Identification and validation of novel amplification target genes in human breast cancer

Inauguraldissertation

zur Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Christian Ruiz
aus Spanien

Basel 2006

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von:

Prof. Dr. Markus A. Rüegg, Prof. Dr. Guido Sauter und Prof. Dr. Ruth Chiquet-Ehrismann

Basel, den 19. Dezember 2006

Prof. Dr. Hans-Peter Hauri
Dekan

Table of contents

Acknowledgements

I am deeply grateful to Prof. Guido Sauter for giving me the opportunity to perform my thesis in his research group and for giving me such an important and exciting research topic.

I am also very grateful to PD Dr. Ronald Simon for supervising me during my thesis and for his unique way of motivation.

I would also like to thank Prof. Luigi Terracciano for hosting me during the last part of my thesis.

Further, I would like to thank Prof. Ruth Chiquet-Ehrismann for being member of my thesis committee, for supporting me during my thesis and for helpful scientific suggestions.

I would also like to express special thanks to PD Dr. Gertraud Orend for the first guidance in the world of science and for being member of my thesis committee.

I am also deeply grateful to Prof. Markus Rüegg for being member of my thesis committee.

Further, I would like to thank Dr. Michel Bihl for scientific discussions and suggestions.

I am very grateful to my colleague Dr. Martin Oeggerli for sharing office, laboratories and scientific knowledge with me, for helping with the design and the layout of this document, and for enjoying so many hundreds of coffees together.

Warm thanks to the members of the laboratories, for technical help and for sharing their time with me: Alex Rufle, Barbara Stalder, Sandra Schneider, Hedvika Novotny and Rosi Chaffard.

Special thanks to my parents for unlimited care and generous support.

Last but not least, I would like to thank Rebecca Schirinzi for supporting me since 1995.

Abbreviations

DCIS	ductal carcinoma in situ
LCIS	lobular carcinoma in situ
IDC	infiltrating ductal carcinoma
ILC	infiltrating lobular carcinoma
ER	estrogen receptor 1 protein
ESR1	estrogen receptor 1 gene
PR	progesterone receptor
HER-2	erythroblastic leukaemia viral oncogene homolog 2 (ERBB2)
CGH	comparative genomic hybridization
aCGH	array based comparative genomic hybridization
FISH	fluorescence in-situ hybridization
SNP	single nucleotide polymorphism
NFIB	nuclear factor 1 B
IHC	immunohistochemistry
cdk	cyclin-dependent kinase
CAM	cell-cell adhesion molecule
MMP	matrix metalloproteinase
DM	double minute
HSR	homogeneously staining region
H&E	haematoxylin and eosin
BAC	bacterial artificial chromosome
RNAi	RNA interference
FCS	fetal calf serum

Summary

Gene amplification is a major mechanism for overexpression of potential oncogenes in cancer. Several amplifications have already been described in breast cancer. In order to find new amplified regions, we screened 30 human breast tumors for gene amplifications using the Affymetrix SNP 10k 2.0 microarrays. For this purpose, we developed a new analysis procedure leading to an increased signal-to-noise ratio which allowed us the discovery of new small single gene amplicons (< 1 Mb). Two of them, the ESR1 gene on 6q25 and the NFIB gene on 9p24, were further investigated. Fluorescence in-situ hybridization of these two genes was performed on a TMA comprising more than 2200 breast cancer samples.

NFIB amplification was found in 5% of all breast cancers analyzed, but with an increased amplification rate in medullary carcinoma (19%). NFIB amplified breast cancers showed a higher Ki67 proliferation index. Functional analysis with RNA interference of the NFIB gene in three tumor cell lines suggested a proliferation supporting role of the NFIB protein in breast cancer.

The ESR1 gene was amplified in 21% of the breast cancer samples analyzed. The ESR1 amplified patients defined a subgroup of ER positive breast cancer patients with prolonged survival, suggesting that patients with ESR1 gene amplification optimally benefit from hormonal therapy. Since this amplification was also found in histological benign and precancerous breast lesion, we suggest that ESR1 gene amplification is an early mechanism in breast cancer development.

