3.11) allow a clear survey of the degree of intra-tumor heterogeneity. Case D4 is one example of a diploid tumor. It presented its major clone with 86% of the cells of the population with a gain of *MYC* (8q) and loss of *DBC2* (8p), indicative of the formation of an isochromosome 8q. An abnormality which has been reported in the literature, has been extensively described and is one of the most frequent cytogenetic alterations in breast cancer.<sup>128</sup> The sample has the lowest instability index (7.8) among our collective and its population is very clonal overall: a great majority of the population displays the same gains and losses. Case D5 is another example of a diploid, genomically stable tumor cell population. 72.3% of cells showed the same major signal pattern with the loss of *CDH1* as the solely observed genetic imbalance. However, we also saw two small clones of 2% each, one of them showing an additional loss of *TP53*, the other an additional gain of *MYC*. Typically, one would expect either clone to be the most dominant in the tumor population, since both exhibit advantages for a malignant cell. One might assume that the smallness of either clone is explained by their only recent development, giving them insufficient time to outcompete the current major clone. Another explanation might be that these



two clones were just not superior to survival compared to the major clone. After all, analysis in this study only comprises eight genes and thus not every aspect of the cell is apparent.

These two diploid cases are in clear contrast with aneuploid case A4, which displays the highest instability index among the aneuploid samples and all cases in this study overall. Its major clone comprises only 9.3% of cells and 27% of cells displayed equal imbalances. Several small sub-populations exhibited aberrant clones with all genes subject to either copy number gains or losses. Overall, the degree of intra-tumor heterogeneity in all aneuploid cases was very high. Case A5 displayed a similarly high degree of heterogeneity as case A4. Situated somewhere in between these extreme examples are cases A1 and A3. Both were found to have a diploid as well as a triploid clone. A1 showed a major clone of 38% which was diploid with two gained oncogenes and a loss of the entire chromosome 17, which contains not only the tumor suppressor gene *TP53*, but also the oncogene *HER2*. The same observation was made in cases D3, A3, and A4, suggesting that the loss of *TP53* is more advantageous to a cell than the gain of *HER2*. However, it may also be an indication that the loss of a chromosome is more easily achieved than the formation of an isochromosome, which requires structural reorganization. On the other hand, findings in cases D2, D4, and A2 lead to the interpretation, that an isochromosome 17q is more beneficial to the cell compared to just a loss of *TP53*.

The cell population of case A1 also included a triploid clone, which showed gains of *MYC*, *CDH1*, and *ZNF217* and, in addition to the loss of chromosome 17, losses of both *COX2* and *DBC2*. The combination of a *MYC* gain and a *DBC2* loss is a reflection of the creation of an isochromosome 8q. Like case A1, A3 also displayed a diploid as well as a triploid clone, both of which account for around 20 % of cells. The diploid clone, with its exclusive loss of *CDH1*, may have been the cancers' origin, acquiring additional imbalances during proliferation. The cell population of case A3 included a near-tetraploid cell with two copies both for *CDH1* and *CCND1*, arisen through genome duplication. Yet more additional losses might then have produced the triploid clone. The possible evolution can be readily appreciated from the FISHtrees analysis in Figure 4.2. The finding of both a diploid

and a triploid clone poses the question why both coexist. Typically, one would expect the triploid clone, with its many gains and losses and therefore higher potential for assertiveness, to outcompete the diploid clone with *CDH1* as its sole loss. If the triploid clone developed much later in the tumor progress, which would be in accordance with our FISHtrees analysis, it might not yet have been able to outperform the diploid clone as major clone.

Related findings were made in other cases. D3 was found to have a major clone of 77% of the population, which showed a gain of *COX2*, and losses of *CDH1* and chromosome 17. A minor clone of 8%, however, showed losses of *CDH1* and *TP53* and a gain of *HER2*, but without a gain of *COX2*. Typically, one would expect the gain of both the oncogenes *COX2* and *HER2* to be most favorable to the tumor cells. The interpretation that suggests itself here is that a gain of *COX2* might in fact be more beneficial to cell proliferation and tumor growth than the gain of *HER2*. On the contrary, this finding may also be explained by an only recent development of this clone and thus its still small size. The major clone in this population is also one of the examples in which we saw a loss of the entire chromosome 17. Thus, the above explanation might also be reasonable here, namely that the loss of a chromosome involves less difficulty than the formation of an isochromosome.

Case D1 is the most heterogeneous one amongst the diploid group with a relatively high instability index of 21.5% and a major signal pattern of only 49.3%. The major clone shows distinct imbalances, with a loss of the three tumor suppressor genes *DBC2*, *CDH1*, and *TP53*, and a gained oncogene *MYC*. Two smaller clones (each 5%) show similar changes, one however lacks the gained *MYC*, the other shows an additional loss of *ZNF217*. Both small clones are suitable examples of the dominance of the major clone over them. The lack of *MYC* gain and the loss of the oncogene *ZNF217* are clearly a disadvantage to the tumor cells' survival. Similarly, the dominance of the 43% clone over the several smaller clones in case A6 is explained. The additional loss of *COX2* is beneficial to the cells' proliferation and survival, hence the 8% and 7% clones are at a disadvantage. The additional gain of *CCND1*, which should be favorable, seems to be balanced out by the loss of *ZNF217*. Merely the non-dominance of the 6%

clone seems unusual, given that the extra gain of *CCND1* should be beneficial. However, this might again be explained by a recent development of this clone.

#### **5.4 Clinical parameters**

We assessed patient age at diagnosis, pre- or postmenopausal status, T stage and tumor size, lymph node status, occurrence of local and distant metastases, hormone receptor status, and overall survival.

Mean age at diagnosis was 62.8 years for the diploid tumor group and 52.0 years for the aneuploid tumor group, a finding which is consistent with a poorer prognosis for patients with younger age at diagnosis.<sup>129</sup> For clinical T stage, four out of five diploid tumors were clinically classified as stage T1, one as T2. However, tumor size revealed all cases to be stage pT1. All aneuploid tumors were graded cT2. Due to tumor size, case A6 was reclassified as pT1 after surgery.

The mean tumor size in the diploid tumor group was 12.6 mm<sup>2</sup>, while it was 24.5 mm<sup>2</sup> in the aneuploid group. This difference can be interpreted as an accelerated growth or a clinically more aggressive behavior of aneuploid tumors. Notably, overall survival outcome did not correlate with tumor size, but with ploidy status. The two patients with the smallest aneuploid tumors (A4 with 20 mm<sup>2</sup> and A6 with 15 mm<sup>2</sup>) died from breast cancer. While the only patient with a diploid tumor to succumb to breast cancer (D2) was also found to have the smallest tumor size (7 mm<sup>2</sup>), the two biggest tumors in this group (D4 and D3 with 18 mm<sup>2</sup> and 20 mm<sup>2</sup>) were still alive after 20 years of follow-up. This leads to the conclusion that rather than tumor size, other factors are significant for patient outcome.

None of the patients with diploid tumors had lymph node metastasis, but half of the patients with aneuploid tumors were lymph node positive. While one of the five patients with a diploid tumor experienced a local relapse, no distant metastases were observed in this group. Whereas three out of six patients with an aneuploid tumor had a local relapse and two patients were diagnosed with distant metastases. These numbers once again illustrate the worse prognosis for patients with aneuploid tumors and are supported by overall survival times. While among

the diploid group no cancer related death was observed, four of the six patients in the aneuploid group succumbed to breast cancer, with a mean survival time of 12.1 years (compared to 20.6 years in the diploid group). Of the four deceased patients, two were lymph node positive, three showed local relapse, and two distant metastases; two tumors were estrogen and progesterone receptor positive, the other two showed a negative receptor status in at least one receptor. Hormone receptor negative status is in general associated with a poorer prognosis compared to a positive hormone receptor status, due to their lack of response to hormone therapy.<sup>130</sup> Comparison of hormone receptor status and overall survival were inconclusive in this study.

