MASTER'S THESIS

MABIO 5900

May 2018

Copy number alterations in breast carcinomas – comparison of different methods for assessment

Geir Andre Kongelf

Department of Cancer Genetics

Institute of Cancer Research

Oslo University Hospital

OsloMet – Oslo Metropolitan University

Copy number alterations in breast carcinomas – comparison of different methods for assessment

Thesis submitted for the master degree in

Biomedicine (60 ECTS)

by

Geir Andre Kongelf

Fakultet for Helsefag, Institutt for naturvitenskapelige helsefag

Department of Cancer Genetics

Institute of Cancer Research

Oslo University Hospital

May 2018

OsloMet – Oslo Metropolitan University

Acknowledgements

The work presented in this thesis was carried out part time at the Department of Cancer Genetics, Institute of Cancer Research at the Norwegian Radium Hospital from 2016-2018. I would like to thank the Department of Pathology at Oslo University Hospital for allowing me the time and financial support to pursue a long standing personal goal and all my colleagues for putting in an extra effort for nearly three years to make up for my absence.

I am truly grateful to my supervisor Hege G. Russnes for accepting me in the program and handing me such an interesting challenge and the support to carry it through. Your enthusiasm and endless knowledge is a true inspiration and being supervised by you has been a humbling experience.

I would also like to extend sincere appreciation to my co-supervisor Inga H. Rye for ever present support in the lab, introduction to microscope techniques and analysis, continuing feedback on all relevant matters, and essential help with R-programming and illustrations in this thesis.

Thank you to Arne V. Pladsen for doing the analysis of sequencing data and providing the copy number alignments, to Helen Vålerhaugen and Bente Risberg for performing the actual sequencing, and to Ole Christian Lingjærde for interesting statistics discussions.

The work in this thesis would not have been possible without the love and support from my family for which I'm truly thankful. A special thank you to my dearest Wibeke and wonderful daughter Alina for filling my life with happiness and providing constant encouragement and motivation through it all.

Abstract

Breast cancer is a disease dominated by copy number alterations, but only *HER2* copy number assessment analyzed by fluorescence *in situ* hybridization (FISH) is today included in the guidelines for breast cancer diagnostics. The anticipated inclusion of copy number estimation for additional DNA segments would be time consuming and cumbersome if analyzed with FISH. The development of Next Generation Sequencing (NGS) has opened the possibility of copy number assessment in multiple regions by targeted sequencing. Introduction of a panel aiming at detecting copy number alterations in breast cancer would need validation. We employed FISH analysis and designed and validated 17 bacterial artificial chromosomes (BAC) probes for selected chromosomal regions which are often subject to copy number alterations in breast cancer. Copy number analysis performed on formalin-fixes paraffin-embedded (FFPE) samples are subject to potential loss of chromosomal material due to sectioning artifacts. NGS uses bulk tumor samples and does not suffer from this effect. We therefore wanted to see if there is a systematic relationship in copy number difference between estimates using FFPE (i.e. sectioned cells) and estimates using whole cell imprints to identify a possible adjusting factor. Linear regression generated coefficients of large variation for the cell lines. We performed FISH analysis with the 17 probes on both imprints and FFPE sections of the breast cancer cell lines HCC1954 and HCC2218 and found that there were in general lower copy number estimates from FFPE sections. The average difference was about one copy number for the most heterogeneous cell line and somewhat less for the other, but there were large variations between the different regions in both cell lines so no specific adjusting factor was found. We then used the same validated probes in FISH analysis of FFPE tumor samples from 12 patients diagnosed with breast cancer and compared the copy number estimates with NGS estimates for the same regions. The copy number estimates by NGS largely aligned with estimates by FISH, but with certain notable discrepancies, reflecting the NGS test still being in a state of development. Further validation of the NGS test against the current gold standard for copy number assessment, FISH, combined with improved data analysis, is a crucial step preceding implementation in routine diagnostics.

Sammendrag

Brystkreft er en sykdom som er dominert av kopitallsendringer, men i tråd med dagens retningslinjer for brystkreftdiagnostikk utføres kun *HER2* kopitallsanalyse ved FISH. En utvidelse av disse retningslinjene til å inkludere flere DNA områder vil være tidkrevende med FISH analyse, men utviklingen av NGS og forestående introduksjon i diagnostikken vil muliggjøre kopitallsestimering ved målrettet sekvensering av en rekke områder i DNA. Før implementering av en slik test i diagnsotikken må den valideres opp mot FISH som er dagens gullstandard for kopitallanalyse. Vi designet og validerte 17 BAC prober med FISH for utvalgte kromosomale regioner som ofte er utsatt for kopitallsendringer i brystkreft. Ved analyse av kopitallsendringer på FFPE materiale er det risiko for tap av DNA materiale grunnet snitting av blokken. NGS bruker biter av hel tumor og er ikke utsatt for et slikt tap av materiale. Vi ønsket derfor å se om det finnes et systematisk forhold mellom kopitallsestimater for snittet FFPE materiale og hele celler, såkalte imprints og mulighet for beregning av en kompensasjonsfaktor. Vi utførte FISH analyse med de 17 validerte probene på imprints og FFPE av de to brystkreftcellelinjene HCC1954 og HCC2218 og fant at det generelt er lavere kopitall i FFPE celler. I gjennomsnitt utgjorde denne forskjellen omtrent 1 kopitall for den mest heterogene av cellelinjene og noe mindre for den andre, men det var stor variasjon mellom regionene i begge cellelinjer, slik at en systematisk forskjell mellom imprints og FFPE ikke ble funnet. Vi utførte deretter FISH analyse på tumorprøver fra 12 pasienter som var diagnostisert med brystkreft og sammenlignet kopitallsestimatene med estimater av kopitall etter NGS. Vi fant at kopitallsestimatene fra NGS i stor grad stemte overens med etimater fra FISH analyse, men med noen påfallende unntak som kan forklares med at NGS analysen for kopitallsestimering fortsatt er under utvikling. Videre validering av NGS analysen for kopitallsestimering, samt videreutvikling av dataanalysepakken er avgjørende neste steg før implementering i dignostikken.

Abbreviations

ATCC	American Type Culture Collection
BAC	bacterial artificial chromosome
BAF	B-allele-frequency
BCS	breast conserving surgery
BER	base excision repair
BFB	breakage-fusion-bridge
BIR	break-induced replication
BRCA1/2	breast cancer 1/2
CN	copy number
CNA	copy number aberrations
Cent17	centromere 17
DAPI	4’,6-diamidino-2-phenylindole
DC	ductal carcinoma
DCIS	ductal carcinoma in situ
DDR	DNA damage response
DNA	deoxyribonucleic acid acid
DSB	double strand breaks
E.coli	Escherichia coli
EGFR	epidermal growth factor receptor
ER	estrogen receptor
FFPE	formalin-fixed paraffin-embedded
FISH	fluorescence in situ hybridization
HER2	human epidermal growth factor receptor 2
HR	homologous recombination
IC	integrated cluster
IHC	immunohistochemistry
ISH	in situ hybridization
LB	Luria Bertani
LOH	loss of heterozygosity
mRNA	messenger ribonucleic acid
NBCG	Norwegian Breast Cancer Group
NER	nucleotide excision repair
NGS	next generation sequencing
NHEJ	non-homologous end joining
PCF	piecewise constant fitting
PCR	polymerase chain reaction
PR	progesterone receptor
RT-qPCR	quantitative reverse transcriptase polymerase chain reaction
SERM	selective estrogen receptor modulators
SNP	single nucleotide polymorphism
SSB	single strand breaks
TDLU	terminal ductal lobular units

Table of Contents

1. Introduction

1.1 Epidemiology

Breast cancer is the second most prevalent type of cancer worldwide, after lung cancer, with approximately 1,67 million new cases in 2012 (1). This represents 25% of all female cancers, excluding non-melanoma skin cancer. The incidence of breast cancer varies four-fold worldwide and with a higher incidence in high-income countries like Northern America, Australia and Western and Northern Europe and lower in Middle Africa and Eastern Asia. Rising incidence rates in Asia and sub-Sahara Africa suggest the importance of environmental factors such as diet, age at parity and reproduction pattern (2). In Norway, the incidence of breast cancer has almost doubled since 1965, and there were 3371 new cases of female breast cancer and 31 cases of male breast cancer in 2016 (3). The number of cases increased in the late 90's following the implementation of the mammographic screening program which inflated the incidence rate because cancers were being diagnosed one to three years earlier than they would have been in the absence of screening and possibly also due to detection of slow-growing tumors. Some of the historic increase in incidence rate also reflects changes in reproductive patterns involving known risk factors. Earlier diagnosis and improvements in treatment have seen the relative survival rate improve from approximately 70% in the 1970's to almost 90% today as illustrated in Figure 1.

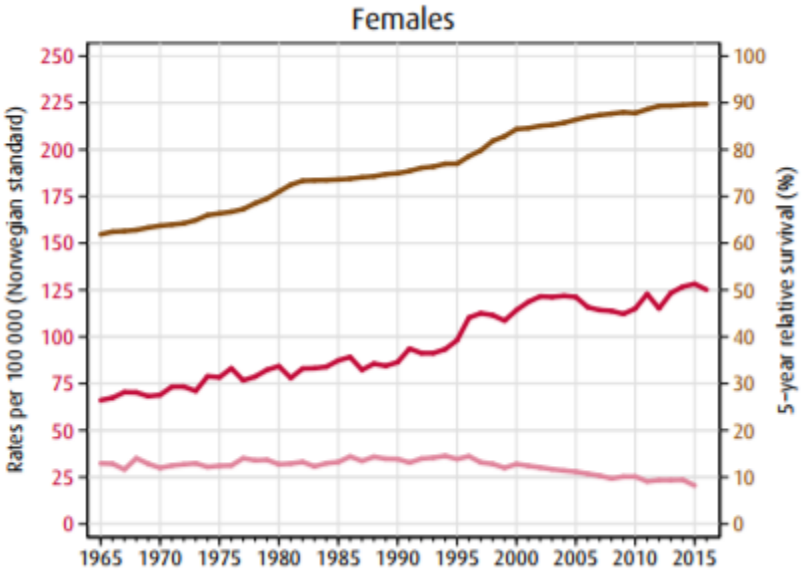


Figure 1: Incidence (red line), mortality rates (pink line), 5-year relative survival (brown) (3).

1.2 Risk factors line

There are several well known risk factors for developing breast cancer. Early menarche, late menopause, low parity, high age at first pregnancy, and short menstrual cycles are all factors which increases estrogen exposure to the mammary epithelium (4-7). Additional risk factors are high age, high alcohol consumption, night shift work (8, 9) and prolonged intake of estrogen supplements. Obesity is a risk factor in postmenopausal women (10, 11). The amount of mammographically dense tissue in the breast differs among women and because of differences in the relative amount of glandular and connective tissue. Women with dense tissue comprising >75% of the breast have a 4 to 5 times increased risk of breast cancer compared to women with less or no dense breast tissue (12). Inherited factors are the cause of 5-10% of breast cancer cases with mutations in high-penetrance genes *BRCA1* and *BRCA2* being the most prevalent. BRCA1/2 are involved in homologous DNA repair after double-strand breaks, transcriptional regulation, chromatin remodelling and cell cycle control (13). Carriers of pathogenic variants of *BRCA1* have a 55-65% risk of breast cancer and 39% risk of ovarian cancer by the age of 70. For carriers of pathogenic BRCA2 variants the risk of breast cancer is 45-47% and ovarian cancer is 11-17%. (Meta-analysis of BRCA1 and BRCA2 penetrance (14). Pathogenic BRCA1 variants are associated with early onset estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, human epidermal growth factor receptor 2 receptor (HER2)-negative, (so called triple-negative), breast cancer which is more aggressive in nature.

Other high-penetrance genetic factors include mutations in *TP53* (Li-Fraumeni syndrome) (15, 16), *PTEN* (Cowden syndrome) (17), *CDH1* (Hereditary diffuse gastric cancer syndrome) (18), and *STK11* (Peutz-Jeghers syndrome) (19, 20). Mutations in these genes comprise a small percentage of the total number of breast cancer cases, but carriers have a significantly increased lifetime risk of developing breast cancer, as well as certain other types of cancer (14).

1.3 Biology of the breast

The female breast, depicted in Figure 2A, contains, in addition to adipose and fibrous connective tissue, branching ducts which connect 15-20 lobes, each comprised of 20-40 terminal ductal lobular units (TDLU). The TDLU forms 10-100 sac-like acinii which produce the milk. The milk is emptied into ducts which converge in the nipple. The epithelium

throughout the TDLU is bilayered with an outer (basal) myoepithelial cell layer surrounding an inner (luminal) epithelial cell layer, as shown in Figure 2B.

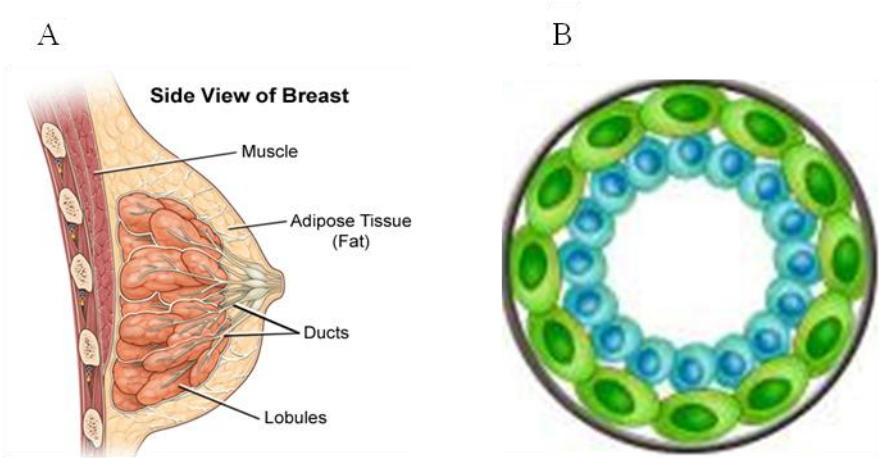


Figure 2: A: Anatomy of the female breast (21). B: Organization of cells in the TDLU. Blue cells depict luminal cells, green cells are myoepithelial cells resting on the basal lamina (22).

The cells in the luminal and basal layer are thought to arise from the same mammary stem cell residing in the basal compartment of the ducts. Cells in these layers have immunohistochemical profiles with distinct expression of cytokeratins and other specific markers (23).

The TDLU are the functional units of the breast. At puberty, the rising levels of estrogen and progesterone initiates growth and differentiation of the ductal structures. Full maturation of the TDLU occurs during pregnancy and lactation.

1.4 Tumorigenesis

Every multicellular organism relies on each cell to function as a coordinated unit with tight control over signaling, proliferation, apoptosis and function. When damage to the DNA allows a cell to bypass these control mechanism, the generation of a cell proliferating on the expense of its host can occur. Cancer is in the simplest of terms a genetic disease, but even though the underlying mechanisms of tumorigenesis are extremely diverse, a set of common traits have emerged. In 2000, Hanahan and Weinberg introduced six hallmarks of cancer and followed up with an additional four hallmarks 11 years later (Figure 3) The first six

established hallmarks of cancer cells included: 1.self-sufficiency in growth signals, 2. insensitivity to growth-inhibitory signals, 3. evasion of programmed cell death, 4. limitless replicative potential, 5. sustained angiogenisis and 6. ability to invade tissue and metastasize. All of these acquired capabilities which allow cancer cells to survive, proliferate and disseminate arise in different tumor types via distinct mechanisms at various times and serve as a framework in understanding the multistep process of tumorigenesis. The driving force behind these acquired capabilities is according to the authors, development of genomic instability which generates random mutations and chromosomal rearrangements, and the tumor promoting inflammatory micro-environment of the (pre-)malignant cells caused by the presence of immune cells. These two factors are termed enabling hallmarks. The Warburg effect describes the abnormal glycolysis in tumors where up-regulation of glucose transporters compensates for the switch to glycolysis even in the presence of oxygen (24). Reprogramming of cellular metabolism to support growth and proliferation seems so widespread in cancer that it's proposed as a hallmark of cancer. Due to unresolved issues about its functional independence from the original core hallmarks, it's deemed an emerging hallmark. A final emerging hallmark involves the cancer's ability to avoid recognition and destruction by the immune system.

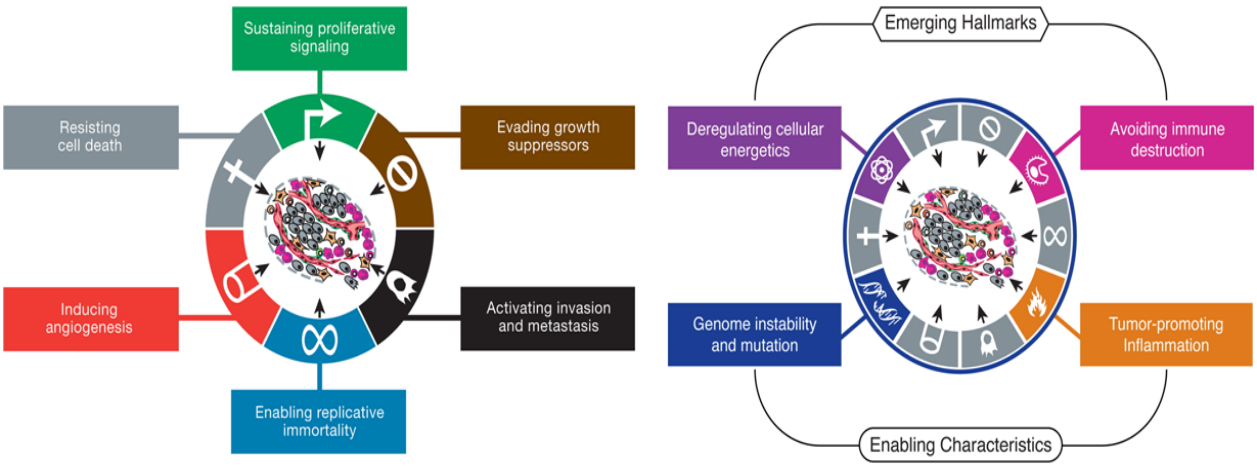


Figure 3: Schematic overview of the ten suggested hallmarks of cancer. The original six hallmarks are illustrated in the left image, the four later added hallmarks are depicted in the right image (25).

Most breast cancers originate from epithelial cells and are thus categorized as carcinomas. Ductal carcinoma not otherwise specified make up the vast majority, about 80%, lobular

carcinomas account for approximately 15% (26, 27). Regardless of subtype, the tumors are subcategorized into histological grades 1 to 3 based on level of differentiation (from well-differentiated, through moderately differentiated, to poorly differentiated, based on architectural features), cytologic features (pleomorphism), and mitotic activity. Ductal carcinoma *in situ* (DCIS) describes proliferation of malignant cells which do not yet have capacity for invasive growth, although they are considered a precursor to invasive carcinoma. DCIS of the breast can be divided into subgroups based on their growth pattern, morphology and level of necrosis.

1.5 Prognostic and predictive markers

When classifying a breast tumor for the purpose of prognosis and treatment, mapping the expression of hormone receptors estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (ErbB2/HER2) and cell proliferation marker Ki-67 by immunohistochemistry (IHC) is the recommended diagnostic practice for all breast cancer cases in Norway (28).

1.5.1 Estrogen receptor

ER regulates transcription of target genes involved with uterine and mammary gland development and growth as well as having other important functions in the immune, skeletal, cardiovascular and central nervous systems (29).

The steroid hormone estradiol (estrogen) crosses the cell membrane and binds to the estrogen receptor (ER). ER is a nuclear transcription factor and forms a dimer upon binding to a ligand. The activated receptor relocates to the nucleus and binds to estrogen response element regulatory sequences in the promoter of estrogen responsive genes and recruits co-regulatory proteins, leading to a physiological response. In addition to the classical “genomic” pathway that occurs over hours, there is an alternative “non-genomic” pathway in which estrogen can exert its effects within seconds or minutes through another class of estrogen receptors in or adjacent to the plasma membrane. These ERs form complexes with receptor tyrosine kinases activating MAPK/ERK and PI3K/AKT pathways. Estrogen can also bind non-ER plasma membrane-associated estrogen-binding protein leading to increased levels of Ca2+ or NO or activation of kinases.

ER expression is associated with more differentiated tumors with favorable prognosis and its detection by IHC as imaged in Figure 4 is part of routine diagnostics in breast cancer.

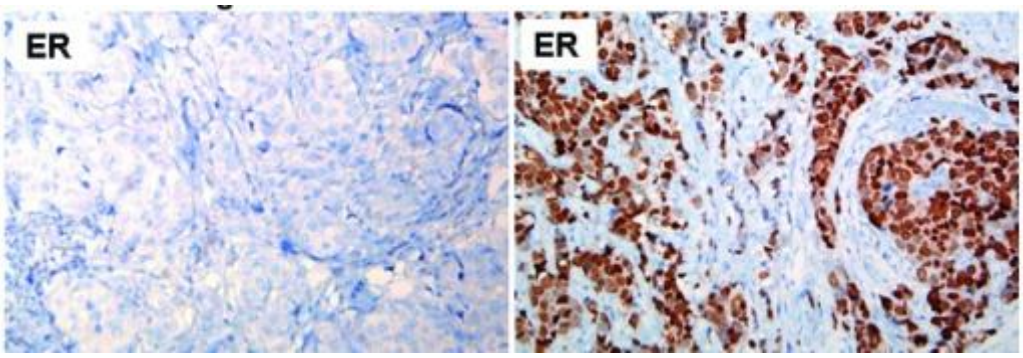


Figure 4: IHC of ER negative (left) and positive (right) breast cancer (30).

A hormone-receptor (ER and/or PR) positive tumor is likely to respond to hormone therapy and the administration of hormone receptor antagonists such as tamoxifen, or compounds blocking the production of estrogen (aromatase inhibitors), has been effective as adjuvant therapy for breast cancer.

1.5.2 Progesterone receptor

Progesterone is an ovarian steroid hormone essential for breast development during puberty and pregnancy in preparation for breast feeding. While progesterone acts as a mitogen in the breast, it has inhibitory function in the reproductive tract and ovary. Like ER, progesterone receptor (PR) is a ligand-activated nuclear transcription factor. There are two main isoforms, PR-A and PR-B, and a lesser known third isoform, PR-C. All are transcribed from the same gene. PR-B is the full length peptide while PR-A and PR-C are truncated due to alternate translational start sites within the mRNA. PR-B is required for normal mammary gland development, while PR-A is essential for uterine development and reproductive function. PR-A can also act as dominant repressor of PR-B and ER. PR-C can enhance PR activity in breast cancer cells or inhibit PR-B in the uterus being important in induction of labor. The ligand-activated A and B isoforms can form both homodimers and heterodimers capable of regulating gene expression by binding directly to DNA at progesterone response elements or indirectly by binding to other transcription factors. It can also exert non-genomic effects

through activation of protein kinases MAPK, PI3K/Akt or c-Src. PR is an ER target gene and a downstream effector of estrogen in the presence of EGF(31), suggesting crosstalk between EGFR, ERBB2 and the steroid receptors, but it can also be expressed independently of ER (32). It has been demonstrated that women taking progestins (synthetic progesterone) in combination with estrogen as part of hormone replacement therapy, had a greater risk of breast cancer with larger tumors of higher grade, than those taking estrogen alone (33). These results are controversial for several reasons (34-36), but still implicate PR in human breast cancer development and progression. The testing of PR status in breast cancer is important for at least two reasons. 3-5% of ER-negative patients are PR-positive and may respond to hormonal therapy. Also, PR status differentiates two classes of ER-positive breast tumors where ER+/PR- tumors are much less likely to respond to tamoxifen than ER+/PR+ tumors (37). Figure 5 is an image of negative and positive PR staining by IHC.