Furthermore, the use of potentially heterogeneous markers, like the Ki67 proliferation index on a breast cancer TMA was investigated. All expected associations between Ki67 and other previously analyzed molecular markers could be reproduced with high statistical significance using a breast cancer TMA containing only one tissue sample per tumor. This leads to the conclusion that associations with cell proliferation can be reliably analyzed in a TMA format.

1. Introduction

1.1. Importance of this work

In the last few years, the identification of gene amplifications has become of special interest since an association of a given amplification with pharmacological effects of drug treatments has been detected. New anti-cancer drugs do not act non-specifically to inhibit dividing cells, but are directed against particular molecular targets. Frequently, a genomic aberration underlies the pro-tumorigenic activity of these drug targets and members of this new class of anti-cancer agents act specifically against the proteins encoded by the affected genes. The first successful agent was Glivec/Gleevec (compound: Imatinib). It is a specific kinase inhibitor targeting the bcr-abl fusion protein, the product of a fusion gene created by a genomic translocation, also known as Philadelphia chromosome. Herceptin (compound: Trastuzumab) was the next agent of this new class of drugs. This medicament is based on a monoclonal antibody against the protein HER-2 and it is only effective if the patient's breast tumor carries a HER-2 gene amplification. Similarly, but still disputed, is the effectiveness of Tarceva (compound: Erlotinib) and Iressa (compound: Gefinitib). Both are suggested to act in non-small cell lung cancer if an EGFR gene amplification is present. These findings, that a gene amplification is able to predict response to an anti-cancer agent and is therefore able to improve decision making on therapy and patient's outcome, encouraged us to focus our work on the discovery of new gene amplifications that in turn could lead to the identification of novel targets for highly specific drug therapies.

1.2. Breast Cancer

1.2.1. Medical view

Each year about 10 million new cases of invasive cancer arise world-wide. More than 10% of these cases arise in the breast. This makes female breast cancer the second most common site of malignant neoplasms after the lung (Parkin, 2001) (Figure 1).

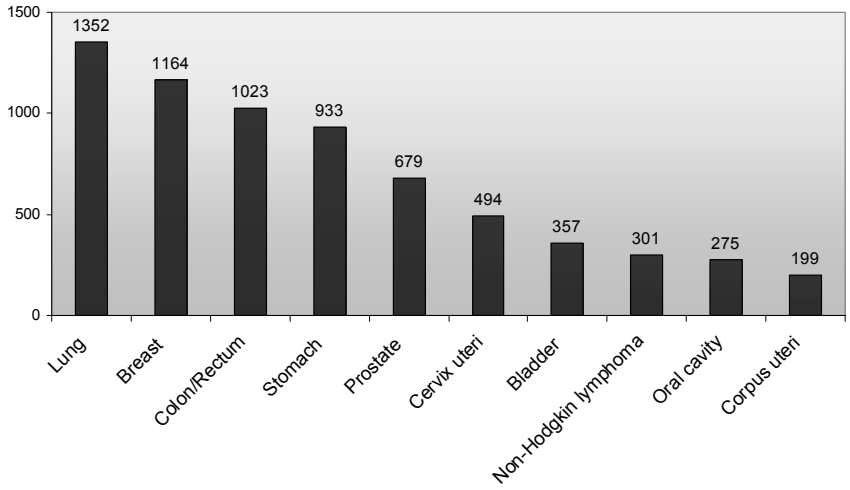


Figure 1. Incidence of new cancer cases (in thousands). Adapted from (Parkin et al., 2005).

Since most of the breast cancer cases affect only women (1% of the cases arise in men), breast cancer is the most common cancer in females. It accounts for about 22% of all new cancers in women (Figure 2). In developed countries, this proportion rises to 27% (Vainio & Bianchini, 2002). The difference in the incidence between developing and developed countries is said to be due to the earlier detection in developed countries (higher screening possibilities) as well as the improvement in prognosis, since more and more therapies have become available.