Strikingly, three of the four patients who succumbed to their cancer, were treated with only endocrine therapy after surgery, the fourth patient received a combination therapy. The other two patients with aneuploid samples who received adjuvant radiation plus endocrine treatment were still alive at the end of follow-up. While these numbers are too small to statistically substantiate, it seems very likely that all patients with aneuploid tumors would have benefitted from a more aggressive adjuvant combination therapy.

Collectively, we found a good concordance of clinical parameters and overall survival, which is also in agreement with literature.<sup>131</sup>

## 5.5 Gene Mutations

Targeted sequencing of 563 cancer-related genes revealed on average 2.8 mutations per case among the diploid tumor samples and 3.3 mutations per case in the aneuploid group. These numbers are not significantly different. Some of the mutated genes have been extensively described and studied previously, for example *ERBB2*, a synonym for *HER2*.<sup>76</sup> *PIK3CA* has been found to be oncogenic, has been implicated in cervical cancers and has been described in several other cancers, among them breast cancer.<sup>132</sup> Most widely known among the genes for which we observed mutations is *BRCA1*, a high-risk gene which is responsible for about 40% of familial breast cancer.<sup>133</sup> However, these mutations are either singular or found in both diploid and aneuploid tumors.

Strikingly, we observed inactivating mutations of the tumor suppressor *TP53* in all aneuploid tumors, but in none of the diploid tumors. This finding is well matched by the fact that all aneuploid tumors also showed a loss of this particular gene. In synopsis, this leads to the interpretation that the loss of function of *TP53* is required for the establishment of aneuploidy.

## 5.6 Conclusion

Four of six patients with aneuploid tumors succumbed to breast cancer during a 20-year follow-up period, whereas none of the patients with diploid tumors died from the disease. Multiplex interphase FISH revealed an elevated degree of intra-tumor heterogeneity, visible in low percentages of cells with the same major signal pattern and considerably higher genomic instability among the aneuploid tumors. Generally, each of the aneuploid tumors displayed numerous genetic imbalances and notably all showed a gain of the oncogene *MYC*, while this was only observed in 20% of the diploid tumors.

The mutation burden did not present itself as a good prognostic marker, as we did not find a significant difference between both groups. However, all aneuploid cancers revealed a mutation of *TP53*, while none of the diploid tumors displayed such a change.

The results suggest that increased intra-tumor heterogeneity, the presence of *MYC* amplification and *TP53* mutations are associated with aneuploidy, faster disease progression, and poorer overall survival.

## 5.7 Outlook

Owing to wide and effective treatment choices, breast cancer is potentially curable. To avoid over- or under-treatment and to reduce toxic effects of adjuvant therapy, the identification of patients at a very low risk of relapse, meaning potentially curable disease with only loco-regional therapy and of those at a high risk, needing an aggressive treatment, is extremely important. Currently, the

choice of therapy for breast cancer is based on prognostic and predictive factors including disease-independent (e.g. age) and disease-related patient characteristics (e.g. tumor size, axillary lymph node status and histological grade), as well as molecular tumor features (hormone receptor-, Her2/neu, Ki67-status). However, it has become evident, that factors other than these greatly influence clinical course and overall survival. Particularly, deviations from a diploid genome have been correlated with poor prognosis.

In this study, we detected amplification of *MYC*, mutations of *TP53*, and an overall heightened degree of genomic heterogeneity, as suitable predictors for aneuploidy and overall poor prognosis.

Therefore, multiplex FISH and targeted sequencing of clinical specimen can be used to augment the validity of to date used parameters and to provide clinicians with a prognostic tool for individual risk stratification and adapted therapy for breast cancer patients.

On the grounds of the small cohort in this project, additional studies are necessary to validate our findings on more extensive cohorts, especially to endorse statistical results.

*MYC* and *TP53* should be elucidated further for their discriminatory power to differentiate between diploid and aneuploid tumors and their potential as prognostic tools ought to be validated in prospectively-designed studies

## 6 Summary

Clinical course and disease-free survival times in breast cancer are extremely variable. Even at an early stage, the alterations in molecular mechanisms affect tumor growth, progression and metastatic potential and therefore limit the prognostic value of the TNM grading system. Additional specific and sensitive prognostic biomarkers might be helpful to facilitate the challenging situation to accurately define individual risk profiles for patients and to set up individualized treatment. Extensive studies point to aneuploidy as a prognostic determinant. We therefore aimed to further study its molecular basis.

Mutation analyses were performed on five diploid and six aneuploid breast cancers with a minimum follow-up of 20 years. By use of multicolor-FISH we analyzed eight genes, which are commonly differentially expressed in solid human malignancies. Each sample was also examined through targeted sequencing of 563 cancer-related genes. Four out of six patients with aneuploid tumors succumbed to breast cancer, whereas none of the patients with diploid tumors died from this cause. Overall survival time of patients with a diploid tumor was significantly longer compared to the aneuploid group ( $p=0.0048$ ). Multiplex-FISH analysis revealed significantly higher genomic instability as represented by instability index ( $p=0.004$ ) and low cell counts for the dominant signal patterns ( $p=0.0002$ ) in aneuploid tumors, overall representing a considerably higher degree of intra-tumor heterogeneity. Our findings were confirmed by FISHTrees analysis, which displayed diploid cases with fewer nodes and less branching compared to the aneuploid cases. All aneuploid tumors, but only one diploid case, displayed a gain of *MYC*. While sequencing results showed only a slightly higher mutation burden for the aneuploid cancers, notably all aneuploid samples had a loss of *TP53*, which was not observed in any of the diploid tumors.

In addition to certain gene mutations that may function as biomarkers to define risk profiles, such as an amplification of *MYC* and a *TP53* mutation, measurements of ploidy status and degree of genomic heterogeneity by miFISH seem to play a pivotal role in the prediction of disease aggressiveness and long-term outcome of patients with breast cancer and have additional prognostic value to guide treatment decisions.

## 7 Reference list

1. Howlader N, Noone AM, Krapcho M, Garshell J, Miller D, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ CK (eds). Cancer of the Breast - SEER Stat Fact Sheets. <http://seer.cancer.gov/statfacts/html/breast.html>. Accessed February 6, 2018.
2. Bettaieb A, Paul C, Plenchette S, Shan J, Chouchane L, Ghiringhelli F. Precision medicine in breast cancer: reality or utopia? *J Transl Med*. 2017;15(1):139. doi:10.1186/s12967-017-1239-z.
3. Ellsworth RE, Decewicz DJ, Shriver CD, Ellsworth DL. Breast cancer in the personal genomics era. *Curr Genomics*. 2010;11(3):146-161. doi:10.2174/138920210791110951.
4. Fallenius AG, Auer GU, Carstensen JM. Prognostic significance of DNA measurements in 409 consecutive breast cancer patients. *Cancer*. 1988;62(2):331-341.
5. International Agency for Research on Cancer. Breast Cancer Fact Sheet. [http://globocan.iarc.fr/Pages/fact\\_sheets\\_cancer.aspx](http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx). Published 2012. Accessed November 21, 2018.
6. Robert Koch Institut. *Gesundheitsberichterstattung Des Bundes, Heft 25: Brustkrebs.*; 2005.
7. Backe J. Deutsches Ärzteblatt: Brustkrebs beim Mann. 2002; 99(17)
8. Hsieh C-C, Trichopoulos D, Katsouyanni K, Yuasa S. Age at menarche, age at menopause, height and obesity as risk factors for breast cancer: Associations and interactions in an international case-control study. *Int J Cancer*. 1990;46(5):796-800. doi:10.1002/ijc.2910460508.
9. McPherson K, Steel CM, Dixon JM. ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. *BMJ*. 2000;321(7261):624-628. doi:10.1136/BMJ.321.7261.624.
10. Lopez-Garcia MA, Geyer FC, Lacroix-Triki M, Marchió C, Reis-Filho JS. Breast cancer precursors revisited: molecular features and progression pathways. *Histopathology*. 2010;57(2):171-192. doi:10.1111/j.1365-2559.2010.03568.x.
11. Lakhani, S.R., Ellis. I.O., Schnitt, S.J., Tan, P.H., van de Vijver MJ. *WHO Classification of Tumours of the Breast, Fourth Edition.*; 2012.
12. Deutsche Krebsgesellschaft. Interdisziplinäre S3-Leitlinie für die Diagnostik, Therapie und Nachsorge des Mammakarzinomas. 2012.
13. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. C. W. Elston & I. O. Ellis. *Histopathology* 1991; 19; 403-410. AUTHOR COMMENTARY. *Histopathology*. 2002;41(3a):151-151. doi:10.1046/j.1365-2559.2002.14691.x.
14. Whelan TJ, Julian J, Wright J, Jadad AR, Levine ML. Does Locoregional Radiation Therapy Improve Survival in Breast Cancer? A Meta-Analysis. *J Clin Oncol*. 2000;18(6):1220-1229. <http://jco.ascopubs.org/content/18/6/1220.short>. Accessed September 1, 2018.
15. Dent R, Clemons M. Adjuvant trastuzumab for breast cancer. *BMJ*. 2005;331(7524):1035-