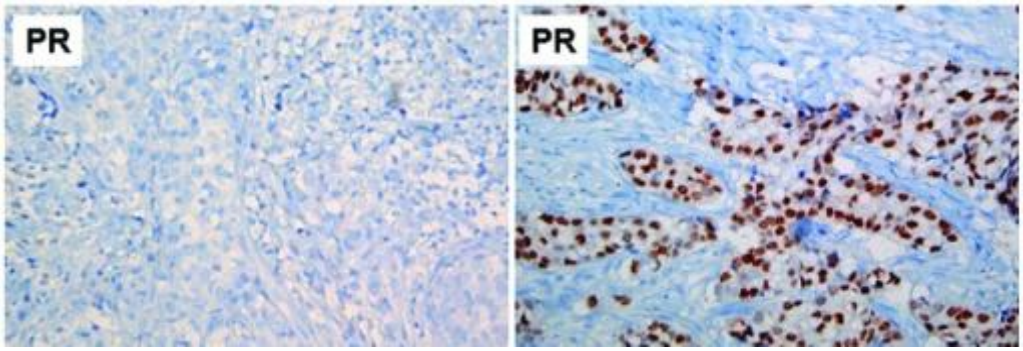


Figure 5: IHC of PR negative (left) and positive (right) breast cancer (30).

1.5.3. HER2/neu

The human epidermal growth factor receptor 2 (HER2), also known as ErbB2, is overexpressed in 15-20% of breast cancer tumors. Overexpression is associated with poorly differentiated, high-grade tumors with high rate of cell proliferation and metastatic capacity, higher recurrence rate, and shortened survival (38, 39).

HER2 is part of the epidermal growth factor family along with HER1 (EGFR), HER3 and HER4. Its intracellular domain has tyrosine kinase activity important in regulation of growth and differentiation (40, 41). HER2 has no known ligand, but a range of growth factors can bind to its family members. Ligand binding initiates receptor dimerization with HER2 being the preferred partner. Heterodimers have higher signaling potential than homodimers, and

heterodimers involving HER2 are more potent than those without. HER2/HER3 dimer is the most potent activator of proliferation and tumor development (42-44). Normal cells have low levels of HER2 at the cell membrane, thus limiting heterodimerization involving HER2 and keeping growth signals at a controllable level.

The humanized monoclonal antibody trastuzumab (45) targeting the extracellular domain of HER2 has been found to have significant impact on disease outcome when administered in combination with chemotherapy both in an adjuvant and neoadjuvant setting (46-52) by reducing cell proliferation and inducing apoptosis (53, 54). It has also proven effective in patients with metastatic disease (55). Trastuzumab is only effective against breast tumors with overexpression of HER2.

Testing for HER2 amplification and overexpression is standard of care for breast cancer patients and is assessed by IHC and/or *in situ* hybridization (ISH) analysis. IHC analysis for HER2 protein expression is visually assigned a score of 0, 1+, 2+ or 3+. Tumors lacking or with weak/incomplete membrane staining is given a score of 0 or 1+ and are deemed negative. Tumors with complete cell membrane staining but weak to moderate intensity are given a score of 2+ and are designated borderline. Borderline cases should be re-tested by complementary method such as ISH. Signals from probes specific for *HER2* and centromere 17 (cent17) is counted manually and a ratio of HER2:cent17 higher than 2.0 is considered amplified. A score of 3+ is given when strong and complete membrane staining is observed in >10% of tumor cells. Figure 6 A and B depicts HER2 detection by IHC and FISH respectively.

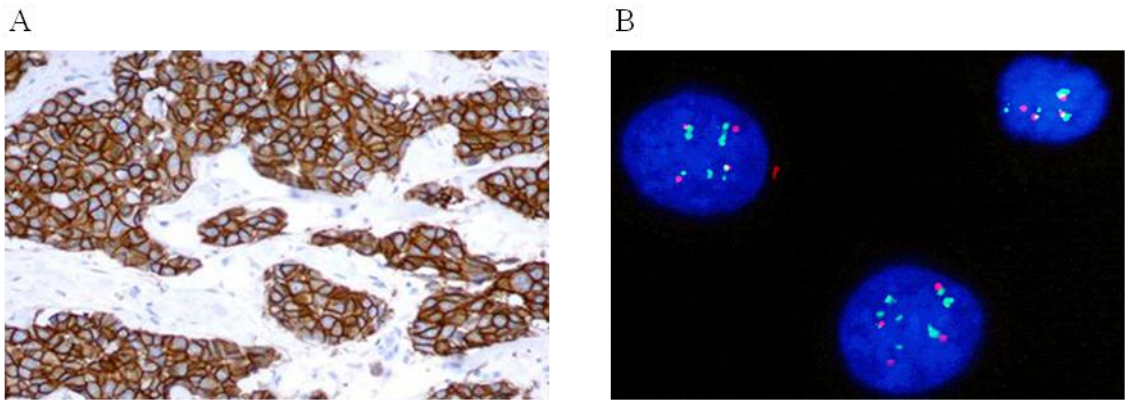


Figure 6: A: IHC of breast cancer with HER2 score of 3+ (56). B: HER2 (green) amplification detected by FISH with chromosome 17 centromeric control probe (red) (57).

A significant number of HER2+ breast tumors may either not initially respond to trastuzumab or eventually develop drug resistance. Possible reasons for drug resistance may be obstacles preventing binding to antigen (truncated, but active HER2 (58), up-regulation of downstream signaling pathways (59, 60), use of alternate signaling pathways (61, 62), or failure to trigger immune-mediated cell death (63, 64).

Several other drugs targeting HER2 and its association with its family members are available in therapy such as lapatinib and neratinib which are dual tyrosine kinase inhibitors interrupting EGFR/HER2. While trastuzumab is inefficient in blocking HER2/HER3 dimerization, the monoclonal antibody pertuzumab binds HER2 in a way that prevents its dimerization with EGFR, HER3 and HER4 (65).

1.5.4 Ki-67

Ki-67 is a nuclear protein associated with cellular proliferation. It's expressed in the G1, S, G2 and peaks in the M phase of the cellular cycle, but it is non-existing in resting G0 phase (66). Normal breast tissue has low levels (<3%) of Ki-67 expression (67), and high levels of Ki-67 in breast tumors are associated with worse prognosis (68). Ki-67 is most readily analyzed with IHC and expression is estimated by manually counting cells. It's recommended as a biomarker to determine optimal treatment in early breast cancer (69) and is included in the Norwegian breast cancer group (NBCG) guidelines (28). Figure 7 is an image of Ki-67 staining by IHC in breast cancer.

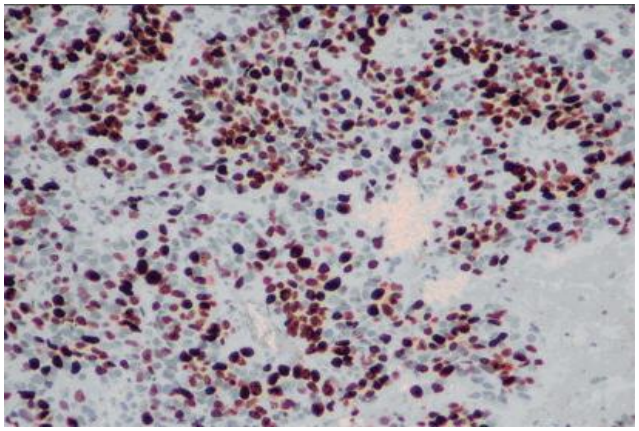


Figure 7: IHC of Ki-67 staining of breast cancer (70).

1.6 Diagnosis and treatment

In Norway, the Norwegian Breast Cancer Group (NBCG) together with the Norwegian Directorate for Health maintain the guidelines for diagnosis and treatment of breast cancer patients (28). In general, a needle biopsy is performed after a tumor in the breast is verified by ultrasound or radiographic imaging. The tumor is classified according to stage, morphology, histological grade, molecular markers such as ER/PR/HER2, size of the tumor and surgical margins by pathological examination. The primary treatment is surgery of the breast with/without removal of lymph nodes in the axilla.

Tumor stage is determined by the TNM system (T=tumor, N=nodes, M=metastasis). Tumor size is graded T1-T4, where T1 denotes tumor less than 2 cm in diameter and T4 marks tumor invading either skin or chest wall. Lymph node involvement is graded N0-N3, and M0-1 describes the presence of distant metastasis. Patients with clinical tumor stage T1-2 N0-1 M0 is considered primary operable. Patients with T3-4 N0-3 M0-1 or T1-2 N2-3 M0-1 describing tumors with more extensive lymph node involvement or distant metastasis are considered primary inoperable. The primary tumor is removed by lumpectomy (breast conserving surgery, BCS) or mastectomy (removal of entire breast). In patients where axillary lymph node involvement is unknown, a sentinel lymph node dissection and assessment of metastatic invasion is performed. The sentinel lymph node is the first lymph node in relation to the tumor that drains the breast and is usually the first location of lymphatic dissemination from the primary site (71). Detection of disseminated cancer cells indicates full axillary clearance (72) Patients with locally advanced, primary inoperable breast cancer undergo neoadjuvant therapy to shrink the tumor and prevent progression prior to surgical removal.

Surgical removal of the primary tumor is the cornerstone of breast cancer treatment, but other options include chemotherapy, radiation therapy, endocrine or targeted therapy.

Systemic adjuvant treatment, chemotherapy, can be administered after surgery based on the patient's age, TNM status and cancer subtype. Standard treatment regimen consists of either anthracyclins or taxanes. Anthracyclins can exert their effects through different mechanisms; inhibition of DNA and RNA synthesis by intercalation, blocking of transcription by inhibition of topoisomerase II, the enzyme which relaxes supercoiled DNA, generation of free radicals that damage DNA, proteins and cell membranes (73), and histone eviction from open chromatin areas (74). Taxanes disrupt microtubule function, thus inhibiting the cell from progressing through the cell cycle (75).

Radiation is therapy directed at reducing relapse after surgery and is offered to patients who have had breast conserving surgery or have cancer cell positive lymph nodes. The radiation therapy aims to induce apoptosis due to catastrophic DNA damage in the remaining cancer cells (76).

Patients with tumors positive for ER and/or PR are offered anti-estrogen therapy. Selective estrogen receptor modulators (SERM) act as an ER antagonist in the breast. The most widely used compound, tamoxifen, binds to ER causing a conformational change that prevents estrogen binding and its ensuing effects in the cell. Several studies show that use of aromatase inhibitors (AI) increases disease-free survival compared to tamoxifen (77, 78). AI inhibits aromatase which is responsible for converting androgen to estrogen. The administration of AI is contraindicated in premenopausal women whose estrogen synthesis occurs in the ovaries. A decrease in estrogen triggers an increased production of androgen. In postmenopausal women, the synthesis of estrogen occurs in other body compartments including the liver, muscles, connective tissue, and skin (79). The use of tamoxifen therapy for 5 years reduces relative mortality by 30% (80), and when extending therapy to 10 years the relative mortality drops another 30% compared to women taking it for only 5 years (81). Targeted therapy aimed at HER2 has proven effective in HER2 amplified tumors, but for triple negative tumors no targeted treatment currently exists and general systemic chemotherapy is the only option. For women with metastatic disease, all treatment options are considered palliative.

1.7 Classification by alterations in gene expression

Breast cancer is a disease driven by multiple genomic alterations in DNA and extensive inter- and intra-tumor heterogeneity (82). By understanding the drivers influencing proliferation, metastatic ability and resistance to systemic treatment in each patient, improved care with prolonged survival may be provided.

By analyzing gene expression profiles in breast carcinoma and selecting genes with a similar expression pattern before and after chemotherapy, an intrinsic gene list presumably reflecting the phenotype of individual tumors was identified (83). Hierarchical clustering of these genes led to classification of the breast tumors into five distinct subclusters, named the intrinsic clusters. The tumors of luminal A type are frequently ER+ and/or PR+, HER2 negative and have low expression of Ki-67 and are associated with good prognosis. Luminal B tumors

have higher expression of genes involved in proliferation and may also express HER2, making the prognosis significantly worse compared to luminal A. The HER2-enriched subgroup has frequently an activation of HER2 related pathways. One subgroup was dominated by ER-, PR- and HER2- (so-called triple negative) tumors with expression of genes characteristic of myoepithelial or basal epithelial cells, such tumors are called basal-like. There is considerable diversity within each group and especially among the basal-like tumors which can be further subdivided into several distinct subgroups (84). Both latter subtypes have rather poor prognosis.

A fifth subgroup of breast carcinomas is called normal-like based (intermediate prognosis) on expression patterns similar to normal breast tissue. Recently, yet another subtype called claudin-low has been identified (85). These tumors are often triple-negative, and show genomic instability with many DNA copy number (CN) gains and losses. The majority of these tumors are associated with poor prognosis.

The Food and Drug Administration approved a refined version of the intrinsic subtypes, the PAM50 test, which was optimized for quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) (86) to analyze genes characteristic of the intrinsic subtypes and assigns samples into luminal A, luminal B, HER2-enriched or basal-like subtypes. This is now commercialized as the Prosigna test (Breast Cancer Prognostic Gene Signature Assay), which in addition to the PAM50 subtypes, gives a numeric score that predicts the risk of relapse after 10 years in postmenopausal women with ER+ tumors (the risk of recurrence (ROR) score).

1.8 Classification by DNA copy number alterations

Classification of breast carcinomas based on pattern of DNA copy number alterations has identified recurring groups termed simplex, complex I (sawtooth), complex II (firestorm), and flat (87). Tumors of the simplex type are characterized by broad segments of duplications and deletions, usually involving whole arms or entire chromosomes. This DNA alteration pattern corresponds often to the luminal A intrinsic subtype. The complex I group with its sawtooth pattern of narrow segments of deletions and duplications often affecting the majority of the chromosomes corresponds to the basal-like subtype. Complex II tumors, which are usually identified as HER2-enriched or luminal B, resemble the simplex type, but have at least one narrow peak of high-level amplification with intermittent deletions (so-called firestorms).

The complex I group with its sawtooth pattern of narrow segments of deletions and duplications often affecting the majority of the chromosomes corresponds to the basal-like subtype. The flat subtype consists of tumors with no obvious amplifications or deletions. The clinical relevance of this subtype is unclear.

1.9 Classification by combining expression and copy number alterations

By analyzing both gene expression and DNA copy number alterations, Curtis et al. selected genes whose expression were correlated to copy number alterations (*in cis*). By using this gene list, 2000 breast tumors were divided into 10 subgroups, termed integrative clusters (IC) (88). The 10 IC have since been further validated (89) and represent one of the most extensive molecular-based taxonomy to date. The integrative cluster approach takes into account major chromosome alterations like specific whole arm gains or losses, narrow amplifications and complex rearrangements (such as firestorms). When classification by this method is compared to classification by PAM50, most of the latter groups get split into subgroups identifying subsets of the five subtypes with differences in prognosis. For instance, the IC 2 comprises tumors identified as intrinsic subtype luminal A or luminal B which in addition to whole arm gains and losses, also have high level amplification of tumor driver genes on chromosome 11q, a typical firestorm event. This subgroup of ER+ patients has poor prognosis. Other ICs dominated by ER+ tumors are IC group 3, 4, 7 and 8 correlating with luminal A tumors with simplex pattern, and IC 1 and IC 6 which mainly comprise luminal B tumors with simplex or firestorm DNA architecture. ER-tumors with DNA alterations of the saw tooth type and chromosome alterations characteristic of basal-like tumors mostly make up IC 10. A subset of basal-like tumors with intermediate prognosis falls into the mixed IC 4. HER2-enriched tumors are mainly grouped into IC 5 which also comprises both ER+ and ER- tumors with poor prognosis. The HER2-enriched tumors in IC 5 have narrow high-level amplification of HER2 and some neighboring genes, while the HER2 when enriched in samples grouped into IC 1, 6 and 9, is part of a broader lower-level amplification with unknown prediction power on HER2-targeted therapy response. IC 9 is mixed with ER+ and ER- tumors with intermediate prognosis. This group has frequent amplification of oncogene MYC and loss of tumor suppressor gene TP53 and RB.

1.10 DNA damage and mechanisms of copy number aberrations

The DNA in a cell is subject to massive damage on a daily basis. Ultraviolet radiation can lead to pyrimidine dimers prone to deamination which, if left unrepaired, will result in cytosine being replaced by thymine. Ionizing radiation can lead to both single strand (SSB) and double strand breaks (DSB) of the DNA backbone. A range of environmental factors such as tobacco smoke, industrial chemicals, some plant and microbial products, and some chemotherapeutic drugs also possess mutagenic effects on DNA. Additionally, endogenous processes in the cell such as spontaneous depurination or deamination, attack by reactive oxygene species generated as a by-product in metabolism, non-enzymatic methylation producing the cytotoxic base 3-methyl-cytosine, as well as proofreading errors in replication, pose a threat to the integrity of the genome. Upon detection of DNA damage, the cell will stop progression through the cell cycle and either repair the damage or enter apoptosis. Malfunctions in the systems for detecting and repairing DNA damage are important contributors in the development of cancer.

Pan-cancer sequencing has revealed an inverse relationship between the amount of mutations and copy number aberrations (CNA). This so-called “cancer genome hyperbola” illustrated in Figure 8 depicts cancers seeming to be dominated by somatic mutations or CNAs (90). Most breast cancer tumors are dominated by CNAs.

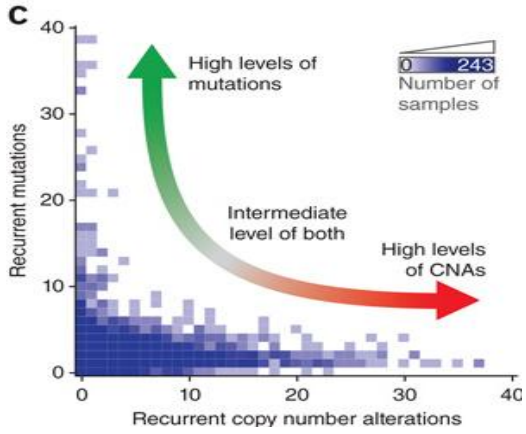


Figure 8: Illustration of the cancer genome hyperbola (90).

Structural variation in the DNA can manifest as deletions or amplifications of DNA segments, inversions, tandem duplications, or more complex alterations such as deletion or rearrangement of chromosomal arms or loss/gain of entire chromosomes. These events may promote tumor progression by amplification of oncogenes or deletion of tumor suppressor

genes. Translocations may create fusion genes or juxtapose a gene next to the regulatory sequence of another gene and in the process result in altered gene expression.

There are multiple endogenous processes which can result in such rearrangements which eventually can drive tumor development. Mechanisms to repair double strand breaks and erroneous replication are known to contribute to structural variation and can be divided into three groups:

1. Homologous recombination
2. Non-replicative, non-homologous repair
3. Replication-based repair mechanisms

The DNA damage response (DDR) which is the system of proteins and pathways to repair DNA is complex and relies on at least six different mechanisms. Three of the mechanisms deal with incorrect bases in one strand using the second strand as template. In base excision repair (BER) the damaged base is removed by a DNA glycosylase followed by removal of the associated sugar-phosphate residue by an endonuclease and a phosphodiesterase, re-synthesis of the base by a DNA polymerase and sealing of the nick (a break in the phosphodiester bond between two adjacent nucleotides) by DNA ligase. More extensive DNA damage such as pyrimidine dimers and SSB are corrected by nucleotide excision repair (NER). This involves excision of a larger stretch of DNA surrounding the damage, re-synthesis and sealing of the nick. An infrequently used mechanism can directly remove methyl groups from incorrectly methylated guanines.

DNA damage involving both strands is repaired either by homologous recombination (HR) or non-replicative, non-homologous repair. HR occurs during late S or G2 of the cell cycle and uses the homologous chromosome as a template in reconstitution of the original sequence (91). The invasion of a homologous strand to act as template may in some cases cause deletion and amplification, and thus copy number variation, in a process called non-allelic homologous recombination. This happens in the event of misalignment between the chromosomes due to non-allelic sequence similarity and may even occur between non-homologous chromosomes (92, 93). There are yet other homology-based pathways known to

cause mutations or rearrangements (94). HR can also contribute to carcinogenesis through loss of heterozygosity (LOH) (95).

The best characterized non-replicative, non-homologous repair mechanism is called non-homologous end-joining (NHEJ). NHEJ functions throughout the cell cycle by bringing together and fusing the broken ends of chromosomes regardless of sequence and generally with loss of a few nucleotides due to trimming of bases at single-stranded ends (96) . In the occurrence of multiple simultaneous DSBs, NHEJ may fuse ends of different chromosomes causing translocations, insertions/deletions or inversions.

Replication based repair mechanisms deal with both mismatch base repair and other replication errors. In the event of a replication fork collapse due to a single-stranded nick, a single-ended double-strand break is formed. This is repaired by break-induced replication (BIR), a process resembling HR. Under normal circumstances this repair is error free, but occasional mistakes can result in LOH, translocation and deletion/duplication and can be an underlying mechanism for CNA and disease development (97).

1.11 Breakage-fusion bridge cycles

The ends of chromosomes are protected from degradation and fusion of chromosome ends by the DNA repair system through multiple repetitive sequences called telomeres (98, 99). In stem cells and germ cells telomerase elongates the telomeres after each round of replication, thus retaining telomere length. In most tissues the telomere is shortened after each successive round of replication (100), thus limiting its proliferative potential. A signal of DSB is activated when chromosome ends lacking telomeres are exposed. This normally triggers TP53-mediated apoptosis. Mice deficient in both telomerase and TP53 show an increase in carcinoma with chromosome rearrangements (101). Upon entering mitosis, uncapped chromosomes will produce two sister chromatids lacking a telomere. The sister chromatids can then fuse with one another. As shown in Figure 9, during anaphase the chromatids will be pulled in opposite direction and break apart, not necessarily at the site of fusion. This produces two uneven chromatids lacking telomeres and the cycle can repeat through subsequent cell divisions until the chromosomes acquire a telomere by translocation with a non-homologous chromosome. This can initiate new breakage-fusion-bridge (BFB) cycles involving the new chromosome lacking a telomere, thus inducing massive chromosomal instability. Cells which have undergone BFB cycles will typically have some DNA regions

with complex rearrangements including high-level amplifications with intermittent deletions. In breast cancer, this is the copy number alteration pattern recognized as “firestorms” by Hicks et al.