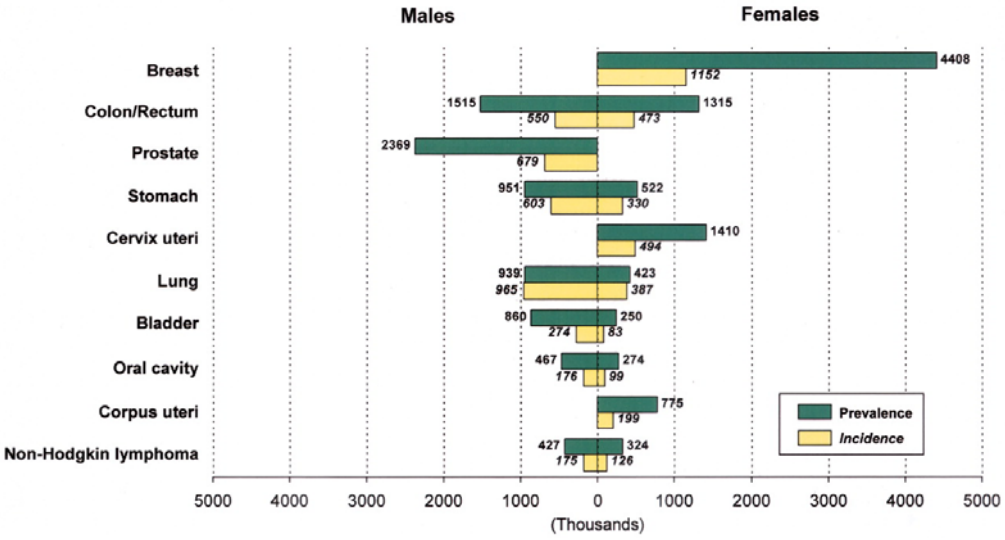


Figure 2. Estimated numbers of new cancer cases (incidence) and prevalent cases in 2002. Adapted from (Parkin et al., 2005).

Nevertheless, there are also environmental factors in the aetiology that contribute to a higher incidence: Low parity and late age at first pregnancy are consistently associated with an increased risk for breast cancer (Vainio & Bianchini, 2002). Increased life expectancy contributes also to a higher risk; as for most epithelial tumors, the risk increases steadily with the age. Although the incidence rate is rising, the mortality rate is decreasing in several western countries. This is due to improvements in diagnosis and therapy (Vainio & Bianchini, 2002), leading to a high prevalence of breast cancer patients in the population (Figure 2). In the United States, survivors of breast cancer constitute 1.5% of the female population (Hewitt et al., 1999).

1.2.1.1. Classification of breast tumors

Breast cancer is a heterogeneous disease in terms of clinical course and microscopic pathology. Although they can start in any tissue of the breast, most of them begin in the ducts (the milk-passages that connect the lobules to the nipple) or in the cells of the lobules (the milk-producing glands). The World Health Organization (WHO) classifies breast cancer due to its histological appearance, into noninvasive and invasive breast cancers (Coleman & Tsongalis, 2001; Fehr et al., 2006; Vainio & Bianchini, 2002):

Noninvasive breast cancers:

There are two main types of noninvasive breast cancers: the ductal carcinoma in situ (DCIS) and the lobular neoplasia (also called lobular carcinoma in situ, LCIS). The cancer cells of these forms are either located inside the ducts (DCIS) or inside the lobules (LCIS). Both types are so-called "in-situ" because they do not invade the surrounding fat tissue, nor spread through other organs in the body. DCIS is classified as precancerous disease, since 30-40% of the DCIS cases would pass into an invasive breast cancer, if they would not be treated. In contrast to DCIS, LCIS is not classified precancerous, since it will only rarely advance to an invasive cancer. But patients with this condition will still have a higher risk of developing an invasive breast cancer, either in the same breast or in the opposite one.