1036. doi:10.1136/bmj.331.7524.1035.
16. Goldhirsch A, Winer EP, Coates AS, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann Oncol.* 2013;24(9):2206-2223. doi:10.1093/annonc/mdt303.
  17. Osborne CK. Tamoxifen in the Treatment of Breast Cancer. Wood AJJ, ed. *N Engl J Med.* 1998;339(22):1609-1618. doi:10.1056/NEJM199811263392207.
  18. American Cancer Society. Chemotherapy for Breast Cancer. <https://www.cancer.org/cancer/breast-cancer/treatment/chemotherapy-for-breast-cancer.html>. Accessed December 3, 2018.
  19. Dunst J, Steil B, Furch S, et al. Prognostic significance of local recurrence in breast cancer after postmastectomy radiotherapy. *Strahlenther Onkol.* 2001;177(10):504-510.
  20. Freedman GM, Fowble BL. Local recurrence after mastectomy or breast-conserving surgery and radiation. *Oncology (Williston Park).* 2000;14(11):1561-81; discussion 1581-2, 1582-1584.
  21. National Cancer Institute. Breast Cancer Treatment (PDQ®) - National Cancer Institute. <http://www.cancer.gov/cancertopics/pdq/treatment/breast/Patient/page5>. Published 2014. Accessed December 4, 2018.
  22. Coradini D, Daidone MG. Biomolecular prognostic factors in breast cancer. *Curr Opin Obstet Gynecol.* 2004;16(1):49-55.
  23. Cooper GM. *The Cell.* 2nd ed. Sunderland, MA: Sinauer Associates; 2000.
  24. Heselmeyer-Haddad K, Berroa Garcia LY, Bradley A, et al. Single-cell genetic analysis of ductal carcinoma in situ and invasive breast cancer reveals enormous tumor heterogeneity yet conserved genomic imbalances and gain of MYC during progression. *Am J Pathol.* 2012;181(5):1807-1822. doi:10.1016/j.ajpath.2012.07.012.
  25. Liu CH, Chang SH, Narko K, et al. Overexpression of cyclooxygenase-2 is sufficient to induce tumorigenesis in transgenic mice. *J Biol Chem.* 2001;276(21):18563-18569. doi:10.1074/jbc.M010787200.
  26. Howe L, Subbaramaiah K. Cyclooxygenase-2: a target for the prevention and treatment of breast cancer. *Endocrine-related ....* 2001;8:97-114.
  27. Nolan RD, Danilowicz RM, Eling TE. Role of arachidonic acid metabolism in the mitogenic response of BALB/c 3T3 fibroblasts to epidermal growth factor. *Mol Pharmacol.* 1988;33(6):650-656.
  28. Huang M, Stolina M, Sharma S, et al. Non-small cell lung cancer cyclooxygenase-2-dependent regulation of cytokine balance in lymphocytes and macrophages: up-regulation of interleukin 10 and down-regulation of interleukin 12 production. *Cancer Res.* 1998;58(6):1208-1216.
  29. Brueggemeier RW, Quinn AL, Parrett ML, Joarder FS, Harris RE, Robertson FM. Correlation of aromatase and cyclooxygenase gene expression in human breast cancer specimens. *Cancer Lett.* 1999;140(1-2):27-35. doi:10.1016/S0304-3835(99)00050-6.

30. Marnett LJ. Aspirin and the potential role of prostaglandins in colon cancer. *Cancer Res.* 1992;52(20):5575-5589.
31. Tsujii M, Kawano S, Tsuji S, Sawaoka H, Hori M, DuBois RN. Cyclooxygenase regulates angiogenesis induced by colon cancer cells. *Cell.* 1998;93(5):705-716.
32. Bennett A, Charlier EM, McDonald AM, Simpson JS, Stamford IF, Zebro T. Prostaglandins and breast cancer. *Lancet (London, England).* 1977;2(8039):624-626.
33. Kim HS, Moon H-G, Han W, et al. COX2 overexpression is a prognostic marker for Stage III breast cancer. *Breast Cancer Res Treat.* 2012;132(1):51-59. doi:10.1007/s10549-011-1521-3.
34. Miglietta A, Toselli M, Ravarino N, et al. COX-2 expression in human breast carcinomas: correlation with clinicopathological features and prognostic molecular markers. *Expert Opin Ther Targets.* 2010;14(7):655-664. doi:10.1517/14728222.2010.486792.
35. Nakopoulou L, Mylona E, Papadaki I, et al. Overexpression of cyclooxygenase-2 is associated with a favorable prognostic phenotype in breast carcinoma. *Pathobiology.* 2005;72(5):241-249. doi:10.1159/000089418.
36. Penault-Llorca F, Abrial C, Raoelfils I, et al. Changes and Predictive and Prognostic Value of the Mitotic Index, Ki-67, Cyclin D1, and Cyclo-oxygenase-2 in 710 Operable Breast Cancer Patients Treated with Neoadjuvant Chemotherapy. *Oncologist.* 2008;13(12):1235-1245. doi:10.1634/theoncologist.2008-0073.
37. Sivula A, Talvensaaari-Mattila A, Lundin J, et al. Association of cyclooxygenase-2 and matrix metalloproteinase-2 expression in human breast cancer. *Breast Cancer Res Treat.* 2005;89(3):215-220. doi:10.1007/s10549-004-0714-4.
38. van Nes JGH, de Kruijf EM, Faratian D, et al. COX2 expression in prognosis and in prediction to endocrine therapy in early breast cancer patients. *Breast Cancer Res Treat.* 2011;125(3):671-685. doi:10.1007/s10549-010-0854-7.
39. Sivula A, Talvensaaari-Mattila A, Lundin J, et al. Association of cyclooxygenase-2 and matrix metalloproteinase-2 expression in human breast cancer. *Breast Cancer Res Treat.* 2005;89(3):215-220. doi:10.1007/s10549-004-0714-4.
40. Nakopoulou L, Mylona E, Papadaki I, et al. Overexpression of cyclooxygenase-2 is associated with a favorable prognostic phenotype in breast carcinoma. *Pathobiology.* 2005;72(5):241-249. doi:10.1159/000089418.
41. Chikman B, Vasyanovich S, Lavy R, et al. COX2 expression in high-grade breast cancer: evidence for prognostic significance in the subset of triple-negative breast cancer patients. *Med Oncol.* 2014;31(6):989. doi:10.1007/s12032-014-0989-1.
42. Hamaguchi M, Meth JL, von Klitzing C, et al. DBC2, a candidate for a tumor suppressor gene involved in breast cancer. *Proc Natl Acad Sci U S A.* 2002;99(21):13647-13652. doi:10.1073/pnas.212516099.
43. Mao H, Qu X, Yang Y, et al. A novel tumor suppressor gene RhoBTB2 (DBC2): frequent loss of expression in sporadic breast cancer. *Mol Carcinog.* 2010;49(3):283-289. doi:10.1002/mc.20598.