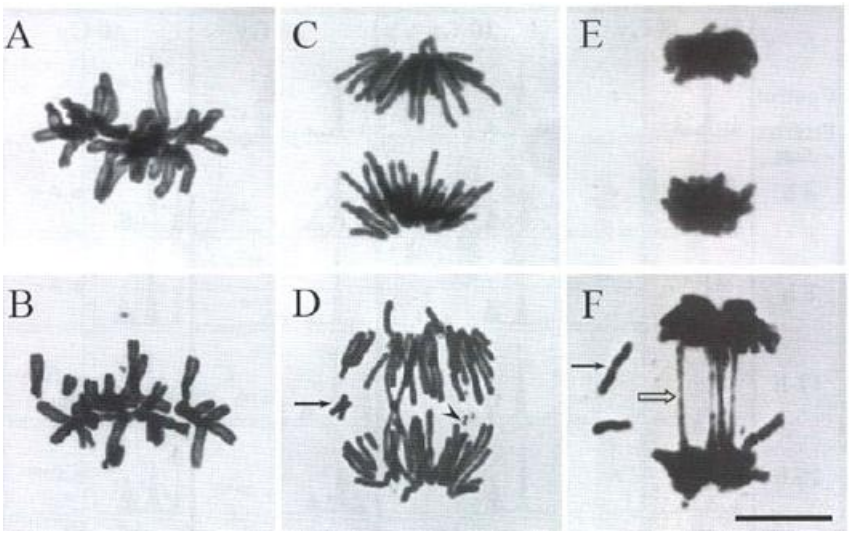


Figure 9: metaphase (A,B), anaphase (C,D) and telophase (E,F) of normal (top row) and irradiated cells (bottom row) (102). Black arrows: acentric chromosomes, arrowhead: double minute chromosome, white arrow: chromosome bridges

1.12 High-level amplifications in breast carcinomas

As mentioned previously, breast cancer is regarded as a neoplasia where DNA copy-number alterations are an important driver event (90). Tumors can have defects in the different repair mechanisms which will result in different types of genomic abberations. Some recurrent alterations are low level gains of large chromosome segments or extra copies of chromosome arms or whole chromosomes. Others are high-level copy number gains where in extreme cases some regions/genes are amplified 10-fold or even more. Such regions are of particular interest as:

1. They are seen even in early stage neoplasia (such as DCIS).
2. They are hallmarks of several biologically and clinically distinct subtypes.
3. Most are highly correlated with increased mRNA and protein level of the amplified genes and represent both established but also potential therapeutic targets.

Today, copy number estimation of only one gene (*HER2*) is included in guidelines for breast cancer diagnostics. If more genes are to be included in daily diagnostics for more optimal prognostication and therapy prediction, it will be both time- and tissue consuming to perform FISH analyses. An alternative would be to apply next-generation sequencing technology for assessment of copy number alterations in multiple regions. NGS is a costly and tissue demanding procedure, but by using a targeted approach where only selected regions of the genome are sequenced, NGS is suitable for diagnostic routine. So far panels for targeted sequencing have been implemented mainly for detection of single nucleotide mutations and small indels. A panel aiming at detecting copy number alterations in breast cancer samples could prove advantageous but will need thorough validation.

2. Aim of study

This work aims at comparing DNA copy number levels assessed by FISH and by next generation sequencing.

This was addressed by:

1. Designing and validating probes detecting frequently amplified DNA regions in breast cancer.
2. Studying the relationship between FISH on intact nuclei (imprints) and sections of nuclei (diagnostic tissue sections).
3. Comparing DNA copy number estimated by FISH with DNA copy number estimated by targeted sequencing.

3. Materials and methods

3.1 Materials

3.1.1 Cell lines

The normal fibroblast cell line was selected as normal control for FISH probes. The cell lines HCC2218 (polyploid) and HCC1954 were selected based on the HER2 positivity and that lymphoblast cell lines from the same patients were available. Lymphoblasts provide internal control. All cell lines were cultured in PenStrep (Gibco, Sigma Aldrich). The imprints were made by adding cells from a culture to glass slides and allowing the cells to settle before fixation by immersion in methanol. The imprints were subsequently stored in 70% ethanol at -20°C. A volume of each cultured cell line was fixated and paraffin embedded by the Department of Pathology. Data for selected cell lines are listed in Table 1.

Table 1: List of cell lines with assorted technical data.

Cell line	Origin	Cell type	Her2 ampl.	Other characteristics	Medium	Manufacturer
Fibroblasts	Blood	Fibroblast	Normal	Normal	MEM w/salts[1] + 10%FBS[2] +L-glutamine[3]	Coriell Inst.
HCC2218	Breast/duct	Invasive ductal carcinoma	Overexpressed	p53 positive ER negative polyploid	RPMI-1640 + 10% FBS	ATCC
Lymphoblast from HCC2218	Blood	Lymphoblasts	Normal		RPMI-1640 + 10% FBS	ATCC
HCC1954	Breast/duct	Invasive ductal carcinoma	Overexpressed	ER negative PgR negative	RPMI-1640 + 10% FBS	ATCC
Lymphoblast from HCC1954	Blood	Lymphoblasts	Normal		RPMI-1640 + 10% FBS	ATCC

3.1.2 Patient samples

Formalin-fixed, paraffin-embedded (FFPE) tissue from primary breast carcinomas from 14 patients were selected based on existing data from genome-wide copy number analyses (tumors displaying a localized amplification). The patients were part of a prospective breast

cancer study (the Oslo2 study, REK 200606181-1). One patient sample was omitted due to sectioning difficulties; the sample was fat rich and dissolved on the water bath microtome and thus required dry sectioning which was not available in the lab. Additional data about patient samples selected for FISH are included in Table 2.

Table 2: Tumor samples selected for FISH with information about ER, PR, HER2 and Ki67 status, histological grade, H&E, tumor type, and ID number.

Oslo2U_#	To FISH	Block	Hist. Grade	Size	ER	PgR	HER2	Ki67		H&E comment:
1116	3	3	3 (3 3 3)	25	0	0	0	80		Ok
1117	4	4	2	10	>50	0	0	31		Ok
1119	4	2, 3, 4	2 (2 3 2)	25	100%	10%	2+ (neg CISH, 1.42)	35		Ok
1122	5 and 7	4, 5, 7	2 (3 2 1)	10 og 9	100%	70 and 80	0	10 og 25	Two tumors	5 and 7: DCIS
1124	6	6, 7, 8	3	22	>50%	0	3	47		Ok
1129	2	1, 2, 3	3 (3 3 3)	18	0	0	3	>60	Medullary	Ok
1138	4	3, 4, 5	2	13	100	0	0	20	Lobular	Ok
1162	2	2	2	25	>50	>10	0	34		Ok
1171	7	7	2	30	100	100	0			Ok
1175	3	3 og 7	2	40	0	0	0	60		Not tumor
1203		9	2 (3 2 1)	18 and 15	>50 100	0	0	25 and 20	Two tumors	Sample omitted
1210	3	3	2	15	>50	>10	0	34		Ok
1240	3	3	3	40	>50	>10	3	33		Ok

From the same clinical cohort 100 patient samples were selected for a separate classification study where both RNA sequencing and PAM50 where performed using fresh frozen biopsies and FFPE tissue. From these, DNA from 68 samples was analyzed by targeted sequencing in a separate project aiming at identifying high-level DNA amplifications and nine of these were also analyzed by FISH.

3.2 Methods

3.2.1 Fluorescence and microscopy

Assessing the basic features of a tissue specimen stained with haematoxylin and eosin requires the use of an optical microscope, or light microscope. The optical microscope uses visible light and a system of lenses to magnify a specimen placed on the stage between the

light source and the lenses which consists of the ocular lens (eye piece) and the objective lens. Light passes through a condenser focusing the light on the specimen.

Fluorescence describes the process where light of a specific wavelength is absorbed by an atom thereby raising an electron from its ground state to an excited state followed by its immediate return to its ground state accompanied by the emission of light of a lower wavelength. The shift to a lower wavelength is called the Stokes shift and is caused by small amounts of energy being lost through vibration of the excited electron. The wavelength, and thus energy, of the incoming light must be equal to the difference between the electron's possible energy states for fluorescence to occur.

Application of a fluorescence microscope for visualization of fluorochromes and fluorophores has become an indispensible technique in molecular biology. Different fluorophores absorb and emit light at different wavelengths allowing multiplexing with several fluorophores to visualize a number of features simultaneously by a simple switch of filter. The epifluorescence microscope uses a high powered light source such as a mercury or xenon arc lamp to illuminate the specimen from above. The light passes through an excitation filter suitable to the fluorophore to be visualized and is focused through an objective lens before exciting the specimen. The emitted light from the excited fluorophores passes back through the same objective lens. A dichroic mirror and an emission filter separate the emitted light from the reflected excitation light transmitting only the fluoresced light to the eye piece and detector. Figure 10 depicts an epi-fluorescence microscope.

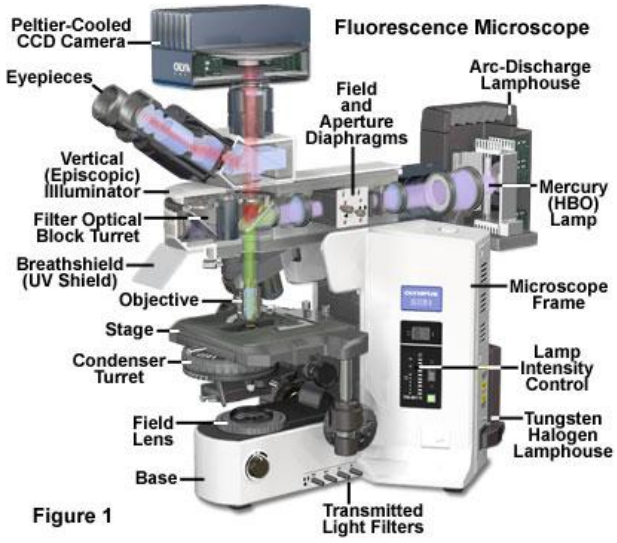


Figure 10: Anatomy of an epi-fluorescence microscope (103).

The epi-fluorescence microscope used for our analyses was a Zeiss Axioplan 2 microscope equipped with an Axio cam MRM CCD camera and Axio Vision software. The selected areas were photographed in z-stacks with 20 images vertically photographed through the specimen.

3.2.2 Fluorescence In situ hybridization

In situ hybridization was first described as a method for detecting specific regions of DNA in 1969 using radioactively labelled probes (104) and has today evolved into a vast array of different applications following better understanding of the physical properties of nucleic acids, vast improvements in fluorescence microscopes, and mapping of the genome coupled with bioinformatics. The various techniques have become indispensible tools for detection of amplifications, deletions, and translocations in nucleic acids. Most of the techniques now use fluorescence for detection. Standard fluorescence *in situ* hybridization (FISH) exploits the basic property of DNA to hybridize with a DNA probe under stringent conditions. The probe can be labelled with fluorescently tagged nucleotides for direct visualization. The nucleus is stained blue with the DNA binding dye DAPI (4',6-diamidino-2-phenylindole).

FISH can be performed using commercially available probes, whole kits, in-house probes made from bacterial artificial chromosomes (BACs), or by PCR based techniques. In-house probes are inexpensive and offer the possibility to easily switch fluorophore for optimal multiplexing.

3.2.3 Selection of regions for customizing DNA FISH probes

As a separate project a customized next-generation DNA targeted sequencing panel for copy number estimation of 19 different DNA loci frequently found to be amplified in breast cancer was developed. Briefly, a TruSeq Custom Amplicon Kit v1.5 (Illumina) for targeted sequencing was developed with the Design Studio software (Illumina). The number of oligo pairs was 1536 (maximum number) and the length of the amplicons was median 155 nucleotides (range: 125-190). The principle behind this approach, as illustrated in Figure 11, was oligo pairs binding to the same DNA strand in a multiplex reaction over a wide range of annealing temperatures (as the temperature gradually were taken down to 40°C after denaturation of DNA). Bound 5' oligos were extended and ligated to the 3'oligos. Universal sequences in the custom oligos were used to amplify and barcode the extension products. Barcoded products were pooled and sequenced using a MiSeq machine (Illumina).

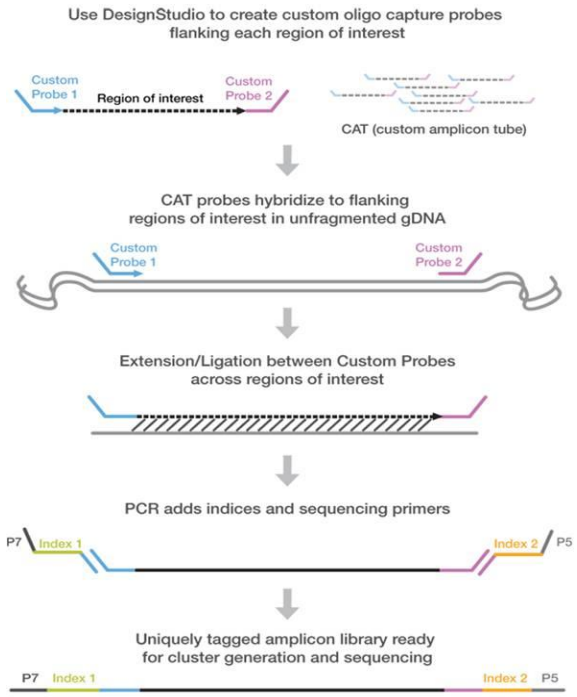


Figure 11: Principle of Illumina TruSeq Custom Amplicon Kit (105).

Briefly, the data analysis was performed as follows:

Sequencing data was obtained from the Illumina MiSeq sequencing machine, and preprocessed through the integrated software (MiSeq Reporter). The output data from this first process contained information on the number of DNA reads from each base pair in each of the 1536 regions included in the panel (regions typically span ~100 base pairs). The first step in the algorithm was to aggregate the data within each regions, so that each region was represented with one mean value for number of reads, as well as one mean B-allele-frequency (BAF) value (pertaining to the regions including SNP-variants). Read values were then log-transformed. For this project, read values from sequencing data from 19 normal (non-abberant) samples were clustered based on a hierarchical clustering method. From this analysis 145 regions systematically presented with low reads, and were interpreted as being of technically low quality and therefore excluded from further analyses. The next step involved segmentation of the log-transformed read counts, to identify breakpoints, and to reassign read-values based on mean read count within each segment. This was performed using the PCF method (piecewise constant fitting) (106). The next step was to calculate integer copy number value for each segment. A central concept in order to find a probable copy number profile was to model a series of non-integer copy number profiles, based on

many different combinations of tumor fraction and mean tumor ploidy. In order to reduce the number of combinations of tumor fraction and mean tumor ploidy, we added one more assumption: that the most frequent read count value for regions with a BAF-value close to 0,5, represented a clonal and balanced genomic region in the tumor with 1+1 or 2+2 maternal/paternal allele distribution. The sum square error between modeled non-integer copy number profiles and the closest integer-copy number profile was then calculated. The model presenting the lowest sum square error value was chosen.

Validation and analysis of BAC probes

Bacterial artificial chromosome (BAC) clones with DNA sequence from several of the regions (see Table 3) was selected using the University of California Santa Cruz genome browser (http://genome.ucsc.edu/). The BAC clones were inserted in plasmids in the chloramphenicol resistant *Escherichia coli* (*E.coli*) and the bacteria were purchased from Source Bioscience. Table 3 lists all probes detailing location and target genes. Probe p9 (CTC-859I3) was discontinued prior to FISH on patient samples. Replacement probe (RP11-939C24) is given the same probe number (p9). Probes for sequences on chromosome 8 target different regions and was hence not used as mutual validation. Probes given a probe number were selected for further experiments.

Table 3: List of probes with location and target genes from hg19 (Human Genome version 19).

Probe	Chr	Base pair start	Base pair end	Target genes	Probe number	Status
CTD-2117M19	1	67227507	67304267	*INSL5/WDR78*		Some satellites
RP11-205N11	1	67318175	67483062	*MIER1*	p1	OK
CTD-3092L3	1	150600571	150658482	*GOLPH3L*		Unspecific
RP11-623P23	1	150447829	150628209	*GOLPH3L*	p2	OK
RP11-148K15	1	204571658	204744566	*MDM4*	p3	OK, some satellites
RP11-155F3	1	204728025	204909902	*NFASC*		Abundant satellites
RP11-466H15	3	178822841	179006790	*PIK3CA*	p4	OK, weak signal
RP11-737O18	3	178633717	178836408	*ZMAT3*	p5	OK
RP11-1141F12	8	38163568	38352086	*WHSC1L1*	p6	OK
RP11-440N18	8	128596756	128777986	*MYC*	p7	OK, some satellites
RP11-1147E20	9	14132115	14263575	*NFIB*	p8	OK, some satellites
RP11-355C15	9	13975754	14148081	*NFIB*		OK, some satellites
CTC-859I3	10	123212761	123337532	*FGFR2*	p9	OK, weak signal
RP11-939C24	10	123312805	123494626	*FGFR2*	p9	OK, some satellites
RP11-825J6	11	68958598	69145055	*FGF4/CCND1 a*	p10	OK
RP11-156B3	11	69106702	69297637	*FGF4/CCND1 b*		OK
RP11-614E9	11	69445317	69614776	*CCND1*	p11	OK
RP11-878N15	12	69235643	69417613	*3'MDM2*	p12	OK, weak signal
RP11-450G15	12	69080492	69256798	*MDM2*		Occasional satellites
RP11-1087J7	15	99251672	99432616	*IGFR1*	p13	OK, weak signal
RP11-1116O4	15	99419812	99595033	*IGF1R*		Satellites
RP11-62N23	17	37725641	37882864	*NEUROD2-GRB7*	p14	OK
RP11-94L15	17	37811836	37973561	*STARD3*		Unspecific
RP11-610O22	17	37862118	38011536	*ERBB2-IKZF3*		Unspecific
CTD-3135I22	17	37953489	37995042	*IKZF3*	p15	OK
RP11-1065N2	17	57646788	57864577	*VMP1*		OK, weak signal
RP11-76K15	17	57855372	58019732	*VMP1*	p16	OK, backgound
CTD-3174P24	20	54892712	55069396	*RTFDC1*	p17	OK
RP11-875M12	20	55066527	55282509	*FAM209B*		Unspecific

The analysis of signals in the nuclei followed a set of parameters based in part by a book about FISH (107), advice from supervisors, and limits set by time, available resources and experience. The following parameters were used as guidelines in analysis of probe signals throughout this thesis:

1. Nuclei from multiple areas with uniformity in signal intensity and signal-to-noise ratio were selected for analysis.
2. Nuclei should not overlap.
3. Apparent signals revealed as artifacts and the associated nuclei were omitted from analysis. Common artifacts include proteinaceous debris and crystals.
4. Nuclei with signals located on the extreme periphery of the nucleus were omitted.
5. Cells with duplicated DNA through S phase display two signals in close proximity reflecting the sister chromatids. These cells were omitted or the signals counted as one.
6. A goal of 50 counted cells per analysis was set.
7. In nuclei where signal count was between 10 and 20, an accurate quantification was sometimes difficult due to close proximity of signals and minor miscalculations may have occurred. Separating these signals as one or more was unbiased causing some over- and underscoring.
8. The number of signals in any given nuclei was capped at 20. Some nuclei displayed a massive amount of signals clearly exceeding 20, but the close proximity of these signals made an accurate quantification impossible.

3.2.4 Plasmid isolation

Plasmid isolation was performed using the low-copy plasmid QIAGEN-tip 500 (Hilden, Germany) kit with very low-copy plasmid protocol. Bacteria containing the desired bacterial artificial chromosome (BAC) were grown on Luria Bertani agarplates (LA) containing chloramphenicol 20 µl/ml and incubated over night at 37°C before storage in refrigerator at 4°C.

Detailed protocol

1. A bacteria colony was picked from the plate and incubated in 2 ml LB medium containing 20 µl/ml chloramphenicol. Shaken vigorously (200 rpm) for 6-8 h at 37°C.

2. The culture was added to 400 ml LB medium containing 20 µl/ml chloramphenicol and shaken vigorously over night (12-16 h) at 37°C.

3. Centrifuged at 6,000 x g for 15 min at 4°C and discard supernatant.

4. Pellet was resuspended in 20 ml Buffer P1 containing RNase A.

5. Lysis of pellet was performed by adding 20 ml alkaline lysis Buffer P2 and mixing thoroughly for 4 min 30 sec.

6. Solution was neutralized and lysis was stopped by adding 20 ml 4°C Buffer P3 (adjusted to pH=5,5 with acetic acid) followed by mixing until lysate turned white. Incubated on ice for 30 min.

7. Centrifuged at 20,000 x g for 30 min at 4°C.

8. Filtered through prewetted gauze and collected supernatant in new centrifuge tubes.

9. Precipitated DNA by adding 42 ml isopropanol and centrifuge at 15,000 x g for 30 min at 4°C and carefully discarded supernatant.

10. DNA pellet was redissolved in 500 µl TE buffer, pH 8,0.11,5 ml Buffer QBT was added to provide optimal DNA binding conditions.

11. QIAGEN tip-500 was equilibrated by applying 10 ml Buffer QBT and allowing column to empty by gravity flow.

12. Applied DNA solution to column and discarded flow through.

13. Column washed with 2 x 30 ml Buffer QC to remove DNA binding proteins.

14. DNA was eluted by adding 15 ml 65°C Buffer QF and collected in small centrifuge tubes.

15. DNA was precipitated by adding 10,5 ml isopropanol, mix and centrifuge at 15,000 x g for 30 min at 4°C. Supernatant was carefully discarded.

16. Marked position of pellet in tube. Washed pellet by adding 10 ml 70% EtOH and centrifuged at 15,000 x g for 15 min at 4°C.

17. Discarded supernatant, air-dried for 10 min and redissolved DNA in 500 µl 10mM Tris-HCl buffer, pH 8,5.

18. Tubes shaken vigorously over night at 4°C and DNA concentration and purity measured on NanoDrop 3000 (ThermoFisher Scientific, MA, USA).

3.2.5 Nick translation

The isolated BAC plamids were tagged with the appropriate fluorophore by nick translation. A master mix containing 5 µl 10x Nick Translation-buffer, 5 µl dNTP-mix (0,5 mM dATP, dCTP, dGTP, 0,1 mM dTTP), 5 µl DTT (100 mM), 5 µl DNase (50 ng) was added to 1 µg plasmid DNA in addition to 1 µl of desired fluorophore-tagged dUTP (Green Flurorescent Protein (GFP), Orange Fluorescent Protein (OFP), Blue Fluorescent Protein (BFP), TexasRed, or InfraRed). After addition of 1 µl DNA polymerase I (10 U/µl, ThermoFisher Scientific) the tubes were incubated at 14°C over night.

Reactions were stopped by adding 5 µl EDTA (pH 7.5) and probes were stored at -20°C.

3.2.6 Precipitation and preparation of nick translated probes

Two pairs of nick translated probes (or triplet+pair) with 2,5 µl of each probe were mixed with 2,5 µl ssDNA (1 mg/ml), 2,5 µl tRNA (1 mg/ml) , 2,5 µl cotDNA and volume adjusted to 200 µl with dH_2O. 20 µl 3M NaAc (pH 5) and 660 µl ice cold abs. EtOH were added, the tubes vortexed and spun down before incubation at -70°C over night.

The precipitated probes were centrifuged at 13400 rpm, 20 mins, at 4°C. After carefully removing the supernatant, pellet was washed with 95 µl 70% EtOH (4°C) and centrifuged for 15 mins with aforementioned settings. The supernatant was discarded and the pellet air dried at room temperature. The pellet was redissolved in 12,5 µl 55% Hybmix. The probemix was finally shaken at room temperature for 2-3 hours followed by 1 hour in 56°C water bath.

3.2.7 Fluorescence in situ hybridization of imprint

FISH on imprinted cell lines was performed according to the following protocol:

1. The slides (stored in 70% EtOH at -20°C) were placed in 1,0xPBS for 5 min.

2. Transfered to 0,4% formaldehyde for 10 min.

3. Transfered to 0,1xPBS for 5 min.

4. Dehydrated slides for 3-5 mins in 70%, 95%, 2 x abs. EtOH.

5. Air-dried for 20 min.

6. Spun down probe mix.

7. Added 10 µl probe mix to each slide and sealed with 18 x 18 mm cover slip.

8. Sealed with Fixogum mounting glue.

9. Placed slides in 47°C for 20 min.

10. Loaded slides into Dako hybridizer running the programmed protocol (87°C for 4 min, 47°C over night). Alternatively, used heat plate 80°C for 10 min followed by incubation in humidity chamber at 47°C over night.

11. Placed slides in pre-warmed 37°C 4xSSPE and carefully removed cover slip. Incubated for 10 min.

12. Transferred to pre-warmed 47°C 4xSSPE for 10 min.

13. Dehydrated slides for 3-5 min in 70%, 95%, 2 x abs. EtOH.

14. Incubated in hexan:isopropanol (60:40) for 10 min.

15. Placed in isopropanol for 5 min.

16. Rehydrated slides for 3-5 min in abs EtOH, 95%, 70%.

17. Transferred slides to 0,1xPBS for 5 min.

18. Air-dried for 5 min.

19. Added a drop of vectashield mounting medium and cover slip. Sealed with nail polish and stored at 4°C.

3.2.8 Fluorescence in situ hybridization of paraffin sections

FISH on FFPE tissue was performed according to the following protocol:

1. Deparaffinized slides 2 x 10 min in xylene.

2. Rehydrated slides for 3-5 min in abs EtOH, 95%, 70%.

3. Air-dried for 5 min.

4. Antigen retrieval was performed by steaming in 10 mM citrate buffer for 45 min.

5. Let slides cool down in room temperature for 10 min.

6. Washed twice in 2 x saline-sodium citrate (SSC) buffer for 5 min.

7. Placed slides in dH_2O for 1 min.

8. Digested tissue with pepsin (0,5%, 10 mg/ml) for 9 min at 37°C in pre-warmed humidity chamber.

9. Washed slides twice in 2xSSC for 5 min.

10. Dehydrate slides for 3-5 min in 70%, 95%, 2 x abs. EtOH.

11. Air-dried

12. Added suitable volume probe mix dependent on area of tissue and added cover slip.

13. Denatured on heat plate at 80°C for 10 min followed by incubation in humidity chamber at 47°C over night.

14. Placed slides in 37°C 2xSSC with 0,1% IGEPAL CA-630 (Sigma Aldrich) for 2 min.

15. Transferred to 73°C 2xSSC 0,1% IGEPAL CA-630 for 2 min.

16. Let slides air-dry.

17. Mounted with vectorshield and cover slip. Stored at 4°C.

Statistics

Statistical analysis was performed in the statistical program R (R Core Team (2017) (108).

The Student's paired t-test was used to calculate differences between mean copy number between imprint and FFPE for the cell lines HCC1954 and HCC2218 and between estimated NGS values and FISH counts for patient material and cell lines.

Correlation was performed between mean FFPE and imprint copy number in order to identify an adjusting factor for copy number in sections (109).

4. Results

4.1 Validation of BAC probes

To verify correct probe localization, two or three overlapping probes nick-translated with different fluorophores were co-hybridized to fibroblasts. We lacked overlapping probes for p6 and p7, which target different sequences on chromosome 8. These had been previously validated in the lab. Probes p6 and p7 were thus only validated for signal quality and not specificity.

Figure 12A is an example of successful overlap between two pairs of probes. Not all probes were distinctly specific; we observed several scenarios in during image analysis which challenged the interpretation. Satellites can be one to several small signals in close proximity to the main signal. As seen in Figure 12B, two overlapping probes with the presence of an additional smaller signal from one probe close to the main signal reflects a split probe which would either lead to discarding of the probe or caution in interpretation of signals in the following experiments.

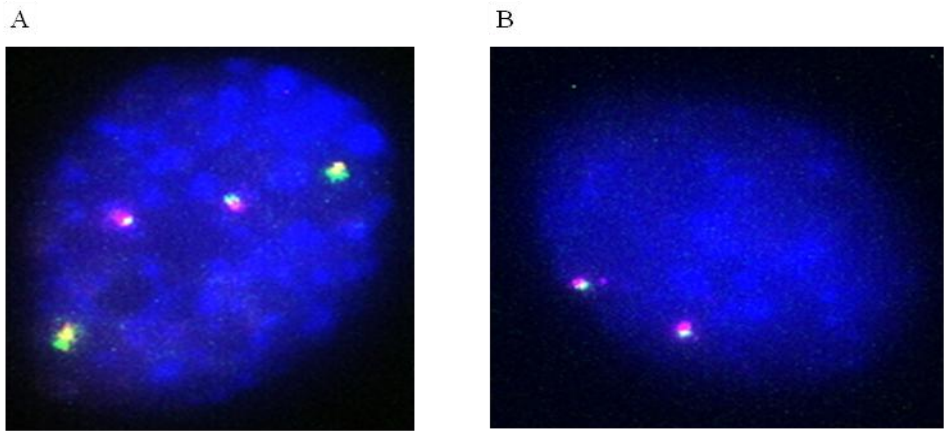


Figure 12: A: RP11-1087J7 (green) and RP11-1116O4 (orange), and RP11-878N15 (cyan) and RP11-450G15 (magenta) show overlapping signals. B: Probe pair CTC-859I3 (green) and RP11-939C24 (magenta) overlapping with the latter split into additional signal.

Probes which did not co-localize, had more than two signals present or had many satellites were deemed unspecific and discarded (Figure 13-14). The satellites were not always present as seen in figure 13B.

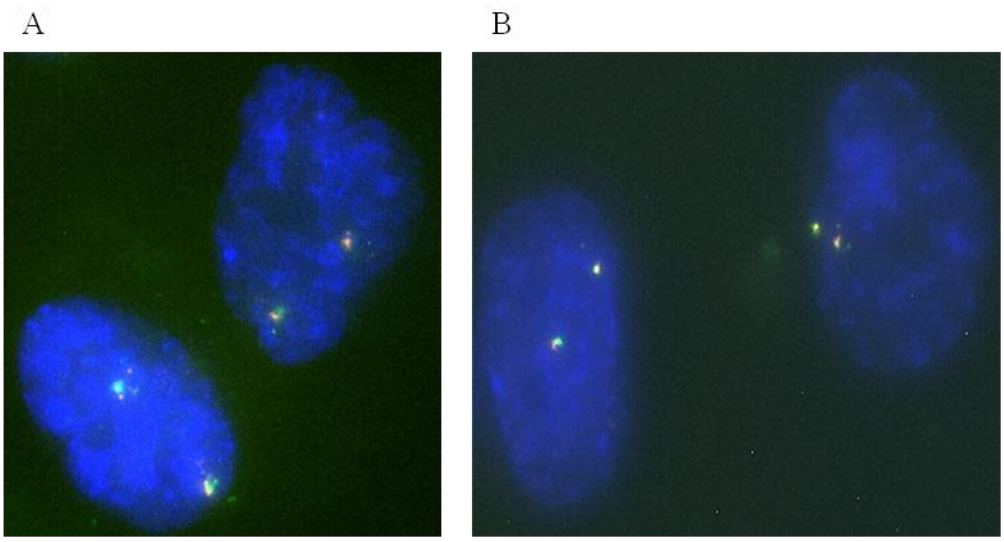


Figure 13: Overlapping probe pair RP11-1147E20 (green) and 355C15 (orange) with satellites.

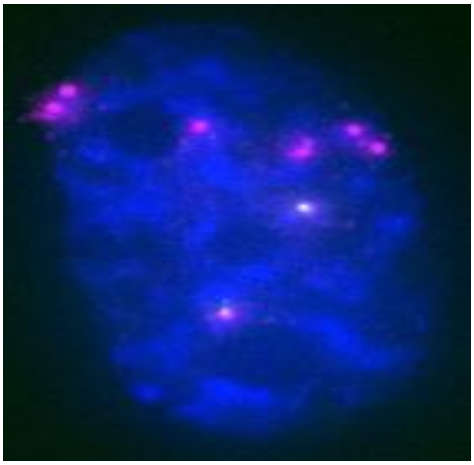


Figure 14: Overlap between probe pair CTD-3174P24 (green) and RP11-875M12 (magenta) in the center of image. Extra signals are due to unspecific binding.

4.2 Comparison of selected probes by FISH on imprint versus FFPE for cell lines HCC1954 and HCC2218.

We wanted to compare copy number for the 17 selected probes (Table 3) on imprint versus FFPE for the two cell lines HCC1954 and HCC2218. Two signals per nuclei were observed in all nuclei deemed countable. (Data not included). Figure 15 and 16 show the comparison between imprint (red boxes) and FFPE (blue boxes) for cell line HCC1954 and HCC2218 respectively, for the probes analyzed. Some probes were omitted from both cell lines due to insufficient staining of imprinted cells. From cell line HCC1954 probe p6, p9 and p12 and

from HCC2218 probes p2 and p16 were omitted due to insufficient staining. Probes are numbered as listed in Table 3. High amplification generated overlapping signals making accurate quantification impossible. Copy number is thus capped at 20.

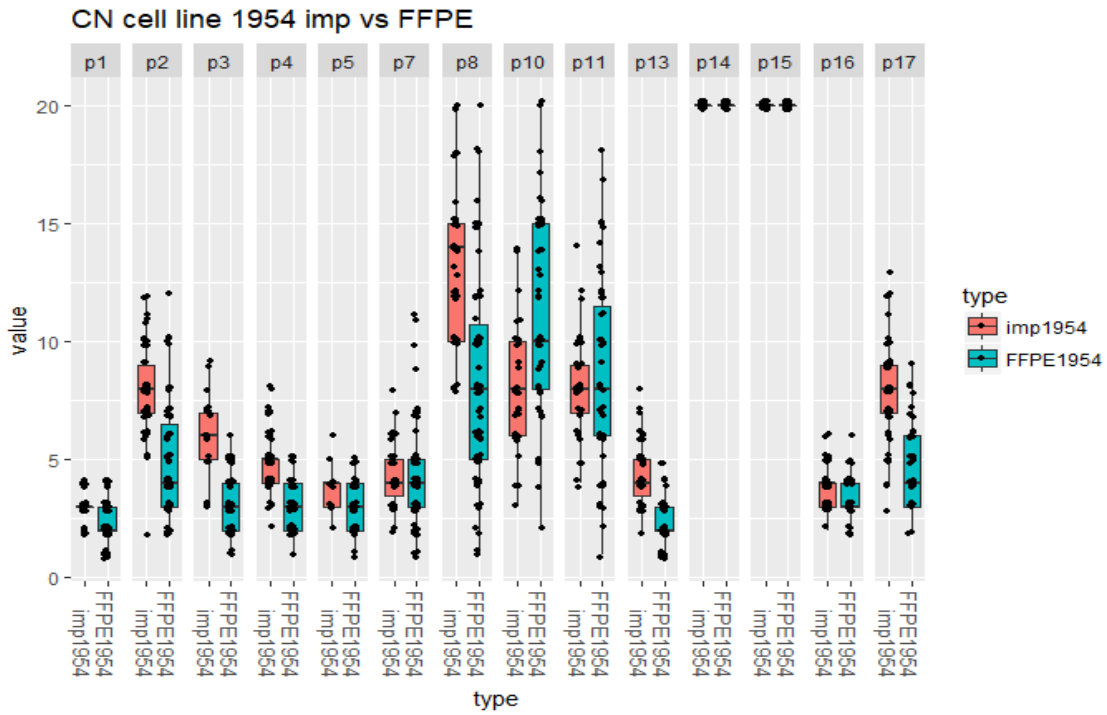


Figure 15: Box plot shows comparison between observed copy number in imprint and FFPE for cell line HCC1954.

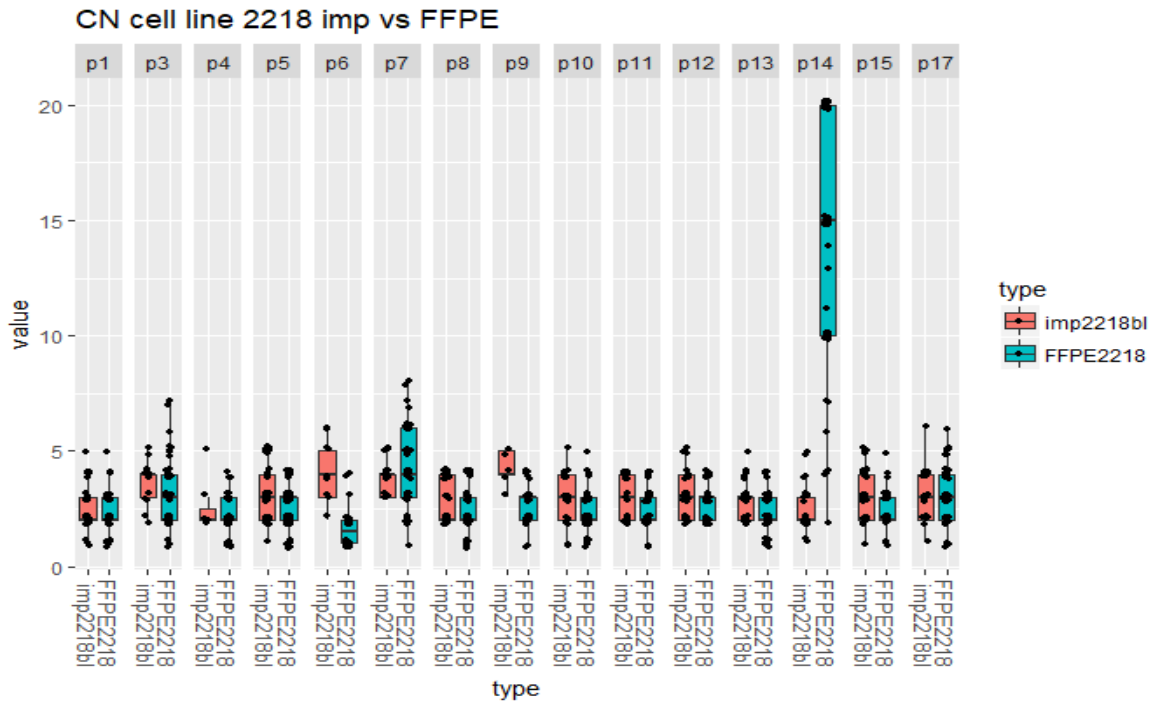


Figure 16: Box plot shows comparison between observed copy number in imprint and FFPE for cell line HCC2218.

The results show a clear trend of imprinted cells having a higher or equal copy number relative to FFPE cells with the notable exceptions of p10 (and to a lesser degree, p11) in HCC1954 and p14 (and to a lesser degree p7) for HCC2218.

A clearer picture emerged when illustrating the difference of copy number between imprint and FFPE cells as mean numbers as shown in Figure 17-18. Probes p14 and p15 in HCC1954 are capped at 20 reflecting non-quantifiable amplification. The dots for imprint and FFPE cells in the plot overlap at 20, but FFPE HCC1954 copy numbers are visible. For HCC2218 the dots for p17 overlap, but only the dot for FFPE cells is visible. While there is a large variance in assessed CN for each probe in HCC1954, the smaller discrepancy between counted CN for HCC2218 indicates a more systematic relationship between imprint and sectioned samples.

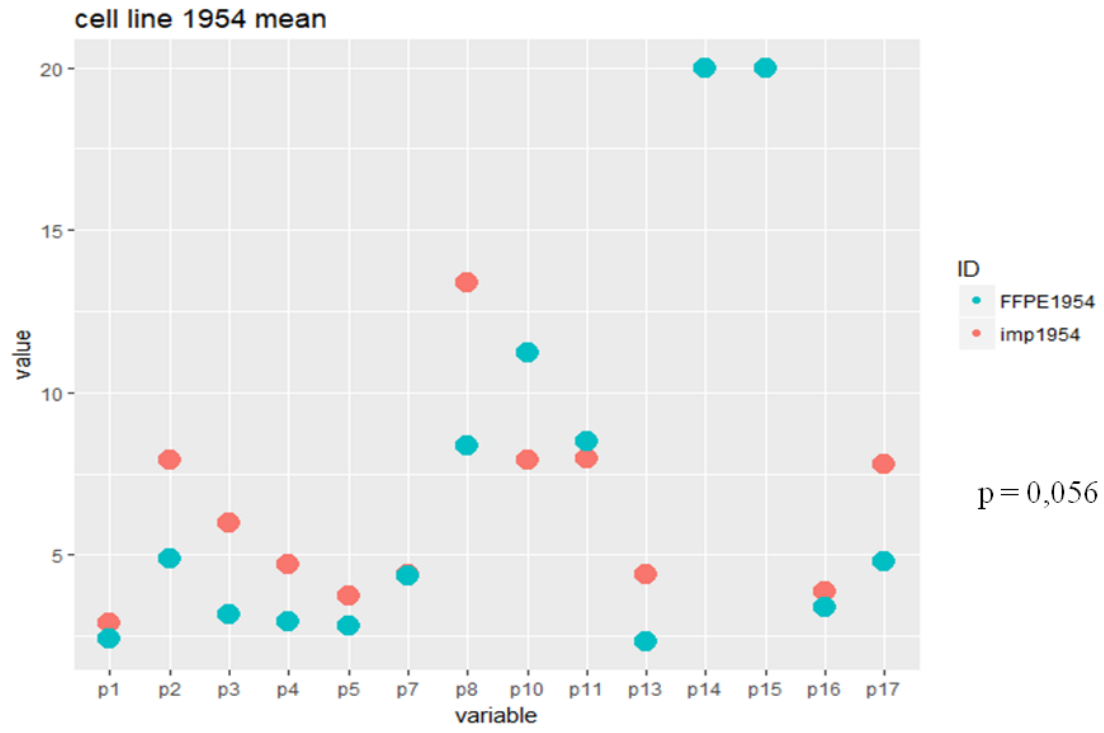


Figure 17: Comparison of mean copy number (on the y-axis) of probes (on the x-axis) between imprint and FFPE for cell line HCC1954. Copy numbers are capped at 20. Omitted probes are due to indeterminate data for imprinted cells.

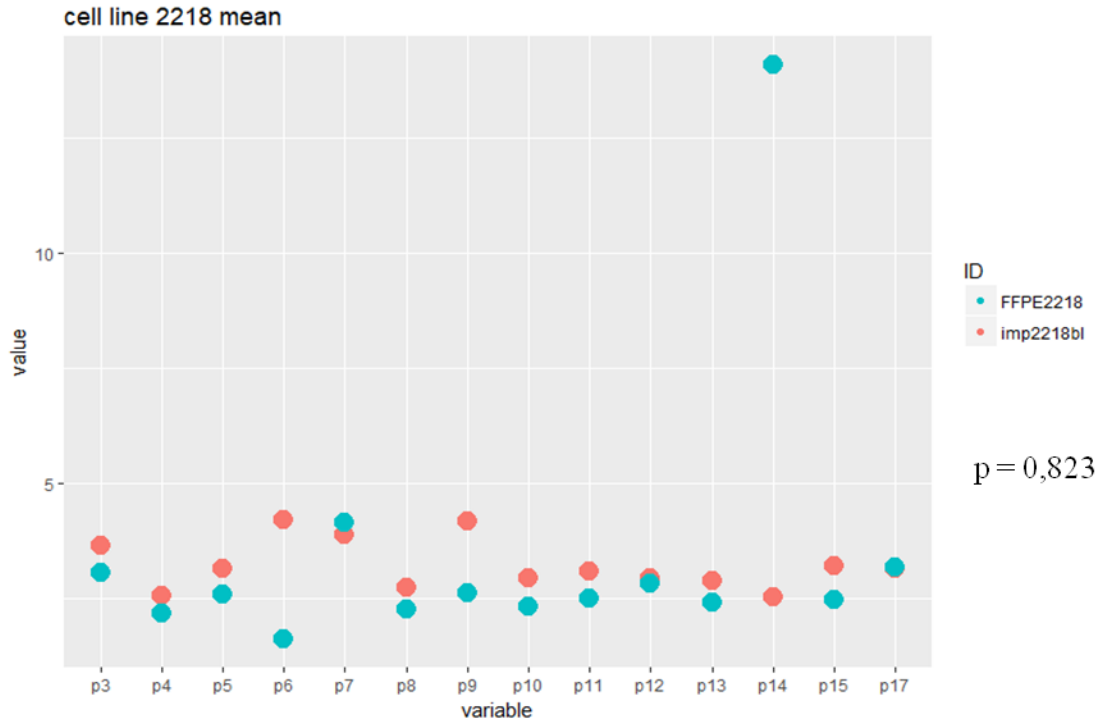


Figure 18: Comparison of mean copy number (on the y-axis) of probes (on the x-axis) between imprint and FFPE for cell line HCC2218. Copy numbers are capped at 20. Omitted probes are due to indeterminate data for imprinted cells.

The copy numbers for the different targeted sequences show substantial variation with a percentage difference ranging from -29 to +90% and an average of 32,5 %. The average difference in copy number is 1,1 with extremes of -3 and +3 (Table 4). Exclusion of p10 (see discussion) leaves the average copy number difference at 1,4 (calculation not shown).

Table 4: probes are listed p1-p17. Mean copy numbers for imprint and FFPE HCC1954 cell line. Difference in copy numbers are given as percentage and numerical value.

	mean CN imp1954	**mean CN FFPE1954**	**%diff 1954impFFPE**	**copy number change**
p1	2.93	2.43	20.3	0.5
p2	7.94	4.89	62.5	3.1
p3	6,00	3.17	89.5	2.8
p4	4.71	2.96	59.1	1.8
p5	3.73	2.82	32.1	0.9
p6	NA	1.7	NA	NA
p7	4.41	4.36	1.1	0
p8	13.4	8.37	60.1	5,0
p9	NA	2.38	NA	NA
p10	7.95	11.22	-29.2	-3.3
p11	7.97	8.49	-6.1	-0.5
p12	NA	2.38	NA	NA
p13	4.42	2.33	90	2.1
p14	20	20	0	0
p15	20	20	0	0
p16	3.88	3.4	14	0.5
p17	7.79	4.81	62	3,0
Average			**32.5**	**1.1**

Table 4 shows comparison of imprinted versus FFPE HCC2218 cells by mean copy number. The difference in CN between imprint and FFPE cells is presented as percentage and numerical value. The copy numbers for the different targeted sequences show substantial variation with a percentage difference ranging from -82 to +160% and an average of 22%. The massive difference in copy number between imprint and FFPE for p14 heavily skews the average for all probes to a seemingly negligible -0,2. Exclusion of the p14 probe puts the average difference between imprint and FFPE of 0,8 (calculation not shown). The striking anomaly of p14 is a topic in the discussion section.

Table 5 probes are listed p1-p17. Mean copy numbers for imprint and FFPE HCC2218 cell line. Difference in copy numbers are given as percentage and numerical value.

	mean CN imp2218bl	**mean CN FFPE2218**	**%diff 2218impFFPE**	**copy number change**
p1	NA	2.2	NA	NA
p2	NA	5.3	NA	NA
p3	3.64	3.07	18.7	0.6
p4	2.57	2.18	18	0.4
p5	3.14	2.58	21.6	0.6
p6	4.22	1.63	159.8	2.6
p7	3.88	4.16	-6.6	-0.3
p8	2.75	2.28	20.6	0.5
p9	4.2	2.64	59.3	1.6
p10	2.96	2.33	27	0.6
p11	3.09	2.52	22.4	0.6
p12	2.95	2.83	4.3	0.1
p13	2.89	2.42	19.6	0.5
p14	2.54	14.1	-82	-11.6
p15	3.21	2.48	29.5	0.7
p16	NA	12.16	NA	NA
p17	3.17	3.18	-0.5	0
Average			**22.3**	**-0.2**

A list of the number of counted nuclei for all probes on cell lines HCC1954 and HC2218 imprint and FFPE is included in ST1. Compared to analysis for HCC1954 imprints, many probes for HCC2218 imprints suffered from low number of countable cells. SF1 and SF2 illustrate plots portraying the standard deviation for signal counts on the cell lines showing a lower deviation in HCC2218 imprint indicating less heterogeneity.