44. Mao H, Zhang L, Yang Y, et al. RhoBTB2 (DBC2) functions as tumor suppressor via inhibiting proliferation, preventing colony formation and inducing apoptosis in breast cancer cells. *Gene*. 2011;486(1-2):74-80. doi:10.1016/j.gene.2011.07.018.
45. NCBI Gene. MYC v-myc avian myelocytomatosis viral oncogene homolog [Homo sapiens (human)] - Gene - NCBI. <http://www.ncbi.nlm.nih.gov/gene/4609>. Published 2014. Accessed December 4, 2017.
46. Wolfer A, Ramaswamy S. MYC and metastasis. *Cancer Res*. 2011;71(6):2034-2037. doi:10.1158/0008-5472.CAN-10-3776.
47. Grandori C, Cowley SM, James LP, Eisenman RN. The Myc/Max/Mad network and the transcriptional control of cell behavior. *Annu Rev Cell Dev Biol*. 2000;16:653-699. doi:10.1146/annurev.cellbio.16.1.653.
48. Iritani BM, Eisenman RN. c-Myc enhances protein synthesis and cell size during B lymphocyte development. *Proc Natl Acad Sci U S A*. 1999;96(23):13180-13185. doi:10.1073/PNAS.96.23.13180.
49. Johnston LA, Prober DA, Edgar BA, Eisenman RN, Gallant P. Drosophila myc regulates cellular growth during development. *Cell*. 1999;98(6):779-790. doi:10.1016/S0092-8674(00)81512-3.
50. Karn J, Watson J V, Lowe AD, Green SM, Vedeckis W. Regulation of cell cycle duration by c-myc levels. *Oncogene*. 1989;4(6):773-787. <http://www.ncbi.nlm.nih.gov/pubmed/2660073>. Accessed December 4, 2017.
51. Sorrentino V, Drozdoff V, McKinney MD, Zeitz L, Fleissner E. Potentiation of growth factor activity by exogenous c-myc expression. *Proc Natl Acad Sci U S A*. 1986;83(21):8167-8171.
52. Ma L, Young J, Prabhala H, et al. miR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis. *Nat Cell Biol*. 2010;12(3):247-256. doi:10.1038/ncb2024.
53. Chaffer CL, Weinberg R a. A perspective on cancer cell metastasis. *Science*. 2011;331(6024):1559-1564. doi:10.1126/science.1203543.
54. Wolfer A, Wittner BS, Irimia D, et al. MYC regulation of a "poor-prognosis" metastatic cancer cell state. *Proc Natl Acad Sci U S A*. 2010;107(8):3698-3703. doi:10.1073/pnas.0914203107.
55. Fu M, Wang C, Li Z, Sakamaki T, Pestell RG. Minireview: Cyclin D1: normal and abnormal functions. *Endocrinology*. 2004;145(12):5439-5447. doi:10.1210/en.2004-0959.
56. Hosokawa Y, Arnold A. Mechanism of cyclin D1 (CCND1, PRAD1) overexpression in human cancer cells: analysis of allele-specific expression. *Genes Chromosomes Cancer*. 1998;22(1):66-71.
57. Baldin V, Lukas J, Marcote MJ, Pagano M, Draetta G. Cyclin D1 is a nuclear protein required for cell cycle progression in G1. *Genes Dev*. 1993;7(5):812-821. doi:10.1101/GAD.7.5.812.
58. Sherr CJ. The Pezcoller Lecture: Cancer Cell Cycles Revisited. *Cancer Res*. 2000;60(14):3689-3695.
59. Wang C, Li Z, Fu M, Bouras T, Pestell RG. Signal transduction mediated by cyclin D1: from

- mitogens to cell proliferation: a molecular target with therapeutic potential. *Cancer Treat Res.* 2004;119:217-237.
60. Wang C, Li Z, Fu M, Bouras T, Pestell RG. Signal transduction mediated by cyclin D1: from mitogens to cell proliferation: a molecular target with therapeutic potential. *Cancer Treat Res.* 2004;119:217-237.
  61. Bartkova J, Lukas J, Müller H, Lützhøft D, Strauss M, Bartek J. Cyclin D1 protein expression and function in human breast cancer. *Int J Cancer.* 1994;57(3):353-361.
  62. Barnes D, Gillett C. Cyclin D1 in breast cancer. *Breast Cancer Res Treat.* 1998;52:1-52.
  63. Han S, Park K, Bae B-N, et al. Cyclin D1 expression and patient outcome after tamoxifen therapy in estrogen receptor positive metastatic breast cancer. *Oncol Rep.* 2003;10(1):141-144.
  64. Hwang TS, Han HS, Hong YC, Lee HJ, Paik N-S. Prognostic value of combined analysis of cyclin D1 and estrogen receptor status in breast cancer patients. *Pathol Int.* 2003;53(2):74-80. doi:10.1046/j.1440-1827.2003.01441.x.
  65. Mylona E, Tzelepis K, Theohari I, Giannopoulou I, Papadimitriou C, Nakopoulou L. Cyclin D1 in invasive breast carcinoma: favourable prognostic significance in unselected patients and within subgroups with an aggressive phenotype. *Histopathology.* 2013;62(3):472-480. doi:10.1111/his.12013.
  66. Stendahl M, Kronblad A, Rydén L, Emdin S, Bengtsson NO, Landberg G. Cyclin D1 overexpression is a negative predictive factor for tamoxifen response in postmenopausal breast cancer patients. *Br J Cancer.* 2004;90(10):1942-1948. doi:10.1038/sj.bjc.6601831.
  67. Cheng C-W, Liu Y-F, Yu J-C, et al. Prognostic significance of cyclin D1,  $\beta$ -catenin, and MTA1 in patients with invasive ductal carcinoma of the breast. *Ann Surg Oncol.* 2012;19(13):4129-4139. doi:10.1245/s10434-012-2541-x.
  68. NCBI Gene. CDH1 cadherin 1, type 1, E-cadherin (epithelial) [Homo sapiens (human)] - Gene - NCBI. <http://www.ncbi.nlm.nih.gov/gene/999>. Published 2014. Accessed December 4, 2017.
  69. Birchmeier W, Behrens J. Cadherin expression in carcinomas: role in the formation of cell junctions and the prevention of invasiveness. *Biochim Biophys Acta (BBA)-Reviews* .... 1994;1198:11-26.
  70. Sarrió D, Pérez-Mies B, Hardisson D, et al. Cytoplasmic localization of p120ctn and E-cadherin loss characterize lobular breast carcinoma from preinvasive to metastatic lesions. *Oncogene.* 2004;23(19):3272-3283. doi:10.1038/sj.onc.1207439.
  71. Yoder BJ, Wilkinson EJ, Massoll NA. Molecular and Morphologic Distinctions between Infiltrating Ductal and Lobular Carcinoma of the Breast. *Breast J.* 2007;13(2):172-179. doi:10.1111/j.1524-4741.2007.00393.x.
  72. Ozawa M, Baribault H, Kemler R. The cytoplasmic domain of the cell adhesion molecule uvomorulin associates with three independent proteins structurally related in different species. *EMBO J.* 1989;8(6):1711-1717.
  73. Behrens J, Mareel MM, Van Roy FM, Birchmeier W. Dissecting tumor cell invasion:

- epithelial cells acquire invasive properties after the loss of uvomorulin-mediated cell-cell adhesion. *J Cell Biol.* 1989;108(6):2435-2447.
74. Frixen UH, Behrens J, Sachs M, et al. E-cadherin-mediated cell-cell adhesion prevents invasiveness of human carcinoma cells. *J Cell Biol.* 1991;113(1):173-185.
  75. Yoshida R, Kimura N, Harada Y, Ohuchi N. The loss of E-cadherin,  $\alpha$ - and  $\beta$ -catenin expression is associated with metastasis and poor prognosis in invasive breast cancer. *Int J Oncol.* 2001;18(3):513-520.
  76. NCBI Gene. ERBB2 erb-b2 receptor tyrosine kinase 2 [ Homo sapiens (human) ]. <https://www.ncbi.nlm.nih.gov/gene/2064>. Accessed September 27, 2017.
  77. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science.* 1987;235(4785):177-182.
  78. Yarden Y, Sliwkowski M. Untangling the ErbB signalling network. *Nat Rev Mol cell Biol.* 2001;2(February).
  79. Ross J, Fletcher J. The HER-2/neu Oncogene in Breast Cancer: Prognostic Factor, Predictive Factor, and Target for Therapy. *Stem Cells.* 1998;2:413-428.
  80. NCBI Gene. TP53 tumor protein p53 [Homo sapiens (human)] - Gene - NCBI. <http://www.ncbi.nlm.nih.gov/gene/7157>. Published 2014. Accessed December 4, 2018.
  81. Strachan T, Read AP. *Human Molecular Genetics.* Wiley-Liss; 1999.
  82. Vousden KH, Lane DP. P53 in Health and Disease. *Nat Rev Mol Cell Biol.* 2007;8(4):275-283. doi:10.1038/nrm2147.
  83. Riley T, Sontag E, Chen P, Levine A. Transcriptional control of human p53-regulated genes. *Nat Rev Mol Cell Biol.* 2008;9(5):402-412. doi:10.1038/nrm2395.
  84. Roger L, Gadea G, Roux P. Control of cell migration: a tumour suppressor function for p53? *Biol Cell.* 2006;98(3):141-152. doi:10.1042/BC20050058.
  85. Teodoro JG, Parker AE, Zhu X, Green MR. p53-mediated inhibition of angiogenesis through up-regulation of a collagen prolyl hydroxylase. *Science.* 2006;313(5789):968-971. doi:10.1126/science.1126391.
  86. Collins C, Rommens JM, Kowbel D, et al. Positional cloning of ZNF 217 and NABC 1: Genes amplified at 20q13.2 and overexpressed in breast carcinoma. *Proc Natl Acad Sci U S A.* 1998;95(July 1998):8703-8708.
  87. Vendrell J a, Thollet A, Nguyen NT, et al. ZNF217 is a marker of poor prognosis in breast cancer that drives epithelial-mesenchymal transition and invasion. *Cancer Res.* 2012;72(14):3593-3606. doi:10.1158/0008-5472.CAN-11-3095.
  88. Huang G, Krig S, Kowbel D, et al. ZNF217 suppresses cell death associated with chemotherapy and telomere dysfunction. *Hum Mol Genet.* 2005;14(21):3219-3225. doi:10.1093/hmg/ddi352.
  89. Nonet GH, Stampfer MR, Chin K, Gray JW, Collins CC, Yaswen P. The ZNF217 gene amplified in breast cancers promotes immortalization of human mammary epithelial cells. *Cancer Res.* 2001;61(4):1250-1254.

90. Thollet A, Vendrell JA, Payen L, et al. ZNF217 confers resistance to the pro-apoptotic signals of paclitaxel and aberrant expression of Aurora-A in breast cancer cells. *Mol Cancer*. 2010;9:291. doi:10.1186/1476-4598-9-291.
91. Tanner MM, Tirkkonen M, Kallioniemi A, et al. Amplification of chromosomal region 20q13 in invasive breast cancer: prognostic implications. *Clin Cancer Res*. 1995;1(12):1455-1461.
92. Kallioniemi A, Kallioniemi O. Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc ....* 1994;91(March 1994):2156-2160.
93. Alitalo K, Schwab M. Oncogene amplification in tumor cells. *Adv Cancer Res*. 1986;47:235-281.
94. Schwab M, Amler LC. Amplification of cellular oncogenes: a predictor of clinical outcome in human cancer. *Genes Chromosomes Cancer*. 1990;1(3):181-193.
95. Bishop JM. Molecular themes in oncogenesis. *Cell*. 1991;64(2):235-248. doi:10.1016/0092-8674(91)90636-D.
96. Habermann JK, Doering J, Hautaniemi S, et al. The gene expression signature of genomic instability in breast cancer is an independent predictor of clinical outcome. *Int J cancer*. 2009;124(7):1552-1564. doi:10.1002/ijc.24017.
97. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*. 2004;351(27):2817-2826. doi:10.1056/NEJMoa041588.
98. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci*. 2001;98(19):10869-10874. doi:10.1073/pnas.191367098.
99. Sotiriou C, Neo S-Y, McShane LM, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci*. 2003;100(18):10393-10398. doi:10.1073/pnas.1732912100.
100. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002;415(6871):530-536. doi:10.1038/415530a.
101. Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *N Engl J Med*. 2016;375(8):717-729. doi:10.1056/NEJMoa1602253.
102. OncotypeQ. About the Oncotype DX Breast Recurrence Score Test. <http://www.oncotypeiq.com/en-US/breast-cancer/healthcare-professionals/oncotype-dx-breast-recurrence-score/about-the-test>. Accessed September 28, 2017.
103. Harris LN, Ismaila N, McShane LM, et al. Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. *J Clin Oncol*. 2016;34(10):1134-1150. doi:10.1200/JCO.2015.65.2289.
104. Kwa M, Makris A, Esteva FJ. Clinical utility of gene-expression signatures in early stage breast cancer. *Nat Rev Clin Oncol*. 2017;14(10):595-610. doi:10.1038/nrclinonc.2017.74.
105. American Cancer Society. Breast Cancer Gene Expression Tests.

- <https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-gene-expression.html>. Accessed September 28, 2017.
106. Senkus, E., Kyriakides, S., Ohno, S., Penault-Llorca, F., Poortmans, P., Rutgers, E., Zackrisson, S., and Cardoso F. Primary Breast Cancer: ESMO Clinical Practice Guidelines | ESMO. <http://www.esmo.org/Guidelines/Breast-Cancer/Primary-Breast-Cancer>. Accessed March 6, 2018.
  107. Auer G, Eriksson E, Azavedo E, Caspersson T, Wallgren A. Prognostic significance of nuclear DNA content in mammary adenocarcinomas in humans. *Cancer Res*. 1984;44(1):394-396.
  108. Cornelisse CJ, van de Velde CJH, Caspers RJC, Moolenaar AJ, Hermans J. DNA ploidy and survival in breast cancer patients. *Cytometry*. 1987;8(2):225-234. doi:10.1002/cyto.990080217.
  109. Kronenwett U, Huwendiek S, Ostring C, et al. Improved grading of breast adenocarcinomas based on genomic instability. *Cancer Res*. 2004;64(3):904-909.
  110. Habermann JK, Doering J, Hautaniemi S, et al. The gene expression signature of genomic instability in breast cancer is an independent predictor of clinical outcome. *Int J Cancer*. 2009;124(7):1552-1564. doi:10.1002/ijc.24017.
  111. National Institutes of Health. The National Institutes of Health (NIH) Consensus Development Program: Adjuvant Therapy for Breast Cancer. In: *The National Institutes of Health Consensus Development Conference: Adjuvant Therapy for Breast Cancer*. ; 2001:1-152. <http://consensus.nih.gov/2000/2000AdjuvantTherapyBreastCancer114html.htm>. Accessed December 4, 2018.
  112. Haibe-Kains B, Desmedt C, Piette F, et al. Comparison of prognostic gene expression signatures for breast cancer. *BMC Genomics*. 2008;9(1):394. doi:10.1186/1471-2164-9-394.
  113. Olivotto IA, Bajdik CD, Ravdin PM, et al. Population-Based Validation of the Prognostic Model ADJUVANT! for Early Breast Cancer. *J Clin Oncol*. 2005;23(12):2716-2725. doi:10.1200/JCO.2005.06.178.
  114. Nalejska E, Mączyńska E, Lewandowska MA. Prognostic and predictive biomarkers: tools in personalized oncology. *Mol Diagn Ther*. 2014;18(3):273-284. doi:10.1007/s40291-013-0077-9.
  115. NCBI. Genome Decoration Page. <https://www.ncbi.nlm.nih.gov/genome/tools/gdp>. Accessed January 5, 2018.
  116. Gertz EM, Chowdhury SA, Lee W-J, et al. FISHtrees 3.0: Tumor Phylogenetics Using a Ploidy Probe. *PLoS One*. 2016;11(6):e0158569. doi:10.1371/journal.pone.0158569.
  117. Chowdhury SA, Gertz EM, Wangsa D, et al. Inferring models of multiscale copy number evolution for single-tumor phylogenetics. *Bioinformatics*. 2015;31(12):i258-i267. doi:10.1093/bioinformatics/btv233.
  118. Boikos SA, Pappo AS, Killian JK, et al. Molecular Subtypes of *KIT/PDGFR*A Wild-Type