There seemed to be heterogeneity in the cell lines with regard to CN as illustrated in Figure 19A. Figure 19B is an image taken during analysis of HCC1954 imprint cells undergoing mitosis which indicate lagging chromosomes during mitosis, which is again is a typical feature of breakage-fusion-bridges (BFB). This is an indication of this type of process being active in the cells, which again can explain heterogeneity within the cell line.

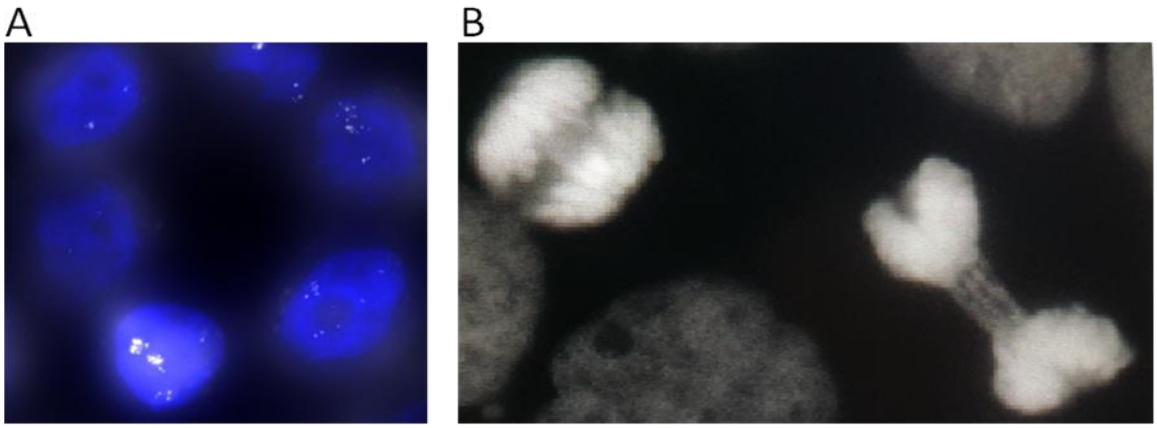


Figure 19 A: HCC2218 FFPE cells of similar size and DAPI intensity with variation in CN. B: DAPI image taken during analysis of HCC1954 imprint of possible BFB.

We performed a linear regression on the CN estimates for imprints and FFPE sections for both cell lines to see if there is a specific factor to adjust for CN loss by sectioning. The coefficient for HC1954 imprints versus FFPE was calculated to 1,01, although excluding the regions with high-level amplification reduced this number to 0,72. For the HCC2218 cell line, the coefficient was calculated to be -1,45, but was modified to 0,20 by excluding the CN for the p14 outlier. This significant difference indicates that no specified adjusting factor for CN loss in FFPE sections can be determined based on the data in this thesis.

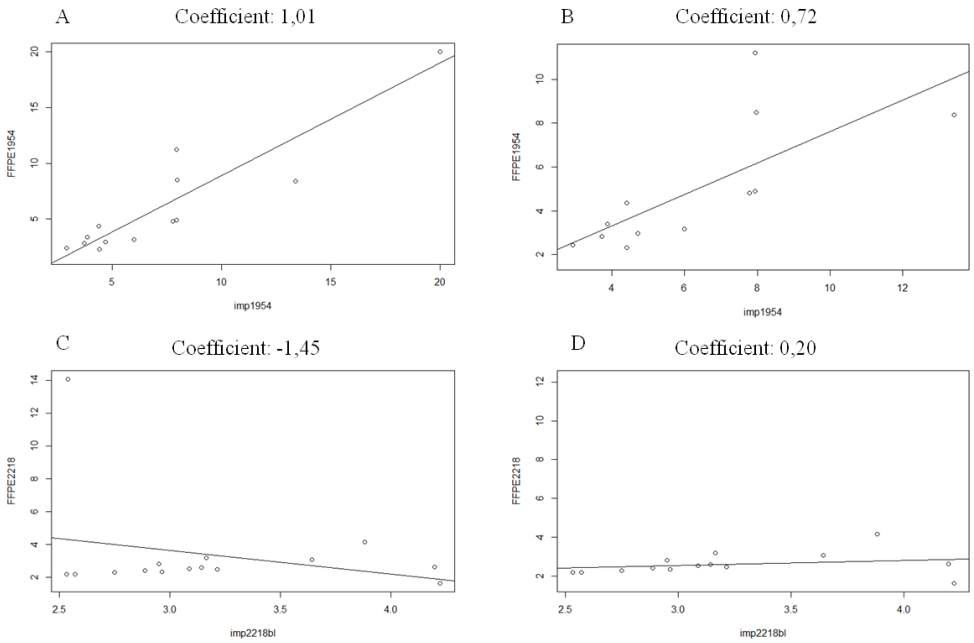


Figure 20: Illustration showing the correlation in CN estimates between imprints on the x-axis and FFPE sections on the y-axis for HCC1954 (A), HCC1954 excluding high-level amplifications (B), HCC2218 (C), and HCC2218 excluding the p14 outlier (D). The coefficients for the plots is shown above each plot

Next, we wanted to compare the copy numbers for both imprints and FFPE for both cell lines HCC1954 and HCC2218 to copy number estimates by NGS, as shown in figure 21. As evident by visual inspection there is large variation in the alignments for the different cell lines, but p-values indicate no significant difference in CN between cell lines and NGS. The alignment for HCC2218 imprints show very good correlation with a p-value >0,91. Mean CN by FISH and CN estimated by NGS for imprints and FFPE of both cell lines are listed in ST2.

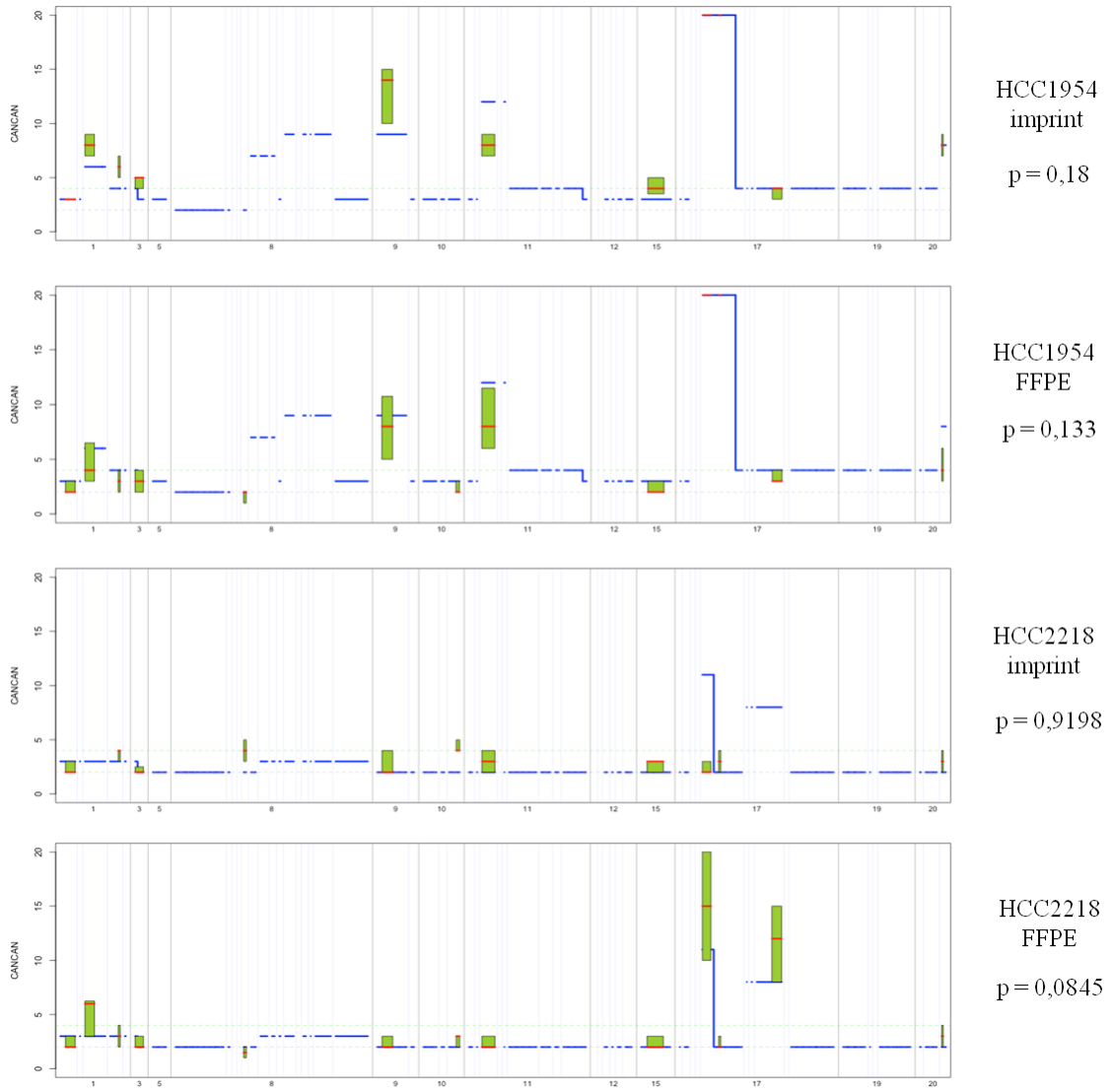


Figure 21: Alignment of CN by FISH against CN estimated by targeted sequencing by NGS for imprints and FFPE of cell lines HCC1954 and HCC2218. CN by FISH is shown by a green box with median CN marked as a red line while the top and bottom represent the 75th and 25th percentile respectively. Mean CN estimated by NGS is illustrated by a blue line. Chromosome number is shown on the x-axis, CN is shown on the y-axis. p-value for statistical comparison of CN by FISH and NGS is shown for each plot.

4.3 FISH on patient samples

Tissue biopsies from 14 patients were selected and stained with H&E for verification and localization of tumor cells. Biopsy from one patient was omitted due to problems sectioning the fat rich tissue. The remaining 13 biopsies were analyzed for copy number with FISH using 17 validated probes (CTC-859I3 from FISH on cell lines was replaced by RP11-939C24 due to weak signal-to-noise ratio, see Table 3). The mean copy numbers from FISH on patient samples are summarized in Table 6. Current guidelines for HER2 status determination by FISH state that six or more copies reflect amplification (110, 111). We therefore defined CN between 2 and 6 as "gain" and CN from 6 and upwards as "amplification". In table 7, all copy numbers defined as amplification (above this threshold) are marked red. CNs equal or less than 1,5 are marked green to illustrate regions with deletions. CN is capped at 20.

Table 6: Patients are designated by their Osl2-number. Probes p1-p17 listed with average copy number. nd: not determinable, marked blue. CN≥6 is marked red and CN≤1,5 is marked green.

Osl2#	**1116**	**1117**	**1119**	**1122-5**	**1122-7**	**1124**	**1129**	**1138**	**1162**	**1171**	**1210**	**1240**	**1163**
p1	2,8	3,8	1,8	1,8	1,5	2,5	1,7	1,6	3,1	1,3	1,8	2,3	nd
p2	2,4	3,4	2,7	2,6	2,1	2,6	2,4	2,2	2,6	2,0	1,8	2,6	nd
p3	3,3	4,5	3,3	3,1	1,9	3,0	2,0	2,2	2,5	1,9	2,0	3,8	nd
p4	2,4	4,0	1,0	1,6	1,5	1,9	2,1	2,2	2,4	1,8	1,8	2,8	2,4
p5	4,0	3,8	2,4	1,9	1,7	2,6	1,9	2,3	2,4	1,5	1,9	3,5	1,8
p6	1,1	1,5	6,6	1,8	1,0	1,2	1,1	20,0	2,5	1,8	13,1	4,5	2,2
p7	nd	1,8	3,0	1,8	1,5	2,6	2,8	2,4	1,8	nd	6,8	5,6	nd
p8	1,9	1,6	1,9	1,7	1,6	2,7	1,8	1,5	1,7	1,5	2,7	2,2	nd
p9	2,5	nd	1,8	1,6	1,7	nd	1,5	1,8	2,2	1,5	1,8	2,6	nd
p10	4,4	3,0	3,8	1,8	1,6	10,0	4,3	1,8	2,2	15,0	15,0	2,3	nd
p11	5,1	3,0	4,1	1,9	1,8	8,7	3,5	1,8	2,9	15,3	10,5	2,7	6,4
p12	3,4	1,5	2,0	1,7	1,4	1,6	2,5	1,7	2,6	1,7	2,4	2,3	nd
p13	1,5	4,9	2,6	1,5	1,3	2,7	2,0	1,8	2,0	1,9	1,9	2,0	nd
p14	1,9	2,5	3,5	1,8	1,6	20,0	20,0	1,9	2,7	2,1	2,1	20,0	nd
p15	1,1	nd	2,0	1,9	1,8	20,0	20,0	1,6	2,9	1,9	2,6	20,0	nd
p16	5,3	2,2	2,6	1,8	1,9	3,7	2,6	1,6	2,7	1,3	2,4	3,0	nd
p17	3,1	1,5	2,3	1,6	1,3	3,5	4,2	1,6	7,0	2,0	4,2	2,7	nd

For patient samples 1116, 1124 and 171 FISH was unsuccessful for one probe, while patient sample 1117 had two unsuccessful hybridizations. FISH results for most probes proved repeatedly suboptimal on tissue in patient sample 1163; only four of the hybridizations resulted in signals discernible from extensive background noise. This sample was therefore

omitted from sequencing. The samples 1122-5 and 1122-7 represent two biopsies from the same patient with DCIS. All other samples were cases of invasive carcinoma. In cases of possible deletions, the CN is set to a minimum of one copy. The number of counted nuclei in the different tumor samples for the selected probes was subject to high variation (ST3). Occasional weak signals, high level of background noise, and morphology of tumor samples influenced the analysis of the samples. Also included in the supplementary, is ST4 listing the standard deviation for signal counts on patient samples.

The copy numbers for each probe for each of the twelve breast cancer patient sample are shown in Figure 22. For each plot a horizontal line at CN of 2 and 6 marks normal CN and amplification respectively.

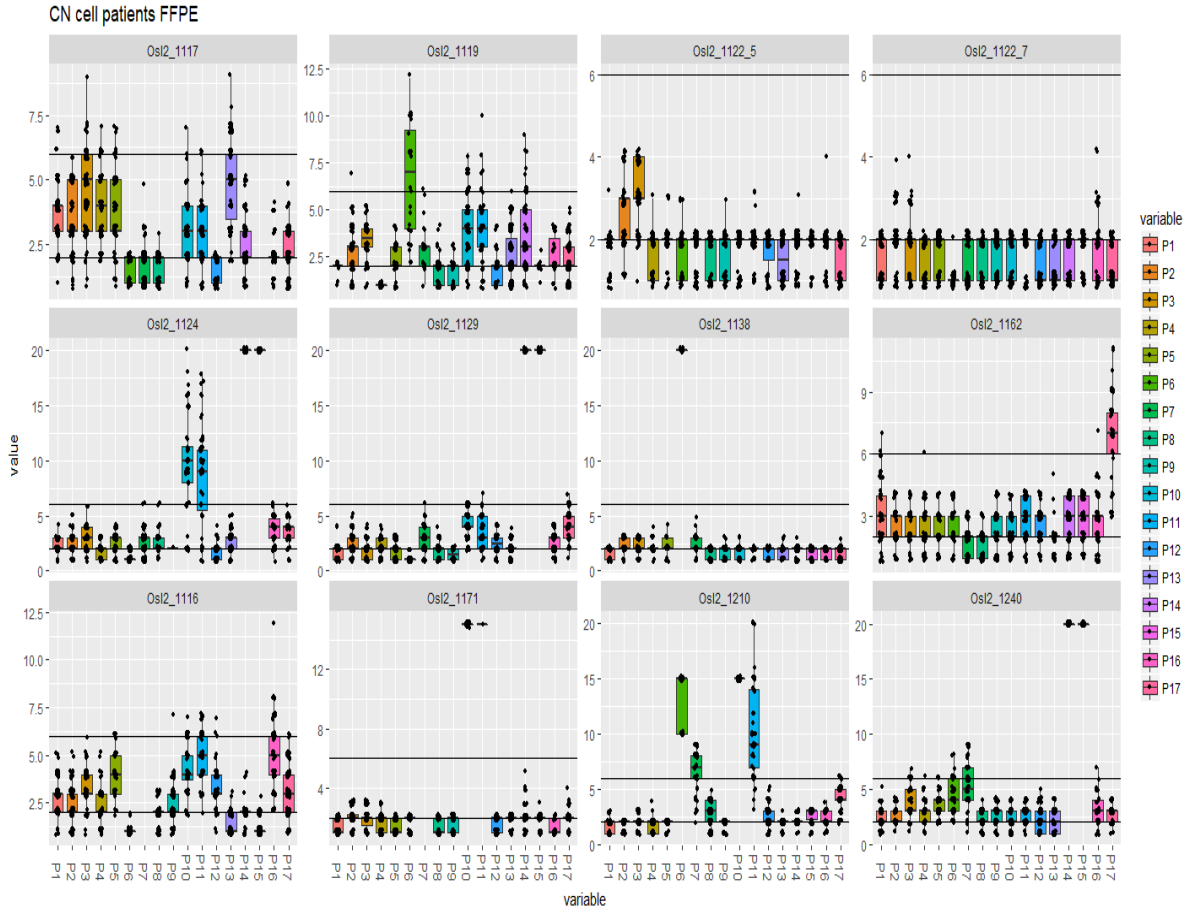


Figure 22: Box plot illustration of copy numbers for the targeted sequences of 17 color coded probes on the x-axis. Patient sample 1163 is omitted. Horizontal lines at 2 and 6 represent normal copy number and amplification respectively. Each dot represents the CN in a separate cell nucleus.

Copy number aberrations were observed for all patient samples. While some samples were dominated by normal CN or deletion (such as patient 1122), others were dominated by CN gain (such as patient 1116, 1117, 1119) or were more heterogenous with both gains and amplifications (such as patients 1124, 1129, 1240). We observed regions with amplifications in several patients and Figure 23 depicts two images of amplified regions. While the copy number in A could be quantified within a small margin of error, the opposite was true in B where the clustered amplification was impossible to count correctly and the CN was set to 20.

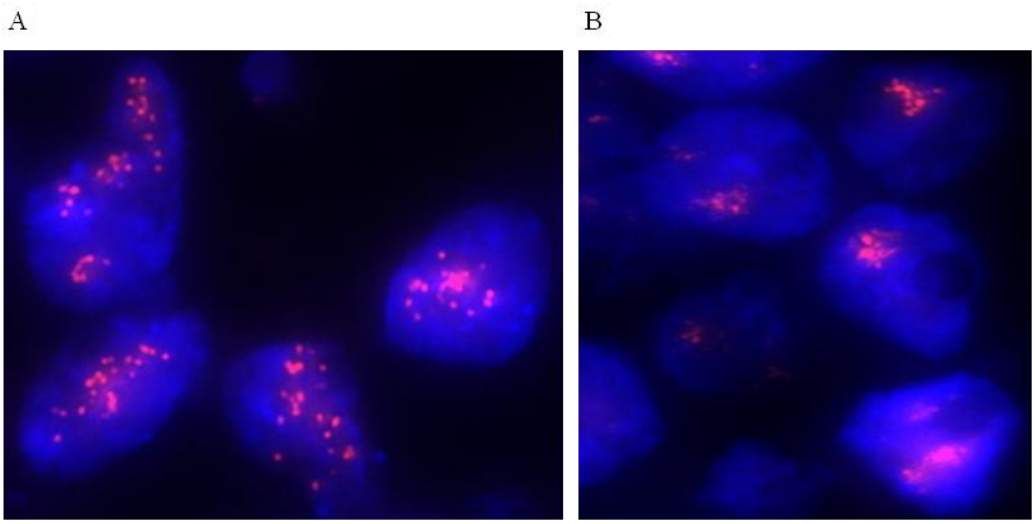


Figure 23: A: Amplification detected with p10 in Osl2_1171. B: Amplification detected with p15 in patient sample 1240.

An example of a patient sample (1210) with heterogeneous CN is shown in Figure 24 where we see CN gains, deletions and amplifications in the same cells.

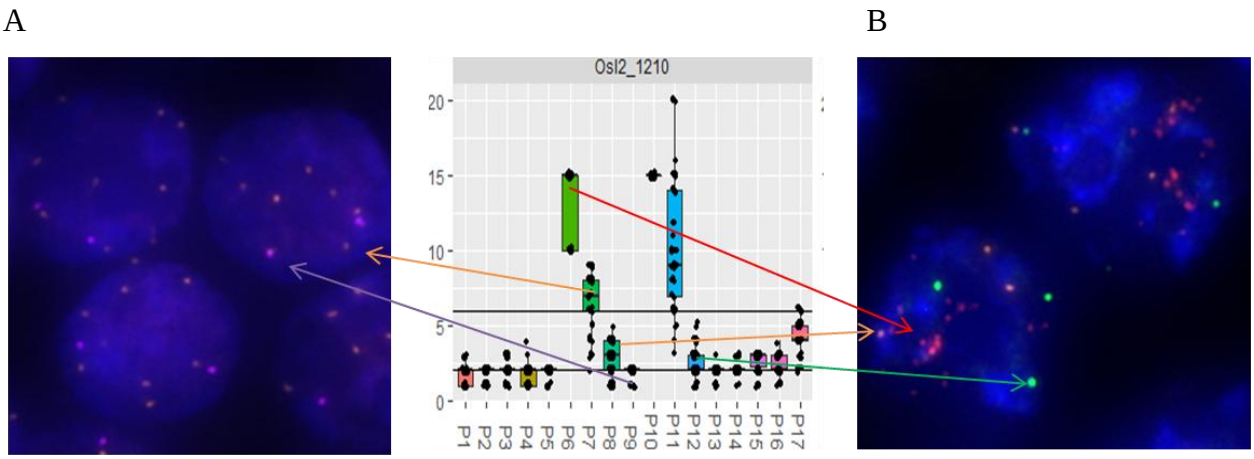


Figure 24: A: Amplification detected with p7 (orange), but not p9 (magenta) in patient sample 1210. B: Amplification detected with p6, gain with p8 and normal CN with p12.

Some patients were dominated by copy number gain as we see in sample 1129 where the majority of the probes had copy numbers between 2 and 6, except from p14 and p15 (Figure 25) which were highly amplified. The latter probes detect sequences in proximity to the *HER2* gene on chromosome 17. The same pattern was observed in patient sample 1240 (Figure 26).

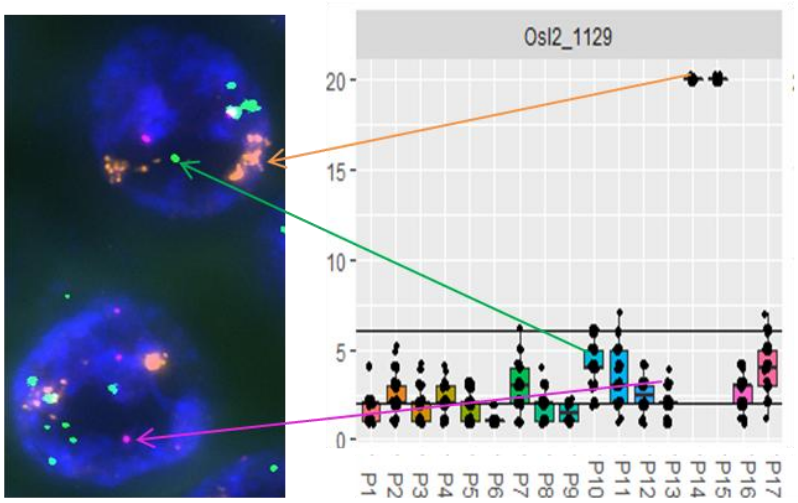


Figure 25: CNA with p14 (orange), p10 (green), p13 (magenta) in patient sample 1129.

Figure 26 illustrates the complexity of CN analysis with different CN in tumor cells with gains/amplifications and closer to normal CN in neighboring cell.

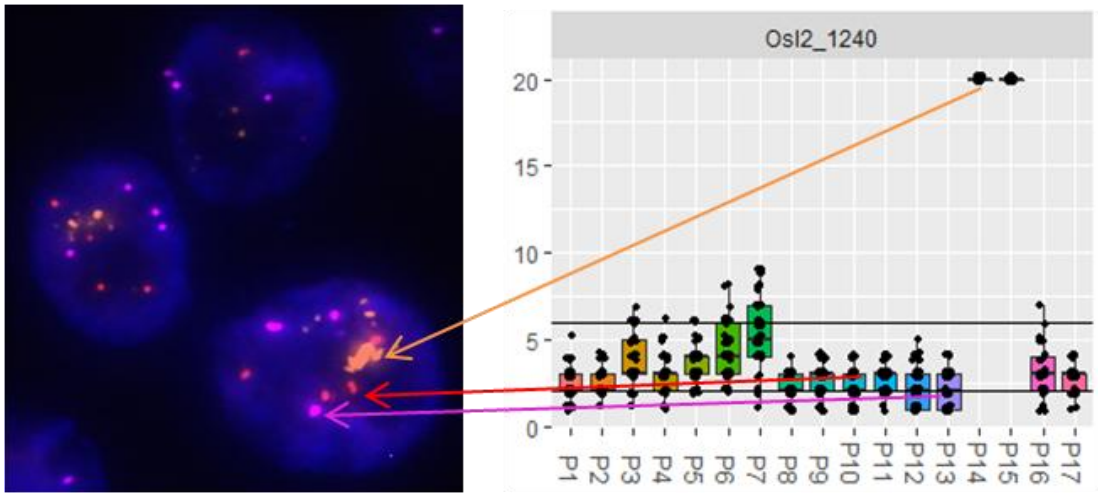


Figure 26: Amplification of p14, gain of p10 and p13 in patient sample 1240.

We observed that several patients had possible deletions in one or more regions. Figure 27 shows two tumor cells both having just one copy of p6 (red) and different CN for p7 (white).

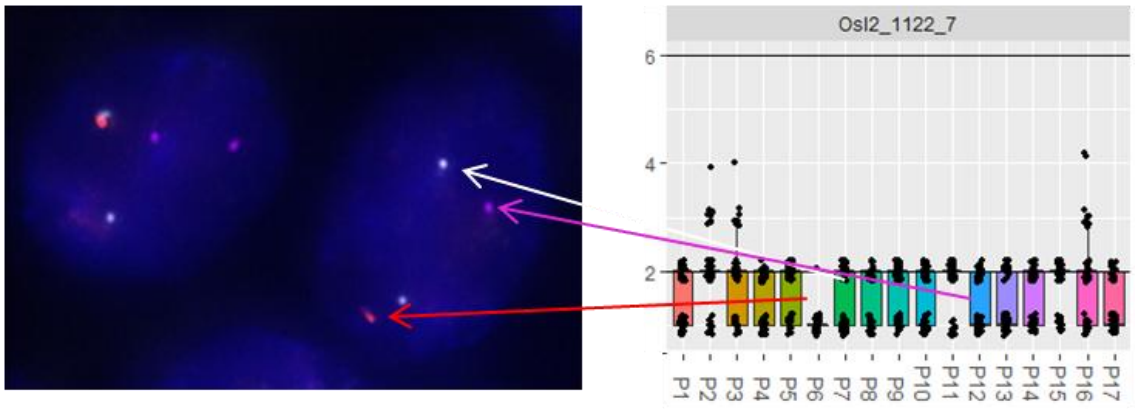


Figure 27: Possible deletion of p6 and p12, and normal CN for p7 in patient sample 122-7.

4.4 Comparison of copy number estimates by FISH and NGS

The patient samples included in the CN analysis by FISH were part of a larger study where the tumor DNA was analyzed by targeted sequencing by NGS for multiple genes frequently altered in breast carcinomas. The regions analyzed by FISH were, with four exceptions, covered by the targeted sequencing and allowed comparison of detected CN between the two methods as illustrated in Figure 27. Remaining five plots are illustrated in SF3, and a table listing the mean CN by FISH and CN estimated by NGS for all nine patient samples is included as ST5.

The CN counts by FISH and estimates by NGS largely seem to correlate with several estimates offering near match, but certain discrepancies are easily noticeable. The CN counts by FISH are both higher and lower for some regions in relation to NGS while other regions offer a better match. Gains and amplifications are detected by both methods with good correlation even though the amplification on chromosome 11 in sample 1210 is lower when estimated by FISH compared to NGS. CN in sample 1138 is systematically estimated higher by NGS reflecting gains in most regions, while corresponding CN by FISH is closer to two copies. Examples of the same tendency of NGS to estimate higher CN are also seen in sample 1210, though one of the amplifications is well correlated. We used a t-test to calculate p-values for the comparison of CN estimates by NGS to CN counts by FISH analysis. A value of 1 would indicate perfect correlation between the two methods of assessment, but the calculated values showed much variation reflecting the different levels of alignment seen in the plots.

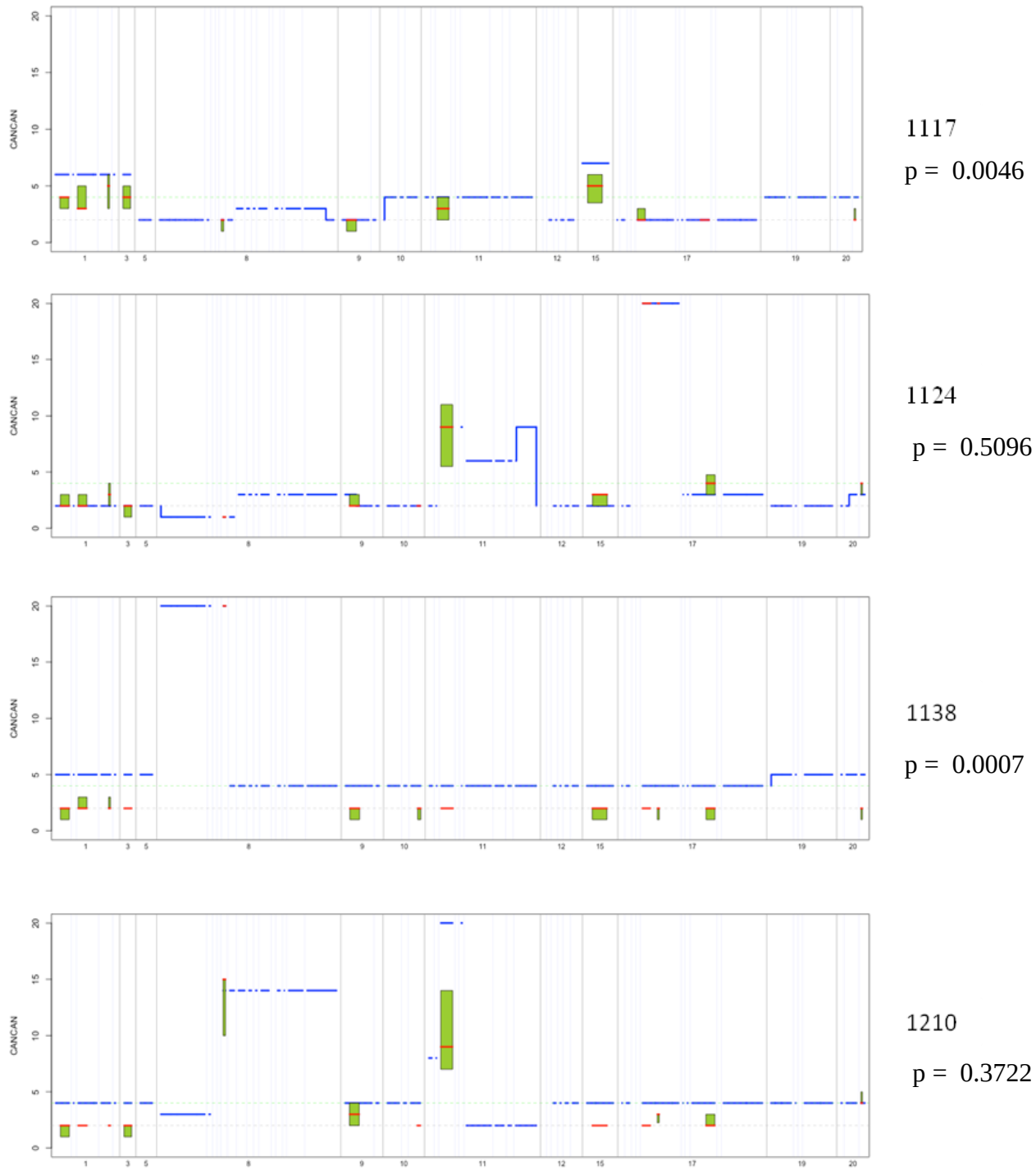


Figure 27: Alignment of CN by FISH against CN estimated by targeted sequencing by NGS in patient samples identified by numbers on the right hand side of the plots. CN by FISH is shown by a green box with median CN marked as a red line while the top and bottom represent the 75th and 25th percentile respectively. Mean CN estimated by NGS is illustrated by a blue line. Chromosome number is shown on the x-axis, CN is shown on the y-axis. p-value for statistical comparison of CN by FISH and NGS is shown for each plot.

5. Discussion

5.1 Validation of BAC probes

29 probes were sought validated with FISH, of which 17 probes were included in further analyses based on correct binding, signal intensity and uniformity of the signal. Most probes were analyzed in pairs as they were overlapping. However, probes for a sequence on chromosome 11 were analyzed as a triplet. Four sequentially overlapping probes for chromosome 17 were isolated, but analyzed as pairs in separate reactions. Also, no overlapping probes were available for the two probes annealing to chromosome 8; these probes had been previously validated in the lab.

Signal intensity was weak for several probes in the first round of FISH analysis and had to be repeated. If a probe signal was continuing to show low intensity, the probe was discarded with a preference for the overlapping probe. In cases where probes gave a split signal or abundant satellites, the probe was either discarded with a preference for the overlapping probe or the ambiguous signal was counted as one during analysis. This was necessary for some of the probes where the best probe was selected for further analysis. Some probes generated a split signal in some cells and a single distinct signal in a neighboring cell. Split signals may be a consequence of condensed chromatin inhibiting a part of the probe to bind. This leaves a gap in the sequence between the other parts of the probe which do hybridize, and the signal can appear as two distinct signals instead of one. Splits were usually evident as an additional smaller signal in near proximity to the main signal and were in these cases counted as one signal. The interpretation became more complicated in nuclei with multiple signals where some of the split signals might have been counted as separate signals. In other cells a true signal in close proximity to another signal may have been counted as one. This ambiguity will be somewhat negated by the number of counted cells for each probe. Only probes with a few satellites were tolerated, and nuclei with more diffuse signals were left out of the analysis.

The varying quality of the BAC probes used to bind the selected gene segments underlines the critical necessity of proper validation. Some probes generated either a few or multiple extra signals due to unspecific binding (Table 3). While several probes consistently generated intense, non-ambiguous signals, other probes produced signals with low signal-to-noise ratio. Weak signals, even among previously approved probes, seemed to be a somewhat random event complicating the experiments. Often a repeat of the FISH with or without a new

attached flurophore produced satisfying results. This did not seem to be a technical issue as three or four other probes in the same probe mix hybridized adequately to its target sequence on the same slide. The fact that all probes were analyzed with a standard protocol might have influenced results, and perhaps the optimization of a custom protocol would improve results for each probe. As all hybridizations were carried out with a mix of three, four or even five probes, the logistics involved in optimizing protocols were beyond the scope of this thesis. Hybridization efficiency and signal quality did not seem to be affected by the amount of different probes in a mix. Multiplexing with two, three, four or even five probes produced the same random results in all cell lines and biopsies. Another factor influencing the analysis was the apparent size of the probes. While most probes produced large signals, some of the probes generated tiny signals in comparison, readily hidden by background noise.

5.2 Comparison of copy number in imprint and FFPE for cell line HCC1954 and HCC2218

There was a trend of higher copy number in imprinted cells versus FFPE cells as evident from Figure 17-18 and Table 4-5. An average difference of one copy number between imprint and FFPE for HCC1954 gives strong indication that chromatin may be lost in some cells during sectioning. This difference was off by several copies for some of the probes, and this may be relevant in a therapeutic setting, which operates with cut-off values to determine diagnosis and treatment. For the HCC2218 cell line the outlier results for probe p14 skewed the average difference in copy number close to zero, but looking at the corresponding numbers for each probe individually, the CN difference largely followed the trend of higher CN in imprint relative to FFPE material. In general, the CN detected by the probes were lower and more homogeneous (see illustration of standard deviance in SF1 and SF2) in HCC2218 than HCC1954 reflecting different amount of CN alterations between the cell lines. The CN comparison between imprints and FFPE cells usually correlated well when CN alterations were low (i.e. close to 2 copies), but there was substantial variation in regions of gains and amplifications which introduced a level of uncertainty due to possibly overlapping signals. Considerable heterogeneity within breast cancer cell lines has been previously reported (112, 113), reflecting the range of varying CN for each locus estimated in our analyses. Additionally, for FFPE material, there is a higher probability of losing target DNA by sectioning if there are multiple copies of a sequence compared to few. The observation of more extensive CNA in HCC1954 compared to HCC2218 could help explain why the CN

observed in HCC1954 has a larger variation for each region than the corresponding region in HCC2218. The possible BFB event depicted in Figure 19 B in the HCC1954 imprint cell line indicates genomic instability in the cell line; if this mechanism has caused some of the observed amplifications, we would expect cell-to-cell variation.

Much of the observed variance is due to heterogeneity of the tumor cells. The varying results imply difficulties in specifying an adjusting factor to compensate for possible loss of CN due to sectioning, at least without more extensive data material providing more reliable results by increased data points. This is supported by the marked difference in linear regression coefficient calculated for the two cell lines. Including an increased number of cell lines with CN assessment for additional regions would possibly clarify if determining such an adjusting factor across cell lines is possible. While an increased number of data points would strengthen the statistical power of the data, it would also demand a massive effort to obtain these data, especially for the probes with a low signal-to-noise ratio where repeated hybridizations had to be done. The magnitude of such analyses would exceed the scope of this thesis, but would be of great value in a follow up study.

The protocol for FISH on imprint and FFPE cells differ significantly and may be a factor in the varying hybridization efficiency, signal intensity and background noise within the same round of hybridization. Some probes may hybridize more favorably to specific tissues under different circumstances. The cells involved were tumor cells which do not have normal chromatin arrangement (114, 115). It is possible that highly condensed chromatin may inhibit annealing of entire probes or parts of probes and account for lower cell to cell copy number estimates and varying signal intensity.

Results of comparison for probe p6, p9 and p12 for HCC1954 are not available due to indeterminate results of hybridization. Hybridization with probe p6 in imprints from both cell lines resulted in scattered background and a low signal-to-noise ratio leaving true signals hard to distinguish. Even though some nuclei enabled an adequate signal count, the total number of countable nuclei remained too low and hence the probe was left out of comparison with FISH on FFPE material. In the FFPE sample the signal was of sufficient strength with low background, reflecting the initial results from validation of the probe on imprinted fibroblasts. Due to low concentration of the isolated plasmid, which also may have been a factor influencing poor FISH results, a second round of plasmid isolation with improved yield was performed preceding FISH on patient material. Probe p9 (CTC-859I3) was, when analyzing

patient samples, replaced by the overlapping probe RP11-939C24 due to poor hybridization efficiency, low signal intensity and high level of background noise. There was an intense background stain in the FFPE samples as well, but signals in a minority of nuclei were easily countable.

The images taken for probe p10 on imprint were suboptimal. The 20-image z-stack was out of focus, potentially missing signals. This was evident by strong signals in the initiating z-stack image. Also, the signals from this probe varied between intense and faint in the same cell, complicating true signal identification. The ambiguity of the results for this probe in HCC1954 casts doubts about the true copy number count and hence the results for this probe in HCC1954 could be considered disregarded.

Other challenging factors included a low number of counted nuclei (ST1). This was primarily caused by low signal-to-noise ratio leaving few qualified nuclei with discernible signals. For the HCC1954 cell line this was especially relevant for probe p3 (21 counted nuclei) and probe p5 (11 counted nuclei) on imprints, where observed CN were variable, displaying a range of 3-9 copies and 3-6 copies respectively. For several probes in the HCC2218 imprints the number of counted nuclei was low. Probe p3 had only 14 countable nuclei in the images taken, while p4, p6, p7 and p9, had seven, nine, 17 and five respectively. This low nuclei count, even if observed CN were subjected to minor standard deviation, remains a source of reduced statistical power. They do, however, offer support to the trend set by the results from the other probes by demonstrating higher CN for imprinted cells.

While HCC1954 are cells that easily adhere to other material, HCC2218 are cells that grow in suspension and thus proved difficult to be made to adhere to glass slides. The slides containing imprinted HCC2218 cells had mainly a few isolated cells or small clusters of cells. The data for these cells were therefore extremely hard to obtain, but a second set of slides with HCC2218 cells with intermittent blood lymphoblast cells were acquired. During microscopic analysis of these slides the lymphoblasts acted as control cells with a normal copy number. They were, however, a complicating factor as these cells should be identified and omitted from analysis. In most situations this was an easy task as the lymphoblasts' nuclei were smaller and more regular in shape compared to the tumor cells. The tumor cells displayed nuclei of a wide variety of shapes and sizes, but also occasionally appearing very similar to lymphoblasts. Most probes were hybridized in a mix of four. This meant that due to increased copy number for at least one of the target sequences, the tumor cells were

unequivocally identified. However, heterogeneity of the tumor cells regarding copy number, shape and size, and the fact that for some slides only minor and sporadic CN increase was observed, misidentifications are probable to some extent. Cells not positively identified as tumor cells were left out, meaning this would mostly impact data by making fewer cells available for analysis. On rare occasions, the possibility of lymphoblasts with normal CN being interpreted as tumor cells would skew the data by lowering the average CN.

As illustrated in Figure 19 A, there was a significant variation of CN for the sequence targeted by probe p14 in multiple adjacent cells of FFPE HCC2218. The cells seem to be of similar size and display near equal intensity of the DAPI stain. The CN for these neighboring cells seem to differ substantially with one cell showing a CN of 2, while the other cells display gains or clustered amplifications. Although in the FFPE sample most cells had amplified CN for the probe p14, the opposite was true in the imprint cell line which only displayed a small gain in CN. The number of imprinted cells analyzed for this probe, however, was low due to previously stated problems with adherence to glass slides. This significant difference in observed CN should mandate repeated FISH analysis to confirm or reject the results, but this was omitted due to time constraints. A lack of validation of this suggests the results to be interpreted with caution.

NGS and FISH CN for both cell lines were found to overlap to a certain degree (Figure 21), however, the difference between NGS and FISH CN was generally larger in HCC1954 compared to HCC2218, again reflecting the observed variation for each probe in this cell line. Alignment of CN estimated by NGS to CN from FISH in HCC1954 revealed an estimated equal or slightly higher CN for most probes. Probes detecting CN gains and amplifications produced more divergent results. HCC2218 displayed less diversity in CN for most sequences and the comparison between FISH and NGS was more aligned (p=0,9198 for HCC2218 imprint). There are, however, two amplified regions which present with somewhat higher CN when analyzed by FISH in HCC2218 FFPE.

5.3 FISH on patient samples and comparison with CN estimate from targeted sequencing

Guidelines for assessing *HER2* amplification without the use of a centromeric control probe state that CN of six or more reflect amplification (110, 111). By extending this recommendation to include all regions analyzed, amplifications of one or more regions were

detected in seven patient samples (excluding sample 1163). Three of the patient samples had amplifications of segments in proximity to *HER2* and these correspond to the patient data in Table 2. Patient sample 1119 had a HER2 score of 2+ by IHC, and correspondingly an elevated CN by FISH for probe p14 in proximity to *HER2*. An average FISH score of 3,5 obscures the cell to cell variation seen in the microscope; a high proportion of cells had CN between 6 and 8, and some even up to 9. Gains were observed to a varying degree in all patient samples, although in some samples (1122-7 and 1171) gains were only detected in a minority of the cells in a few regions. This, as well as the large diversity in observed CN for some loci (especially evident in samples 1117, 1119, and 1124), may reflect intratumoral heterogeneity analogous to the cell-to-cell variation seen in the HCC1954 and HCC2218 cell lines. Intratumoral heterogeneity is a well-know feature in breast carcinomas and seem to be more prominent in some tumors (22).

Distinguishing nuclei with loss of both copies of a region from nuclei with unsuccessful hybridization posed an analytical challenge. Because of this, deletions were only detected in cells with a minimum of 1 copy. In some cases (i.e. probe p4 in patient sample 1119 and probe p6 in patient sample 1122-7) where CN is set to one, no nuclei were non-ambiguously observed to have more than one copy, but a great proportion of nuclei were blank, even if they were been positive for other probes. The images in Figure 23-27 illustrate some of the complexity in CN analysis in tumor cells. For many probes the number of signals in a nucleus is beyond visual separation making an accurate count impossible. In cases of low CN, overlapping signals are discernible by careful analysis of the z-stack, but this is a time consuming exercise. However, heavily amplified segments still remained a challenge. An upper limit of 20 signals per nuclei was set as signal separation was challenging when there were high level amplifications. Another challenge when analyzing tumor cells by FISH is the separation of individual nuclei. In most samples, with the cases of DCIS a notable exception, a great number of nuclei were overlapping with cells proliferating in a disorganized pattern. Much time was spent using the microscope to locate a suitable region of the tumor for image generation and even then, some images provided a slim number of nuclei meeting the previously stated criteria for analysis. Highly irregular nuclei morphology and DAPI stain, and instances of overdigestion of FFPE tissue prior to hybridization, excluded an additional number of cells.