- Gastrointestinal Stromal Tumors. *JAMA Oncol.* 2016;2(7):922.  
doi:10.1001/jamaoncol.2016.0256.
119. Broad Institute. GATK | Best Practices. <https://software.broadinstitute.org/gatk/best-practices/>. Published 2017. Accessed September 29, 2017.
  120. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-1760. doi:10.1093/bioinformatics/btp324.
  121. Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin).* 2012;6(2):80-92. doi:10.4161/fly.19695.
  122. Liu X, Jian X, Boerwinkle E. dbNSFP: A lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum Mutat.* 2011;32(8):894-899. doi:10.1002/humu.21517.
  123. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001;29(1):308-311. <http://www.ncbi.nlm.nih.gov/pubmed/11125122>. Accessed March 6, 2018.
  124. Forbes SA, Beare D, Boutselakis H, et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* 2017;45(D1):D777-D783. doi:10.1093/nar/gkw1121.
  125. Kroese M, Zimmern RL, Pinder SE. HER2 status in breast cancer--an example of pharmacogenetic testing. *J R Soc Med.* 2007;100(7):326-329. doi:10.1177/014107680710000715.
  126. Tanner MM, Tirkkonen M, Kallioniemi A, et al. Amplification of chromosomal region 20q13 in invasive breast cancer: prognostic implications. *Clin Cancer Res.* 1995;1(12):1455-1461.
  127. Hosokawa Y, Arnold A. Mechanism of cyclin D1 (CCND1, PRAD1) overexpression in human cancer cells: Analysis of allele-specific expression. *Genes, Chromosom Cancer.* 1998;22(1):66-71. doi:10.1002/(SICI)1098-2264(199805)22:1<66::AID-GCC9>3.0.CO;2-5.
  128. Mertens F, Johansson B, Mitelman F. Isochromosomes in neoplasia. *Genes Chromosomes Cancer.* 1994;10(4):221-230.
  129. El Saghir NS, Seoud M, Khalil MK, et al. Effects of young age at presentation on survival in breast cancer. *BMC Cancer.* 2006;6(1):194. doi:10.1186/1471-2407-6-194.
  130. Dunnwald LK, Rossing MA, Li CI. Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. *Breast Cancer Res.* 2007;9(1):R6. doi:10.1186/bcr1639.
  131. Soerjomataram I, Louwman MWJ, Ribot JG, Roukema JA, Coebergh JWW. An overview of prognostic factors for long-term survivors of breast cancer. *Breast Cancer Res Treat.* 2008;107(3):309-330. doi:10.1007/s10549-007-9556-1.
  132. NCBI Gene. PIK3CA phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha [ Homo sapiens (human) ]. <https://www.ncbi.nlm.nih.gov/gene/5290>. Accessed August 6, 2019.
  133. NCBI Gene. BRCA1 BRCA1, DNA repair associated [ Homo sapiens (human) ]. <http://www.ncbi.nlm.nih.gov/gene/672>. Accessed March 5, 2018.



## **8 Appendix**

### **8.1 List of tables and figures**

#### Figures

Fig. 1: Gene models, page 24-25

Fig. 2.1: DNA histogram of cases D1, page 41

Fig. 2.2: DNA histogram of cases D2, page 43

Fig. 2.3: DNA histogram of cases D3, page 45

Fig. 2.4: DNA histogram of cases D4, page 46

Fig. 2.5: DNA histogram of cases D5, page 48

Fig. 2.6: DNA histogram of cases A1, page 49

Fig. 2.7: DNA histogram of cases A2, page 51

Fig. 2.8: DNA histogram of cases A3, page 52

Fig. 2.9: DNA histogram of cases A4, page 54

Fig. 2.10: DNA histogram of cases A5, page 55

Fig. 2.11: DNA histogram of cases A6, page 57

Fig. 3.1: Color displays of miFISH analysis for case D1, page 42

Fig. 3.2: Color displays of miFISH analysis for case D2, page 44

Fig. 3.3: Color displays of miFISH analysis for case D3, page 45

Fig. 3.4: Color displays of miFISH analysis for case D4, page 47

Fig. 3.5: Color displays of miFISH analysis for case D5, page 48

Fig. 3.6: Color displays of miFISH analysis for case A1, page 50

Fig. 3.7: Color displays of miFISH analysis for case A2, page 51

Fig. 3.8: Color displays of miFISH analysis for case A3, page 53

Fig. 3.9: Color displays of miFISH analysis for case A4, page 54

Fig. 3.10: Color displays of miFISH analysis for case A5, page 56

Fig. 3.11: Color displays of miFISH analysis for case A6, page 57

Fig. 4.1: FISHtrees analysis of case D3, page 60

Fig. 4.2: FISHtrees analysis of case A3, page 61

## Tables

Table 1.1: UICC (Union Internationale Contre le Cancer), 7<sup>th</sup> edition - Classification of primary tumor (T), page 3

Table 1.2: UICC (Union Internationale Contre le Cancer), 7<sup>th</sup> edition - Regional lymph nodes Clinical (cN), page 4

Table 1.3: UICC (Union Internationale Contre le Cancer), 7<sup>th</sup> edition - Regional lymph nodes Pathological (pN), page 5

Table 1.4: UICC (Union Internationale Contre le Cancer), 7<sup>th</sup> edition – tumor stage for breast cancer, page 6

Table 2: Summary of clinical data

	Case No.	Age (years)	pre- or post-menopausal	Clinical T Stage	Tumor Size (mm <sup>2</sup> )	ER Status	PR Status	Lymph Node Status (pos. nodes in nodes resected)	Treatment	Local relapse	Distant mets 1	Distant mets 2	Alive/ Dead after 20 years (years after diagnosis)	Overall Survival in years
Diploid	D1	66	Post	T1	7	Pos.	Pos.	Neg. (0/10)	R+E	0	0	0	Alive	22.4
	D2	72	Post	T1	7	Neg.	Neg.	Neg. (0/9)	E	0	0	0	Dead by CCVD (15)	15
	D3	45	Pre	T1	20	Pos.	Pos.	Neg. (0/15)	RE	ipsi-lateral breast	0	0	Alive	22
	D4	62	Post	T2	18	Neg.	Neg.	Neg. (0/3)	E	0	0	0	Alive	21.6
	D5	69	Post	T1	11	Pos.	Pos.	Neg. (0/11)	RE	0	0	0	Alive	21.6
	Averages and Totals	62.8			12,6	3/5	3/5	0/5		1/5	0/5	0/5	None Dead by BC	20.6
Aneuploid	A1	40	Pre	T2	30	Neg.	Neg.	Neg. (0/11)	RE	0	0	0	Alive	22.1
	A2	60	Post	T2	30	Pos.	Pos.	Pos. (1/3)	E	lymph nodes	bones	liver	Dead by BC (8.5)	8.5
	A3	53	Post	T2	27	Pos.	Neg.	Pos. (14/18)	RE	0	0	0	Dead by BC (1.6)	1.6
	A4	49	Pre	T2	20	Pos.	Pos.	Neg. (0/4)	E	ipsi-lateral breast	media-stinum	0	Dead by BC (4.8)	4.8
	A5	64	Post	T2	25	Pos.	Neg.	Pos. (1/10)	RE	0	0	0	Alive	21.6
	A6	46	Post	T2	15	Neg.	Neg.	Neg. (0/10)	E	lymph nodes	0	0	Dead by BC (13.9)	13.9
	Averages and Totals	52.0			24,5	4/6	2/6	3/6		3/6	2/6	1/6	4 Dead by BC	12.1