The number of counted nuclei in the different tumor samples for the selected probes was subject to significant variation. Occasional weak signals, high level of background noise, and

morphology of tumor samples introduced a level of uncertainty to take into account when comparing copy number count in FISH versus sequencing. ST3 lists the number of counted signals per probe for each tumor sample. For a majority of the patients an adequate nuclei count was achieved, but two patients offered a low count for one or two probes. Patient sample 1119, however, was subject to highly variable signal quality and hence, seven probes have countable nuclei below 30. For FISH analysis on the patient samples, probe p9 was replaced by its overlapping validation probe as previously stated. Several probes were by necessity incorporated with a different flurophore and many hybridizations were repeated once or twice to achieve a successful result. Still, some patient samples lack results from FISH analysis for one or two probes as illustrated in table 6. A special case was that of patient sample 1163 of which proved particularly difficult to obtain FISH signals for many of the probes. Even after multiple rounds of FISH, only four probes were subject to analysis. All other hybridizations were either deemed negative or riddled with intense background noise obscuring FISH signals for all but a few cells in each image. Reasons for this are probably tissue specific. Possible degradation of tissue prior to fixation or suboptimal fixation may be cause of poor hybridization efficiency or excessive background stain.

The FISH copy number and NGS estimated copy number did overlap to a certain degree, however for some patient samples (ie. 1117, 1138, 1129, 1162) the difference in copy number was significantly different ($p<0.05$). Comparison plots of FISH versus NGS for four selected patients in Figure 27 illustrate some of the challenges involved while the remaining plots are included in the supplementary (SF3). CN estimated by NGS are in many cases a close reflection of the CN analyzed by FISH both in identifying gains and amplifications, as well as normal CN. The CN in cases of seemingly diploid states estimated by FISH is often somewhat below 2. It is to be expected that a diploid CN would appear slightly reduced due to chromatin lost during sectioning. A proportion of the cells will present only one copy even if the majority have two. If none or few cells have a gain, the average number will be lower than two. Provided that the CN analysis revealing the diploid status, which is indicated by the blue line closely following the stapled line, is a correct representation of reality, the compared results are in concordance. However, for some samples (i.e. patient sample 1138) the NGS systematically overestimates CN for regions which by FISH are close to 2 copies. Accurately estimating CN by NGS in cases of few segments, as seen in patient sample 1138, poses an analytical challenge, although a similar profile with better concordance is seen in sample 1122. This is a complicated technical issue and reflects the NGS test still being in a state of

development. FISH, although it too has analytical pitfalls challenging even experienced pathologists, is considered the gold standard in CN analyses. The technical difficulties discussed in previous sections, as well as inter- and intraobserver variability introduce a certain degree of uncertainty, but the CN for all probes estimated by FISH are results of careful and thorough analyses. NGS is based in part on advanced mathematical algorithms and data processing as well as being performed using DNA from bulk-tumor samples. There is considerable amount of noise stemming from non-neoplastic cells which occasionally mask data. The fraction of tumor cells to normal cells in any given sample is also of critical importance as the algorithm relies on a certain fraction in later calculations and data processing is to some degree a process of trial and error.

Technical challenges in analysis of FISH

Occasionally the guidelines for FISH signal analysis were circumvented due to few cells with high-intensity signals. In cases of weak signals, low number of countable cells, and high level of background noise, signals were counted with caution, but the low number of counted cells results in weakened statistical power. Though the number of counted cells should preferably exceed 50, it was in some cases considerably lower (ST1 and ST3). The nucleus membrane should be intact and have clear borders. Nuclei with signs of over-digestion were analyzed with caution or omitted as these nuclei may have lost chromatin. This also applied to sectioned samples where nuclei presented with varying size. A small nucleus with dim DAPI stain indicates that parts of the nucleus have been sectioned off, possibly losing chromatin in the process, no longer reflecting the complete copy number. In some cases, the validity of a signal was either confirmed or rejected through analysis of the 20 image z-stack of the nucleus. By sifting through the stack of 20 images, split probes generating an extra signal could be identified by being in the same focus plane as the main signal. If the signals appeared to localize in different levels of the nuclei, they were likely reflecting multiple copies. Artifacts could also be identified by this method, as these generally were much larger and persisted through multiple levels of the z-stack. Nuclei with ambiguous signals were omitted from analysis. Overlapping nuclei was tolerated if the counted signals were non-ambiguously determined to reside in one or the other nuclei. Presence of signals in overlapping regions of two or more nuclei disqualified the nuclei from further analysis.

6. Conclusions

The proper validation and selection of probes for FISH are of critical necessity as many probes are unspecific, produce abundant satellite or generate a weak signal-to-noise ratio, all complicating visual interpretation of copy numbers.

While imprints from cell lines in large part, as expected, were estimated to have somewhat higher CN than FFPE sections from the cells lines, this was not universally true.

Additionally, the estimated CN in both imprints and FFPE sections displayed a high degree of heterogeneity further complicating matters. Regions with gains and, in particular, amplifications displayed a larger diversity between the two methods. A high mean CN was often accompanied by a wide spread in low and high CN values. A high degree of cell-to-cell variation reflects a need to analyze a sufficient number of cells. Specifying a factor to adjust for sectioning artifacts thus proved difficult based on the cell lines and data points in this thesis. However, the overall trend observed supports the logical assumption of higher observed CN for imprint cells relative to FFPE cells and could inspire a more comprehensive study involving additional cell lines and chromosomal regions to further investigate this relationship. As a side note, the tissue specific challenges seen in one of the patient samples, underlines the essential value of having multiple methods of CN assessment.

Detection of *HER2* amplification by IHC and/or FISH remains to date the only CN analysis included in routine diagnostics of breast carcinomas. In the anticipated future of personalized medicine, a more thorough molecular characterization of individual tumors will exceed the capacity of FISH analyses and require more time- and cost efficient technologies. While FISH is the current gold standard for CN analyses, the introduction of NGS into routine diagnostics is imminent. However, methodical, technical and analytical challenges still need to be worked out. The ongoing development of the targeted sequencing test to better differentiate breast carcinomas could potentially be modified to include other forms of neoplasia where CNA are the driver events. Augmenting the current capabilities of FISH analysis is the development and introduction of advanced image analysis software. This will allow easier identification of tumor cells, signal count and artifacts, providing more accurate analysis in a shorter time frame.

The CN estimates by NGS were in large part in agreement with the corresponding CN estimates by FISH. While minor discrepancies were observed throughout most patient samples, occasional larger discordance was identified. Systematic overestimating was evident

in some cases. While FISH is more reliable in detecting deletions which may be masked by noise in NGS, the latter is better at estimating massive amplifications which extend beyond the resolution of FISH analysis. The diagnostic implications of accurate quantification of such amplifications remain unclear. The impact of accurate CN analysis may be vital in cases where threshold values will determine therapy, accentuating the importance of extensive validation of NGS-based methods in diagnostics.

References

1. globocan. Breast Cancer: Estimated Incidence, Mortality and Prevalence Worldwide in 2012 2018, may 15 [Available from: http://globocan.iarc.fr/old/FactSheets/cancers/breast-new.asp.
2. Jemal A, Center MM, DeSantis C, Ward EM. Global patterns of cancer incidence and mortality rates and trends. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2010;19(8):1893-907.
3. Kreftregisteret. Cancer in Norway 2016 2017, october 23 [Available from: https://www.kreftregisteret.no/Generelt/Publikasjoner/Cancer-in-Norway/cancer-in-norway-2016/.
4. Rosner B, Colditz GA, Willett WC. Reproductive risk factors in a prospective study of breast cancer: the Nurses' Health Study. American journal of epidemiology. 1994;139(8):819-35.
5. Kelsey JL, Gammon MD, John EM. Reproductive factors and breast cancer. Epidemiologic reviews. 1993;15(1):36-47.
6. Ritte R, Lukanova A, Tjonneland A, Olsen A, Overvad K, Mesrine S, et al. Height, age at menarche and risk of hormone receptor-positive and -negative breast cancer: a cohort study. Int J Cancer. 2013;132(11):2619-29.
7. Russo J, Tay LK, Russo IH. Differentiation of the mammary gland and susceptibility to carcinogenesis. Breast Cancer Res Treat. 1982;2(1):5-73.
8. Megdal SP, Kroenke CH, Laden F, Pukkala E, Schernhammer ES. Night work and breast cancer risk: a systematic review and meta-analysis. European journal of cancer (Oxford, England : 1990). 2005;41(13):2023-32.
9. Hansen J, Stevens RG. Case-control study of shift-work and breast cancer risk in Danish nurses: impact of shift systems. European journal of cancer (Oxford, England : 1990). 2012;48(11):1722-9.
10. Morimoto LM, White E, Chen Z, Chlebowski RT, Hays J, Kuller L, et al. Obesity, body size, and risk of postmenopausal breast cancer: the Women's Health Initiative (United States). Cancer Causes Control. 2002;13(8):741-51.
11. Eliassen AH, Colditz GA, Rosner B, Willett WC, Hankinson SE. Adult weight change and risk of postmenopausal breast cancer. Jama. 2006;296(2):193-201.
12. Boyd NF, Rommens JM, Vogt K, Lee V, Hopper JL, Yaffe MJ, et al. Mammographic breast density as an intermediate phenotype for breast cancer. The Lancet Oncology. 2005;6(10):798-808.
13. Narod SA, Foulkes WD. BRCA1 and BRCA2: 1994 and beyond. Nat Rev Cancer. 2004;4(9):665-76.
14. National Cancer Institute. High-Penetrance Breast and/or Gynecologic Cancer Susceptibility Genes [
15. Masciari S, Dillon DA, Rath M, Robson M, Weitzel JN, Balmana J, et al. Breast cancer phenotype in women with TP53 germline mutations: a Li-Fraumeni syndrome consortium effort. Breast Cancer Res Treat. 2012;133(3):1125-30.
16. Hisada M, Garber JE, Fung CY, Fraumeni JF, Jr., Li FP. Multiple primary cancers in families with Li-Fraumeni syndrome. Journal of the National Cancer Institute. 1998;90(8):606-11.
17. Tan MH, Mester JL, Ngeow J, Rybicki LA, Orloff MS, Eng C. Lifetime cancer risks in individuals with germline PTEN mutations. Clin Cancer Res. 2012;18(2):400-7.
18. Guilford P, Humar B, Blair V. Hereditary diffuse gastric cancer: translation of CDH1 germline mutations into clinical practice. Gastric cancer : official journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association. 2010;13(1):1-10.
19. Giardiello FM, Brensinger JD, Tersmette AC, Goodman SN, Petersen GM, Booker SV, et al. Very high risk of cancer in familial Peutz-Jeghers syndrome. Gastroenterology. 2000;119(6):1447-53.
20. Beggs AD, Latchford AR, Vasen HF, Moslein G, Alonso A, Aretz S, et al. Peutz-Jeghers syndrome: a systematic review and recommendations for management. Gut. 2010;59(7):975-86.

21. Center UoRM. Health Enclypedia, Anatomy of the Breast,. 2018.
22. Polyak K. Breast cancer: origins and evolution. J Clin Invest. 2007;117(11):3155-63.
23. Zaha DC. Significance of immunohistochemistry in breast cancer. World journal of clinical oncology. 2014;5(3):382-92.
24. Warburg O. On respiratory impairment in cancer cells. Science (New York, NY). 1956;124(3215):269-70.
25. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.
26. Korkola JE, DeVries S, Fridlyand J, Hwang ES, Estep AL, Chen YY, et al. Differentiation of lobular versus ductal breast carcinomas by expression microarray analysis. Cancer Res. 2003;63(21):7167-75.
27. Turashvili G, Bouchal J, Ehrmann J, Fridman E, Skarda J, Kolar Z. Novel immunohistochemical markers for the differentiation of lobular and ductal invasive breast carcinomas. Biomedical papers of the Medical Faculty of the University Palacky, Olomouc, Czechoslovakia. 2007;151(1):59-64.
28. Helsedirektoratet. Nasjonalt handlingsprogram med retningslinjer for diagnostikk, behandling og oppfølging av pasienter med brystkreft. 2017(10).
29. Bocchinfuso WP, Korach KS. Mammary gland development and tumorigenesis in estrogen receptor knockout mice. Journal of mammary gland biology and neoplasia. 1997;2(4):323-34.
30. Li XX, Gao SY, Wang PY, Zhou X, Li YJ, Yu Y, et al. Reduced expression levels of let-7c in human breast cancer patients. Oncology letters. 2015;9(3):1207-12.
31. Ankrapp DP, Bennett JM, Haslam SZ. Role of epidermal growth factor in the acquisition of ovarian steroid hormone responsiveness in the normal mouse mammary gland. Journal of cellular physiology. 1998;174(2):251-60.
32. Hewitt SC, Korach KS. Progesterone action and responses in the alphaERKO mouse. Steroids. 2000;65(10-11):551-7.
33. Beral V. Breast cancer and hormone-replacement therapy in the Million Women Study. Lancet (London, England). 2003;362(9382):419-27.
34. Birrell SN, Butler LM, Harris JM, Buchanan G, Tilley WD. Disruption of androgen receptor signaling by synthetic progestins may increase risk of developing breast cancer. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2007;21(10):2285-93.
35. Courtin A, Communal L, Vilasco M, Cimino D, Mourra N, de Bortoli M, et al. Glucocorticoid receptor activity discriminates between progesterone and medroxyprogesterone acetate effects in breast cells. Breast Cancer Res Treat. 2012;131(1):49-63.
36. Fournier A, Berrino F, Riboli E, Avenel V, Clavel-Chapelon F. Breast cancer risk in relation to different types of hormone replacement therapy in the E3N-EPIC cohort. Int J Cancer. 2005;114(3):448-54.
37. Osborne CK, Yochmowitz MG, Knight WA, 3rd, McGuire WL. The value of estrogen and progesterone receptors in the treatment of breast cancer. Cancer. 1980;46(12 Suppl):2884-8.
38. 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 (New York, NY). 1987;235(4785):177-82.
39. Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science (New York, NY). 1989;244(4905):707-12.
40. Yarden Y, Sliwkowski MX. Untangling the ErbB signalling network. Nature reviews Molecular cell biology. 2001;2(2):127-37.
41. Cho HS, Mason K, Ramyar KX, Stanley AM, Gabelli SB, Denney DW, Jr., et al. Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. Nature. 2003;421(6924):756-60.

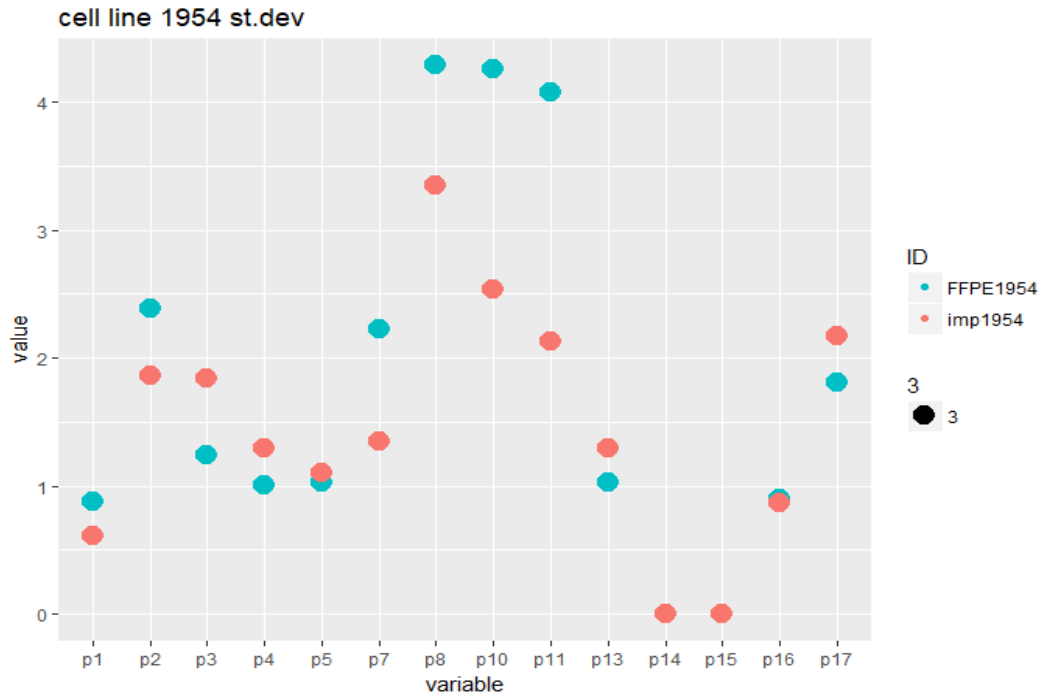
42. Klapper LN, Kirschbaum MH, Sela M, Yarden Y. Biochemical and clinical implications of the ErbB/HER signaling network of growth factor receptors. Advances in cancer research. 2000;77:25-79.
43. Alimandi M, Romano A, Curia MC, Muraro R, Fedi P, Aaronson SA, et al. Cooperative signaling of ErbB3 and ErbB2 in neoplastic transformation and human mammary carcinomas. Oncogene. 1995;10(9):1813-21.
44. Pinkas-Kramarski R, Lenferink AE, Bacus SS, Lyass L, van de Poll ML, Klapper LN, et al. The oncogenic ErbB-2/ErbB-3 heterodimer is a surrogate receptor of the epidermal growth factor and betacellulin. Oncogene. 1998;16(10):1249-58.
45. Carter P, Presta L, Gorman CM, Ridgway JB, Henner D, Wong WL, et al. Humanization of an anti-p185HER2 antibody for human cancer therapy. Proc Natl Acad Sci U S A. 1992;89(10):4285-9.
46. Madarnas Y, Trudeau M, Franek JA, McCready D, Pritchard KI, Messersmith H. Adjuvant/neoadjuvant trastuzumab therapy in women with HER-2/neu-overexpressing breast cancer: a systematic review. Cancer treatment reviews. 2008;34(6):539-57.
47. Romond EH, Perez EA, Bryant J, Suman VJ, Geyer CE, Jr., Davidson NE, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. The New England journal of medicine. 2005;353(16):1673-84.
48. Perez EA, Romond EH, Suman VJ, Jeong JH, Davidson NE, Geyer CE, Jr., et al. Four-year follow-up of trastuzumab plus adjuvant chemotherapy for operable human epidermal growth factor receptor 2-positive breast cancer: joint analysis of data from NCCTG N9831 and NSABP B-31. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2011;29(25):3366-73.
49. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, Goldhirsch A, Untch M, Smith I, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. The New England journal of medicine. 2005;353(16):1659-72.
50. Mariani G, Fasolo A, De Benedictis E, Gianni L. Trastuzumab as adjuvant systemic therapy for HER2-positive breast cancer. Nat Clin Pract Oncol. 2009;6(2):93-104.
51. Untch M, Rezai M, Loibl S, Fasching PA, Huober J, Tesch H, et al. Neoadjuvant treatment with trastuzumab in HER2-positive breast cancer: results from the GeparQuattro study. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2010;28(12):2024-31.
52. Buzdar AU, Ibrahim NK, Francis D, Booser DJ, Thomas ES, Theriault RL, et al. Significantly higher pathologic complete remission rate after neoadjuvant therapy with trastuzumab, paclitaxel, and epirubicin chemotherapy: results of a randomized trial in human epidermal growth factor receptor 2-positive operable breast cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2005;23(16):3676-85.
53. Nuti M, Bellati F, Visconti V, Napoletano C, Domenici L, Caccetta J, et al. Immune effects of trastuzumab. Journal of Cancer. 2011;2:317-23.
54. Spector NL, Blackwell KL. Understanding the mechanisms behind trastuzumab therapy for human epidermal growth factor receptor 2-positive breast cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2009;27(34):5838-47.
55. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. The New England journal of medicine. 2001;344(11):783-92.
56. Leica Biosystems. Bond Oracle HER2 IHC System [Available from: https://www.leicabiosystems.com/ihc-ish-fish/companion-diagnostics/theranostics/products/leica-oracle-her2-bond-ihc-system-usa-breast-only/.
57. Genemed Biotechnologies. HER2 green and Chromosome 17 Centromere Red FISH Probe [Available from: http://www.genemed.com/products/Fluorescent-In-Situ-Hybridization-Probes/HER2-Green-and-Chromosome-17-Centromere-Red-FISH-Probe.html.
58. Scaltriti M, Rojo F, Ocana A, Anido J, Guzman M, Cortes J, et al. Expression of p95HER2, a truncated form of the HER2 receptor, and response to anti-HER2 therapies in breast cancer. Journal of the National Cancer Institute. 2007;99(8):628-38.

59. Nagata Y, Lan KH, Zhou X, Tan M, Esteva FJ, Sahin AA, et al. PTEN activation contributes to tumor inhibition by trastuzumab, and loss of PTEN predicts trastuzumab resistance in patients. Cancer Cell. 2004;6(2):117-27.
60. Berns K, Horlings HM, Hennessy BT, Madiredjo M, Hijmans EM, Beelen K, et al. A functional genetic approach identifies the PI3K pathway as a major determinant of trastuzumab resistance in breast cancer. Cancer Cell. 2007;12(4):395-402.
61. Lu Y, Zi X, Pollak M. Molecular mechanisms underlying IGF-I-induced attenuation of the growth-inhibitory activity of trastuzumab (Herceptin) on SKBR3 breast cancer cells. Int J Cancer. 2004;108(3):334-41.
62. Nahta R, Takahashi T, Ueno NT, Hung MC, Esteva FJ. P27(kip1) down-regulation is associated with trastuzumab resistance in breast cancer cells. Cancer Res. 2004;64(11):3981-6.
63. Shields RL, Namenuk AK, Hong K, Meng YG, Rae J, Briggs J, et al. High resolution mapping of the binding site on human IgG1 for Fc gamma RI, Fc gamma RII, Fc gamma RIII, and FcRn and design of IgG1 variants with improved binding to the Fc gamma R. The Journal of biological chemistry. 2001;276(9):6591-604.
64. Musolino A, Naldi N, Bortesi B, Pezzuolo D, Capelletti M, Missale G, et al. Immunoglobulin G fragment C receptor polymorphisms and clinical efficacy of trastuzumab-based therapy in patients with HER-2/neu-positive metastatic breast cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2008;26(11):1789-96.
65. O'Sullivan CC, Connolly RM. Pertuzumab and its accelerated approval: evolving treatment paradigms and new challenges in the management of HER2-positive breast cancer. Oncology (Williston Park, NY). 2014;28(3):186-94, 96.
66. Gerdes J, Lemke H, Baisch H, Wacker HH, Schwab U, Stein H. Cell cycle analysis of a cell proliferation-associated human nuclear antigen defined by the monoclonal antibody Ki-67. Journal of immunology (Baltimore, Md : 1950). 1984;133(4):1710-5.
67. Urruticoechea A, Smith IE, Dowsett M. Proliferation marker Ki-67 in early breast cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2005;23(28):7212-20.
68. de Azambuja E, Cardoso F, de Castro G, Jr., Colozza M, Mano MS, Durbecq V, et al. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. British journal of cancer. 2007;96(10):1504-13.
69. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. Ann Oncol. 2011;22(8):1736-47.
70. Yerushalmi R, Woods R, Ravdin PM, Hayes MM, Gelmon KA. Ki67 in breast cancer: prognostic and predictive potential. The Lancet Oncology. 2010;11(2):174-83.
71. Giuliano AE. Mapping a pathway for axillary staging: a personal perspective on the current status of sentinel lymph node dissection for breast cancer. Archives of surgery (Chicago, Ill : 1960). 1999;134(2):195-9.
72. Lyman GH, Giuliano AE, Somerfield MR, Benson AB, 3rd, Bodurka DC, Burstein HJ, et al. American Society of Clinical Oncology guideline recommendations for sentinel lymph node biopsy in early-stage breast cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2005;23(30):7703-20.
73. Khasraw M, Bell R, Dang C. Epirubicin: is it like doxorubicin in breast cancer? A clinical review. Breast (Edinburgh, Scotland). 2012;21(2):142-9.
74. Pang B, Qiao X, Janssen L, Velds A, Groothuis T, Kerkhoven R, et al. Drug-induced histone eviction from open chromatin contributes to the chemotherapeutic effects of doxorubicin. Nat Commun. 2013;4:1908.
75. Fitzpatrick JM, de Wit R. Taxane mechanisms of action: potential implications for treatment sequencing in metastatic castration-resistant prostate cancer. European urology. 2014;65(6):1198-204.