**Table 2: Summary of clinical data. Overview of clinical data of all patients included in this study. Data was collected during treatment and follow-up, with a minimum follow up of 20 years for each patient. It comprises information about patient age at diagnosis, pre- or postmenopausal status at time of diagnosis. Clinical T stage, tumor size in mm<sup>2</sup>, estrogen receptor (ER), progesterone receptor (PR) and lymph node status; all patients had surgery and in addition received either adjuvant endocrine therapy (E) or a combination of radiation and endocrine therapy (RE). The table also includes occurrence and location of local relapse and distant metastases, if any. Overall survival is denoted in years after diagnosis and includes cause of death (BC=breast cancer; CCVD=chronic cardiovascular disease).**

Table 3: Detailed list of gene probes and BAC (Bacterial Artificial Chromosome) clones, page 24

Table 4: Hybridization panels I and II, page 29

Table 5: Comparison of gain and loss frequency for diploid and aneuploid tumors, page 39

Table 6: Major signal pattern (genes in the following order: Ploidy-*COX-DBC-MYC-CCND-CDH-TP53-HER-ZNF*) and gene gains and losses for each case, page 40

Table 7: Targeted sequencing analysis results, page 63

Table 8: Summary of clinical data, multiplex FISH and mutation analysis

Case No.	Clinical Data										Multiplex FISH Analysis					Targeted Sequencing Analysis				
	Age (yrs)	pre- or post-menopausal	Clinical T Stage	Tumor Size (mm <sup>3</sup> )	ER Status	PR Status	Lymph Node Status (pos. nodes in nodes resected)	Local relapse	Distant mets1	Distant mets2	Alive/Dead after 20 years (months alive after diagnosis)	Overall Survival in days	Instability Index	% cells with major signal pattern	Major signal patterns in following order Ploidy-COX2-DBCC2-MYC-CCND1-CDH1-TP53-HER2-ZNF217	Gains	Losses	TP53 mutation	Mutated genes and their mutation sites	Tumor cell frequency %
D1	66	Post	T1	7	Pos.	Pos.	Neg. (0/10)	0	0	0	8058	21.5	49.3	2-2-1-4-2-1-1-2-2	MYC	DBCC2, CDH1, TP53	No	ERBB2 (V777L), FDFE (V148*), PIK3CA (H1047R)	70	175.78
D2	72	Post	T1	7	Neg.	Neg.	Neg. (0/9)	0	0	0	5479	15.3	67.0	2-2-2-2-3-2-1-3-2	CCND1, HER2	TP53	No	BCOR (R342*), PIK3CA (R1086del)	70	393.45
D3	45	Pre	T1	20	Pos.	Pos.	Neg. (0/15)	0	0	0	8012	11.8	75.3	2-4-2-2-2-1-1-1-2	COX2	CDH1, TP53, HER2	No	FLCN (R194W), PIK3CA (H1047R), PTPRD (R995H)	50	381.16
D4	62	Post	T2	18	Neg.	Neg.	Neg. (0/3)	0	0	0	7886	7.8	84.5	2-2-2-2-2-2-1-3-2	HER2	TP53	No	ATP7B (Q1210frameshift), FGFR4 (V510M)	70	79.23
D5	69	Post	T1	11	Pos.	Pos.	Neg. (0/11)	0	0	0	7886	14.3	72.3	2-2-2-2-2-1-2-2-2		CDH1	No	ABL1 (P918L), CSF1R (W159*), MYB (T758M), PIK3CA (E545K)	50	35.36
Averages and Totals	62.8			12.6	3/5	3/5	0/5	1/5	0/5	0/5	7464	14.1	69.7		1	1.8	0/5	average no. of mutations: 2.8		213.00
A1	40	Pre	T2	30	Neg.	Neg.	Neg. (0/11)	0	0	0	8049	35.5	44.4	(2-3-2-4-2-2-1-1-2)	COX2, MYC, CDH1, ZNF217	COX2, DBCC2, TP53, HER2	Yes	EPH6E (S172_175dup), SKNSA (R1195C), TP53 (G154frameshift)	40	165.04
A2	60	Post	T2	30	Pos.	Pos.	Pos. (1/3)	lymph nodes	bones	liver	3090	44.8	13.0	3-3-2-12-6-2-2-4-4	MYC, CCND1, HER2, ZNF217	DBCC2, CDH1, TP53	Yes	EPHA6 (R268C), FGFR4 (R411frameshift), TP53 (R209frameshift)	60	54.37
A3	53	Post	T2	27	Pos.	Neg.	Pos. (14/18)	0	0	0	584	39.0	37.0	(2-2-2-2-2-1-2-2-2)	COX2, DBCC2, MYC	CCND1, CDH1, TP53, HER2	Yes	AKT1 (E17K), CDH1 (Q23*), PDE4DIP (D1912N), TP53 (Q192*)	25	204.97
A4	49	Pre	T2	20	Pos.	Pos.	Neg. (0/4)	ipsilateral breast	mediastinum	0	1782	82.0	9.3	5-6-2-8-8-5-3-3-8	COX2, MYC, CCND1, ZNF217	DBCC2, TP53, HER2	Yes	BRCA1 (Q12frameshift), LRP1B (G4525V), PIK3CA (E542K), TP53 (Q16frameshift)	50	127.93
A5	64	Post	T2	25	Pos.	Neg.	Pos. (1/10)	0	0	0	7896	49.3	15.5	3-4-1-4-3-2-2-4-2	COX2, MYC, HER2	DBCC2, CDH1, TP53, ZNF217	Yes	AFF3 (T619F), MAT2 (D251N), NF1 (L2671frameshift), TP53 (R213L)	40	179.93
A6	46	Post	T2	15	Neg.	Neg.	Neg. (0/10)	lymph nodes	0	0	5112	36.5	11.3	3-6-2-4-3-2-2-3-3	COX2, MYC	DBCC2, CDH1, TP53	Yes	ORM2 (F144frameshift), TP53 (R196*)	25	41.15
Averages and Totals	52.0			24.5	4/6	2/6	3/6	3/6	2/6	1/5	4419	47.9	11.0		3.3	3.5	6/6	average no. of mutations: 3.3		128.90