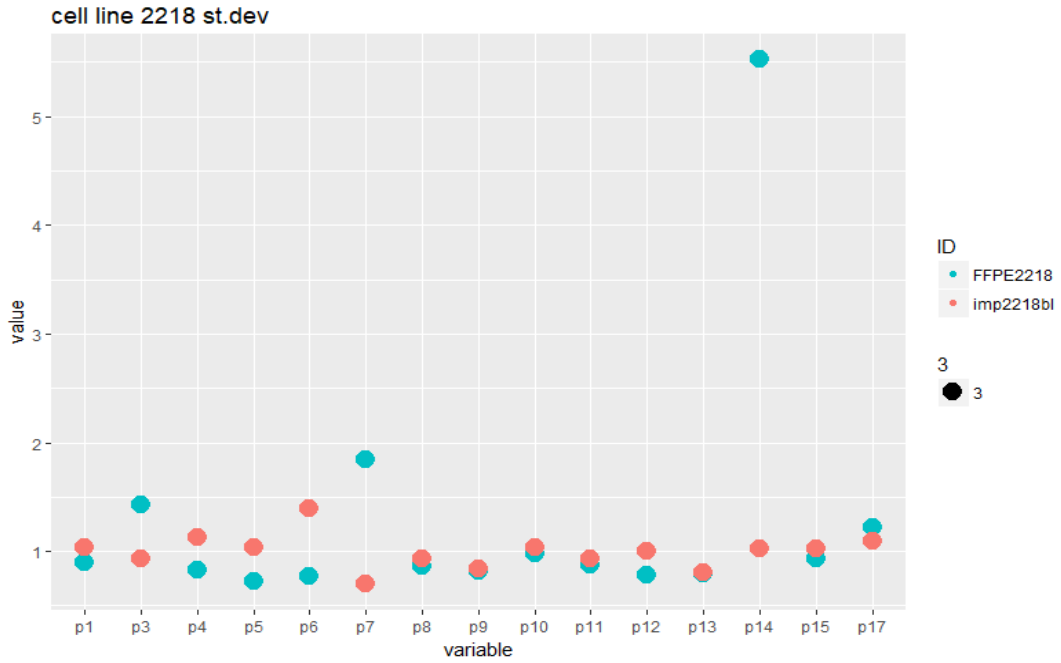
76. Baskar R, Lee KA, Yeo R, Yeoh KW. Cancer and radiation therapy: current advances and future directions. International journal of medical sciences. 2012;9(3):193-9.
77. Regan MM, Neven P, Giobbie-Hurder A, Goldhirsch A, Ejlertsen B, Mauriac L, et al. Assessment of letrozole and tamoxifen alone and in sequence for postmenopausal women with steroid hormone receptor-positive breast cancer: the BIG 1-98 randomised clinical trial at 8.1 years median follow-up. The Lancet Oncology. 2011;12(12):1101-8.
78. Coates AS, Keshaviah A, Thurlimann B, Mouridsen H, Mauriac L, Forbes JF, et al. Five years of letrozole compared with tamoxifen as initial adjuvant therapy for postmenopausal women with endocrine-responsive early breast cancer: update of study BIG 1-98. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2007;25(5):486-92.
79. Lonning PE, Eikesdal HP. Aromatase inhibition 2013: clinical state of the art and questions that remain to be solved. Endocrine-related cancer. 2013;20(4):R183-201.
80. Davies C, Godwin J, Gray R, Clarke M, Cutter D, Darby S, et al. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. Lancet (London, England). 2011;378(9793):771-84.
81. Davies C, Pan H, Godwin J, Gray R, Arriagada R, Raina V, et al. Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. Lancet (London, England). 2013;381(9869):805-16.
82. Teixeira MR, Pandis N, Heim S. Cytogenetic clues to breast carcinogenesis. Genes Chromosomes Cancer. 2002;33(1):1-16.
83. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747-52.
84. Kreike B, van Kouwenhove M, Horlings H, Weigelt B, Peterse H, Bartelink H, et al. Gene expression profiling and histopathological characterization of triple-negative/basal-like breast carcinomas. Breast Cancer Res. 2007;9(5):R65.
85. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. 2010;12(5):R68.
86. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2009;27(8):1160-7.
87. Hicks J, Krasnitz A, Lakshmi B, Navin NE, Riggs M, Leibu E, et al. Novel patterns of genome rearrangement and their association with survival in breast cancer. Genome Res. 2006;16(12):1465-79.
88. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012;486(7403):346-52.
89. Ali HR, Rueda OM, Chin SF, Curtis C, Dunning MJ, Aparicio SA, et al. Genome-driven integrated classification of breast cancer validated in over 7,500 samples. Genome Biol. 2014;15(8):431.
90. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. Nat Genet. 2013;45(10):1127-33.
91. Thompson LH, Schild D. Homologous recombinational repair of DNA ensures mammalian chromosome stability. Mutation research. 2001;477(1-2):131-53.
92. Shaw CJ, Lupski JR. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. Human molecular genetics. 2004;13 Spec No 1:R57-64.
93. Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. PathoGenetics. 2008;1(1):4.
94. Chen JM, Cooper DN, Ferec C, Kehrer-Sawatzki H, Patrinos GP. Genomic rearrangements in inherited disease and cancer. Seminars in cancer biology. 2010;20(4):222-33.

95. Lasko D, Cavenee W, Nordenskjold M. Loss of constitutional heterozygosity in human cancer. Annual review of genetics. 1991;25:281-314.
96. Haber JE. Partners and pathwaysrepairing a double-strand break. Trends Genet. 2000;16(6):259-64.
97. Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. Nat Rev Genet. 2009;10(8):551-64.
98. Blackburn EH, Gall JG. A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. Journal of molecular biology. 1978;120(1):33-53.
99. Blackburn EH, Chiou SS. Non-nucleosomal packaging of a tandemly repeated DNA sequence at termini of extrachromosomal DNA coding for rRNA in Tetrahymena. Proc Natl Acad Sci U S A. 1981;78(4):2263-7.
100. Harley CB, Futcher AB, Greider CW. Telomeres shorten during ageing of human fibroblasts. Nature. 1990;345(6274):458-60.
101. Artandi SE, Chang S, Lee SL, Alson S, Gottlieb GJ, Chin L, et al. Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice. Nature. 2000;406(6796):641-5.
102. Carballo JA, Pincheira J, de la Torre C. The G2 checkpoint activated by DNA damage does not prevent genome instability in plant cells. Biological research. 2006;39(2):331-40.
103. Olympus Lifescience. Anatomy of the Fluorescence Microscope [Available from: https://www.olympus-lifescience.com/en/microscope-resource/primer/techniques/fluorescence/anatomy/fluoromicroanatomy/.
104. Pardue ML, Gall JG. Molecular hybridization of radioactive DNA to the DNA of cytological preparations. Proc Natl Acad Sci U S A. 1969;64(2):600-4.
105. Illumina. TruSeq Custom Amplicon v1.5 [Available from: https://support.illumina.com.cn/content/illumina-marketing/en/products/truseq_custom_amplicon.html.
106. Nilsen G, Liestol K, Van Loo P, Moen Vollan HK, Eide MB, Rueda OM, et al. Copynumber: Efficient algorithms for single- and multi-track copy number segmentation. BMC Genomics. 2012;13:591.
107. Beatty BG, Mai, Sabine. FISH A Practical Approach. 2002:195-7.
108. R.
109. D.G A. Practical statistics for medical research. 1991.
110. Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2013;31(31):3997-4013.
111. Walker RA, Bartlett JM, Dowsett M, Ellis IO, Hanby AM, Jasani B, et al. HER2 testing in the UK: further update to recommendations. Journal of clinical pathology. 2008;61(7):818-24.
112. Voet T, Kumar P, Van Loo P, Cooke SL, Marshall J, Lin ML, et al. Single-cell paired-end genome sequencing reveals structural variation per cell cycle. Nucleic acids research. 2013;41(12):6119-38.
113. Szulwach KE, Chen P, Wang X, Wang J, Weaver LS, Gonzales ML, et al. Single-Cell Genetic Analysis Using Automated Microfluidics to Resolve Somatic Mosaicism. PLoS One. 2015;10(8):e0135007.
114. Ferraro A. Altered primary chromatin structures and their implications in cancer development. Cellular oncology (Dordrecht). 2016;39(3):195-210.
115. Brock MV, Herman JG, Baylin SB. Cancer as a manifestation of aberrant chromatin structure. Cancer journal (Sudbury, Mass). 2007;13(1):3-8.

Supplementary



SF1: Illustration of the standard deviance of CN estimates for cell line HCC1954. Probes are listed on the x-axis, st.dev. value is marked on the y-axis.



SF2: Illustration of the standard deviance of CN estimates for cell line HCC2218. Probes are listed on the x-axis, st.dev. value is marked on the y-axis

ST1: Number of counted nuclei for cell lines HCC1954 and HCC2218 imprint and FFPE. Number below or equal to 30 are marked green, while lacking data are marked blue.

	imp1954	**FFPE1954**	**imp2218bl**	**FFPE2218**
p1	40	58	nd	45
p2	54	63	nd	40
p3	21	66	14	58
p4	63	54	7	39
p5	11	56	49	89
p6	nd	43	9	40
p7	39	61	17	45
p8	45	70	36	50
p9	nd	40	5	33
p10	37	51	28	48
p11	37	51	34	44
p12	nd	42	41	35
p13	43	46	27	53
p14	39	39	26	61
p15	39	39	41	27
p16	57	40	nd	45
p17	48	47	26	44

ST2: Median CN by FISH and estimated CN by NGS for probes p1-p17 for HCC1954 and HCC2218 imprints and FFPE.

1954imp	FISH_median	NGS CN		1954FFPE	FISH_median	NGS CN
p1	3	2		**p1**	2	3
p2	8	2		**p2**	4	6
p3	6	2		**p3**	3	4
p4	5	2		**p4**	3	3.1
p5	4	NA		**p5**	3	NA
p6	NA	2		**p6**	2	2
p7	4	NA		**p7**	4	NA
p8	14	2		**p8**	8	9
p9	NA	2		**p9**	2	3
p10	8	NA		**p10**	10	NA
p11	8	2		**p11**	8	12
p12	NA	NA		**p12**	2	NA
p13	4	2		**p13**	2	3
p14	20	126		**p14**	20	126
p15	20	126		**p15**	20	126
p16	4	2		**p16**	3	4
p17	8	2		**p17**	4	8
2218imp	**FISH_median**	**NGS CN**		**2218FFPE**	**FISH_median**	**NGS CN**
p1	2	3		**p1**	2	3
p2	NA	3		**p2**	6	3
p3	4	3		**p3**	3	3
p4	2	2.1		**p4**	2	2.1
p5	3	NA		**p5**	3	NA
p6	4	2		**p6**	1.5	2
p7	4	NA		**p7**	4	NA
p8	2	2		**p8**	2	2
p9	4	2		**p9**	3	2
p10	3	NA		**p10**	2	NA
p11	3	2		**p11**	2	2
p12	3	NA		**p12**	3	NA
p13	3	2		**p13**	2	2
p14	2	11		**p14**	15	11
p15	3	2		**p15**	2	2
p16	NA	8		**p16**	12	8
p17	3	2		**p17**	3	2

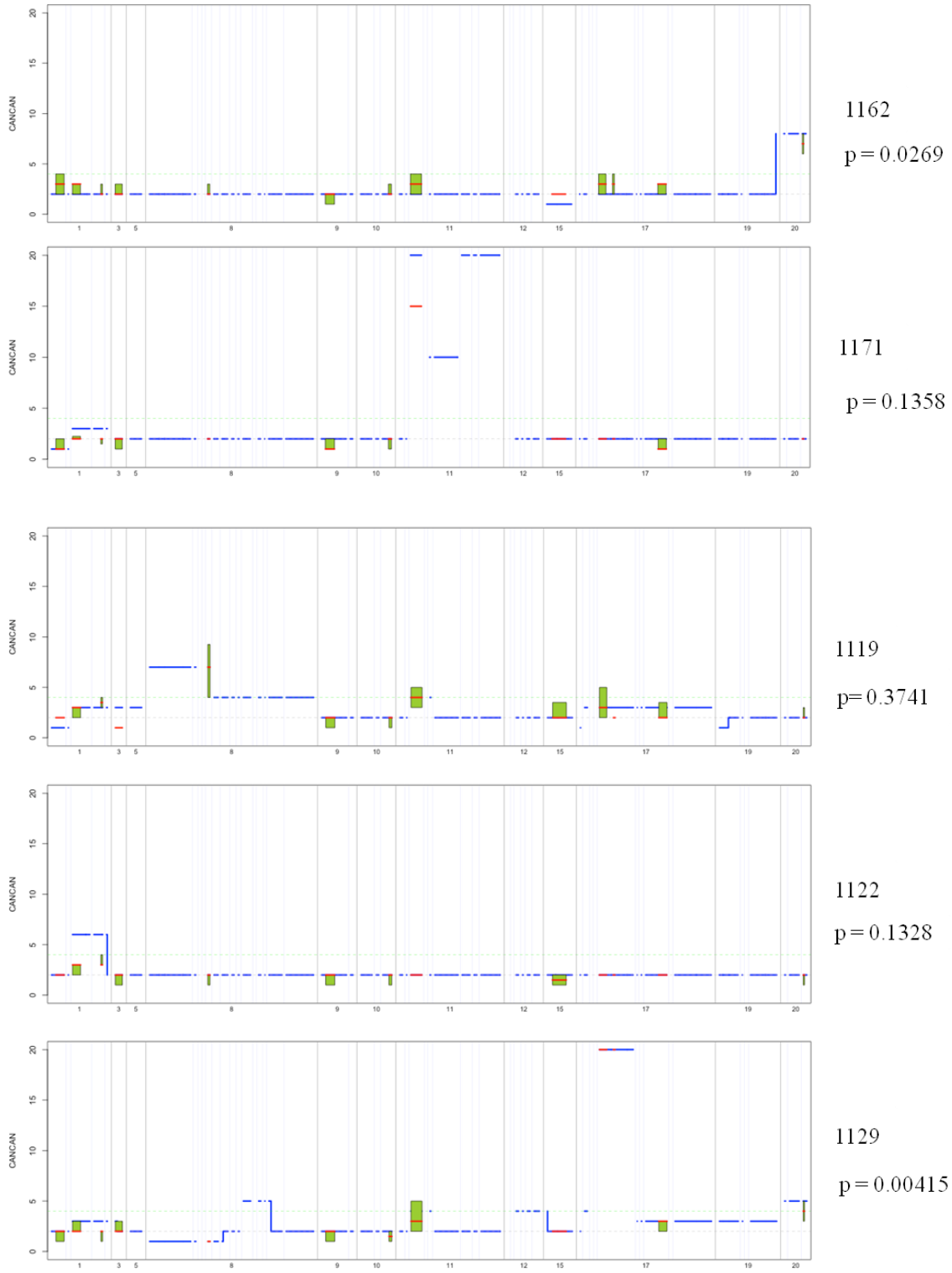
ST3: Overview of the number of counted nuclei for each probe for all patient samples. Numbers below or equal to 30 are marked green while indeterminate samples are marked blue.

Osl2_#	1116	1117	1119	1122_5	1122_7	1124	1129	1138	1162	1171	1210	1240	X
p1	50	52	10	40	50	31	54	46	51	61	47	56	nd
p2	84	65	79	49	58	55	73	22	41	44	41	52	nd
p3	64	92	50	54	49	62	64	52	36	35	45	43	nd
p4	40	55	19	42	43	37	64	19	62	46	68	48	38
p5	28	63	50	62	62	52	75	17	45	58	41	50	22
p6	49	46	24	48	33	65	42	43	48	44	42	40	43
p7	nd	79	25	64	66	79	51	34	56	nd	48	51	nd
p8	52	64	64	39	44	35	55	51	57	48	92	44	nd
p9	94	nd	25	40	76	nd	56	40	64	65	45	55	nd
p10	39	67	69	48	64	48	66	48	67	57	41	50	nd
p11	54	59	51	66	64	51	57	31	63	57	40	47	22
p12	55	57	31	55	58	67	52	22	57	46	84	50	nd
p13	80	75	119	60	50	48	58	37	43	57	48	51	nd
p14	44	113	79	95	47	54	41	46	61	44	51	43	nd
p15	95	nd	14	41	64	41	41	47	71	46	66	43	nd
p16	62	38	23	54	48	46	47	43	69	22	41	45	nd
p17	87	87	116	58	64	53	45	50	41	53	53	45	nd

ST4: List of standard deviations for probes p1-p17 for the patient samples.

Pat.ID	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17
Osl2_1117	1.3	1.3	1.5	1.4	1.6	0.5	0.7	0.5	NA	1.2	1.2	0.5	1.6	0.9	NA	0.8	0.9
Osl2_1119	0.4	1.2	1.1	0	0.8	3,0	1.3	0.8	0.7	1.7	1.9	0.9	1.2	1.9	0.4	1,0	1,0
Osl2_1122_5	0.5	1,0	0.8	0.5	0.4	0.5	0.4	0.5	0.5	0.4	0.4	0.4	0.5	0.4	0.4	0.5	0.5
Osl2_1122_7	0.5	0.7	0.7	0.5	0.5	0.2	0.5	0.5	0.5	0.5	0.4	0.5	0.5	0.5	0.4	0.8	0.5
Osl2_1124	0.7	0.9	1.1	0.7	0.9	0.4	1.1	1.2	NA	4.4	4.4	0.7	1,0	0	0	1.2	1.1
Osl2_1129	0.7	0.9	0.8	0.8	0.7	0.3	1.3	0.6	0.5	1.1	1.4	0.9	0.6	0	0	0.9	1.5
Osl2_1138	0.5	0.7	0.7	0.7	0.8	0	0.8	0.5	0.6	0.5	0.4	0.5	0.6	0.4	0.5	0.5	0.5
Osl2_1162	1.5	0.9	0.9	1.1	0.9	0.9	0.7	0.6	0.9	0.8	0.9	0.8	0.7	1.1	1,0	1.2	2
Osl2_1116	1,0	0.9	0.9	1.1	1.1	0.3	NA	0.6	1,0	1.3	1.2	1.2	0.6	0.7	0.4	1.9	1.2
Osl2_1171	0.5	0.7	0.7	0.6	0.5	0.4	NA	0.5	0.5	0	NA	0.5	0.4	0.9	0.4	0.5	0.7
Osl2_1210	0.7	0.4	0.6	0.5	0.3	2.5	2,0	1.1	0.4	0	4.6	0.9	0.4	0.4	0.7	0.7	1.1
Osl2_1240	1,0	0.8	1.4	1.2	1.1	1.6	2,0	0.8	1,0	1,0	0.9	1.1	1,0	0	0	1.4	0.8

1
2
2.1
NA



SF3: Alignment of CN by FISH against corresponding targeted sequencing by NGS in remaing 5 patient samples identified by numbers on the right hand side of the plots. CN by FISH is shown by a green box with median CN marked as a red line while the top and bottom represent the 75th and 25th percentile respectively. Mean CN estimated by NGS is illustrated by a blue line. Chromosome number is shown on the x-axis, CN is shown on the y-axis. p-value included for each plot.

ST5: Median CN by FISH and estimated CN by NGS for probes p1-p17 for 9 patient samples.

1117	FISH median	NGS CN		1119	FISH median	NGS CN		1122	FISH median	NGS CN
p1	4	6		**p1**	2	1		**p1**	2	2
p2	3	6		**p2**	3	3		**p2**	3	6
p3	5	6		**p3**	3.5	3		**p3**	3	6
p4	4	6		**p4**	1	3		**p4**	2	2
p5	3	NA		**p5**	2	NA		**p5**	2	NA
p6	2	2		**p6**	7	7		**p6**	2	2
p7	2	NA		**p7**	3	NA		**p7**	2	NA
p8	2	2		**p8**	2	2		**p8**	2	2
p9	NA	4		**p9**	2	2		**p9**	2	2
p10	3	NA		**p10**	4	NA		**p10**	2	NA
p11	3	4		**p11**	4	4		**p11**	2	2
p12	2	NA		**p12**	2	NA		**p12**	2	NA
p13	5	7		**p13**	2	2		**p13**	1.5	2
p14	2	2		**p14**	3	3		**p14**	2	2
p15	NA	2		**p15**	2	3		**p15**	2	2
p16	2	2		**p16**	2	3		**p16**	2	2
p17	2	4		**p17**	2	2		**p17**	2	2
1124	**FISH median**	**NGS CN**		**1129**	**FISH median**	**NGS CN**		**1138**	**FISH median**	**NGS CN**
p1	2	2		**p1**	2	2		**p1**	2	5
p2	2	2		**p2**	2	3		**p2**	2	5
p3	3	2		**p3**	2	3		**p3**	2	5
p4	2	2		**p4**	2	2.1		**p4**	2	5
p5	3	NA		**p5**	2	NA		**p5**	2	NA
p6	1	1		**p6**	1	1		**p6**	20	31
p7	3	NA		**p7**	3	NA		**p7**	2	NA
p8	2	2.2		**p8**	2	2		**p8**	2	4
p9	2	2		**p9**	1,5	2		**p9**	2	4
p10	10	NA		**p10**	4	NA		**p10**	2	NA
p11	9	9		**p11**	3	4		**p11**	2	4
p12	2	NA		**p12**	2,5	NA		**p12**	2	NA
p13	3	2		**p13**	2	2		**p13**	2	4
p14	20	24		**p14**	20	25		**p14**	2	4
p15	20	24		**p15**	20	25		**p15**	2	4
p16	4	3		**p16**	3	3		**p16**	2	4
p17	4	3		**p17**	4	5		**p17**	2	5

1162	FISH median	NGS CN		1171	FISH median	NGS CN		1210	FISH median	NGS CN
p1	3	2		**p1**	1	1		**p1**	2	4
p2	3	2		**p2**	2	3		**p2**	2	4
p3	2	2		**p3**	2	3		**p3**	2	4
p4	2	2		**p4**	2	2		**p4**	2	4
p5	2	NA		**p5**	1	NA		**p5**	2	NA
p6	2	2		**p6**	2	2		**p6**	15	14
p7	2	NA		**p7**	NA	NA		**p7**	7	NA
p8	2	2		**p8**	1	2		**p8**	3	4
p9	2	2		**p9**	2	2		**p9**	2	4
p10	2	NA		**p10**	15	NA		**p10**	15	NA
p11	3	2		**p11**	15	22		**p11**	9	21
p12	3	NA		**p12**	2	NA		**p12**	2	NA
p13	2	1		**p13**	2	2		**p13**	2	4
p14	3	2		**p14**	2	2		**p14**	2	4
p15	3	2		**p15**	2	2		**p15**	3	4
p16	3	2		**p16**	1	2		**p16**	2	4
p17	7	8		**p17**	2	2		**p17**	4	4