**Table 8: Summary of clinical data, multiplex FISH and mutation analysis. Patient age (years) at diagnosis, pre- or postmenopausal status at time of diagnosis. Clinical T stage, tumor size in mm<sup>2</sup>, estrogen receptor (ER), progesterone receptor (PR) and lymph node status; all patients had surgery and in addition received either adjuvant endocrine therapy (E) or a combination of radiation and endocrine therapy (RE). Occurrence and location of local relapse and distant metastases, if any. Overall survival is denoted in years after diagnosis and includes cause of death (BC=breast cancer; CCVD=chronic cardiovascular disease). The instability index is calculated as number of signal patterns x 100 / number of cells analyzed (see Material and Methods). Major signal patterns: Copy numbers higher than the ploidy (first number) are bolded (gains), copy numbers lower than the ploidy are in light gray (losses). Patterns in parenthesis are diploid clones observed as a subpopulation in an aneuploid lesion and are therefore not included in the average percentage of the major signal pattern clone for the aneuploid lesions. Percentages for the diploid and aneuploid signal pattern clones in cases A1 and A3 are calculated separately within the diploid and the aneuploid cell populations of these cases. Gains and losses tabulated in this table were derived from all major signal patterns listed for the case. True mutations were filtered from sequencing data as being listed in cosmic or frameshift or stop mutations.**

## 8.2 Preparations

### Nick Translation

#### **0.5mM ACG's Mix**

1  $\mu$ l of dATP (conc. 10  $\mu$ moles/100 $\mu$ l)

1  $\mu$ l of dCTP (conc. 10  $\mu$ moles/100 $\mu$ l)

1  $\mu$ l of dGTP (conc. 10  $\mu$ moles/100 $\mu$ l)

197  $\mu$ l sterile H<sub>2</sub>O

#### **0.1mM dTTP**

1  $\mu$ l of dTTP (conc. 10  $\mu$ moles/100 $\mu$ l)

99  $\mu$ l sterile H<sub>2</sub>O

#### **Labeled dUTP Mix**

labeled dUTP	1.25 $\mu$ l
0.1 mM dTTP	3.75 $\mu$ l
dH <sub>2</sub> O	5.0 $\mu$ l
Final volume	10.0 $\mu$ l

#### **10X NT Buffer**

Tris-HCl, 1M, pH 8.0	500 $\mu$ l	[final concentration 0.5 M]
MgCl <sub>2</sub> , 0.5 M	100 $\mu$ l	[final concentration 50 mM]
BSA, 10 mg/ml	50 $\mu$ l	[final concentration 0.5 mg/ml]
Sterile H <sub>2</sub> O	350 $\mu$ l	
Final volume	1000 $\mu$ l	

#### **DNase I stock solution, 1mg/ml**

DNase I	10 mg	
NaCl, 1 M	1.5 ml	[final concentration 0.15 M]
Glycerol	5.0 ml	[final concentration 50%]
Sterile water	3.5 ml	

**0.1M  $\beta$ -mercaptoethanol (BME) stock**

99% solution (14.4 M)	34.7 $\mu$ l
Sterile water	bring up to 5 ml

**Master Mix (Nick translation)**

10X NT buffer	10 $\mu$ l
0.5mM ACG Mix	10 $\mu$ l
0.1M BME stock	10 $\mu$ l
Final volume	30 $\mu$ l

**Procedure:** For each DNA sample add to an Eppendorf tube:

X $\mu$ l	DNA (equivalent to 2ug)
30 $\mu$ l	Master Mix
10 $\mu$ l	labeled dUTP Mix
X $\mu$ l	sterile dH <sub>2</sub> O (adjust to a final reaction volume of 100 $\mu$ l)
2 $\mu$ l	DNA Polymerase
2-4 $\mu$ l	DNase I (diluted 1:1000)
100 $\mu$ l	Final volume

**Precipitation and Hybridization****Master Mix (Precipitation)**

Dextran sulfate, 50%	40 ml	[final concentration 20%]
20X SSC	20 ml	[final concentration 4X SSC]
Sterile dH <sub>2</sub> O	40 ml	
Final volume	100 ml	

**70% Formamide/ 2X SSC**

20X SSC	10 ml	[final concentration 2X SSC]
Sterile dH <sub>2</sub> O	20 ml	
Deionized formamide	70 ml	[final concentration 70%]
Final volume	100 ml	



## 9 Acknowledgement

Ich danke Herrn Prof. Dr. med. T. Keck für die Möglichkeit, meine Doktorarbeit in seiner Klinik durchführen zu können. Herrn Prof. Dr. Dr. med. Jens Habermann danke ich für die Vergabe des Themas und die Betreuung meiner Arbeit in Lübeck und in Washington D.C. ebenso, wie für Gespräche und Anregungen, die meine Forschungsarbeit entscheidend geprägt haben.

Prof. Dr. Thomas Ried hat mich sehr großzügig und fachkundig in seinem Labor in Maryland aufgenommen und unterstützt; allen anderen Mitarbeitern der Arbeitsgruppe danke ich für die freundschaftliche Arbeitsatmosphäre im Forschungslabor und die entgegengebrachte Hilfsbereitschaft. Mein Dank gilt dabei insbesondere Dr. Kerstin Heselmeyer-Haddad sowie Leanora S. Hernandez, die mich mit viel Geduld in die FISH-Methodik eingearbeitet und fortlaufend unterstützt haben. Buddy Chen und Yue Hu danke ich für den IT Support. Dr. Rüdiger Meyer möchte ich für die anregenden Gespräche und den Austausch auch außerhalb des Labors danken.

Bei Prof. Dr. Dr. Gerd Auer vom Karolinska Institut möchte ich mich für das zur Verfügungstellen der Präparate bedanken. Bei Anders Höög bedanke ich mich für die histopathologische Befundung der ausgewählten Präparate.

Meinen Freunden möchte ich mich für die fortlaufende Unterstützung und Motivation danken. Besonderer Dank gilt zudem meiner Familie für das Ermöglichen meines Forschungsaufenthaltes in den USA sowie für die Unterstützung und den Glauben an meine Fähigkeiten. Ohne sie wäre dieses Projekt nicht möglich gewesen.

## 10 Curriculum Vitae

### Persönliche Daten

Vor- und Zunahme	Johanna Oltmann
Geburtsdatum	06.12.1990
Geburtsort	Wilhelmshaven

### Ausbildung

10/2010 – 06/2017	<b>Universität zu Lübeck, Lübeck</b> Studium der Humanmedizin 1. Abschnitt der Ärztlichen Prüfung 2012 2. Abschnitt der Ärztlichen Prüfung 2016 3. Abschnitt der Ärztlichen Prüfung 2017
2003 – 2010	<b>Lothar-Meyer-Gymnasium, Varel</b>

### Praktisches Jahr

05/16 – 09/16	<b>Klinik für Neurochirurgie</b> , Universitätsklinikum Schleswig-Holstein, Campus Lübeck
09/16 – 10/16	<b>Department of Infectious Diseases</b> , The Royal London Hospital, Queen Mary University of London
10/16 – 12/16	<b>Medizinische Klinik III</b> , Hämatologie, internistische Onkologie, Immunologie und Palliativmedizin, Sana Kliniken Lübeck
12/16 – 04/17	<b>Klinik für Allgemeinchirurgie und Gefäßchirurgie</b> , Universitätsklinikum Schleswig-Holstein, Lübeck

## Promotion und Berufstätigkeit

08/2014	Beginn Promotion
10/2014 – 03/2015	Forschungsaufenthalte am National Cancer Institute,
08/2015 – 09/2015	National Institutes of Health, Bethesda, MD, USA unterstützt durch ein PROMOS-Stipendium des DAAD
Seit 01/2018	Assistenzärztin an der Klinik für Neurochirurgie Universitätsklinikum Schleswig-Holstein, Campus Lübeck

## Veröffentlichung

Oltmann J\*, Heselmeyer-Haddad K\*, Hernandez LS, Torres I, Hu Y, Doberstein N, Killian JK, Petersen D, Zhu YJ, Edelman DC, Meltzer PS, Schwartz R, Gertz EM, Schäffer AA, Auer G, Habermann JK, Ried T

**Aneuploidy, TP53 mutation, and amplification of MYC correlate with increased intratumor heterogeneity and poor prognosis of breast cancer patients.** *Genes Chromosomes Cancer*. 2018;57(4).

\* = geteilte Erstautorenschaft