Invasive breast cancers:

Most of the invasive breast cancers (about 80%) are infiltrating (or invasive) ductal carcinomas (IDCs). These tumors start in the duct of the breast, brake through the wall of the duct and invade the surrounding fat tissue, from where they can spread through the lymphatic system or bloodstream. The other main type of invasive breast cancers (about 10-15%) is the infiltrating (or invasive) lobular carcinoma (ILC). These cancers begin in the lobules of the breast and act then similarly to the IDCs. Less frequent than the ILCs are the medullary carcinomas. The proportion described in the literature ranges between 5% and 10%. This special type of invasive breast cancer was named in this way for its similarity in color to brain tissue (medulla). Nevertheless, they are difficult to distinguish from IDC and are therefore often treated the same way, although the prognosis for medullary breast carcinomas is better than for other types of invasive breast cancer (Rubens et al., 1990). Other rare types of invasive breast cancer are inflammatory breast cancer and tubular carcinoma (each about 2%), and even rarer (less than 1%) are mucinous carcinomas, paget disease of the nipple and phyllodes tumor.

1.2.2. Molecular biological view

The human body can be seen as a micro-ecosystem consisting of billions of cells, each with a defined role behaving in a socially responsible manner. In such a system it is not surprising that cell growth is tightly controlled (Alberts et al., 2000). A molecular disturbance, i.e. a heritable change like a mutation, can give to the single cell carrying this mutation the ability to circumvent the strict cell growth control mechanisms. This can allow the cell to divide more quickly than its adjacent cells and become the founder of a new cell mass, called a tumor. As long as these tumor cells do not have the ability to invade the surrounding tissue, they are called benign. The conversion of such a cell into a neoplastic (also invasive) state is a multi-step process and requires the accumulation of several mutations (Coleman & Tsongalis, 2001). Indeed, up to 10 mutations are needed to become malignant and to have the ability to invade the surrounding tissue or to metastasize using the bloodstream or the lymphatic system as a carrier and to colonize other organs of the human body (Alberts et al., 2000; Coleman & Tsongalis, 2001). These tumor-promoting mutations typically affect three major classes of genes: proto-oncogenes, tumor suppressor genes and stability genes (caretakers) (Vogelstein & Kinzler, 2004). The latter act in a different way than oncogenes or tumor suppressor genes, because they do not influence cell growth or cell proliferation directly. The normal function of stability genes is to keep genetic alterations to a minimum (Vogelstein & Kinzler, 2004). They are either involved in DNA repair mechanisms (mismatch repair, nucleotide-excision repair, base-excision repair) or in the control of processes, which involve larger portions of chromosomes, like mitotic recombination and chromosomal segregation. When these genes cannot function in a proper way (because of a mutation or of a loss), the cells become genetically instable and genetic alterations in other genes can occur at a higher rate (Cahill et al., 1999; Vogelstein & Kinzler, 2004). Genome instability is a process which may develop after tumor initiation, i.e. during tumor progression, for example by a malfunction of p53, the so-called "guardian of the genome" (Alberts et al., 2000). Nevertheless, the progressive destabilization of the genome of a cancer cell is a common event in all cancers and is a prerequisite for the tumor cells to gain the abilities that are necessary for malignant growth. These so-called six hallmarks of cancer (Hanahan & Weinberg, 2000) are said to be the critical features that are

responsible for the phenotype that we recognize as cancer: self-sufficiency in growth signals, insensitivity to anti-growth signals, evasion of apoptosis, limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis. These properties may also apply to breast cancer:

Self-sufficiency in growth signals:

As mentioned before, cell proliferation is a process that is rigidly controlled. Growth signals are required to switch the cells from a quiescent state into a proliferative state (Hanahan & Weinberg, 2000). These growth signals are ligands of receptors that will induce the activation of specific signaling pathways, thus leading to activation of gene transcription. Growth factors, like the transforming growth factor alpha (TGF-α) and insulin-like growth factor-1 (IGF-1) stimulate benign breast growth. Uncontrolled expression of these signals can become important drivers of self-sufficient growth in breast cancer (Humphreys & Hennighausen, 2000; Schroeder & Lee, 1997; Sledge & Miller, 2003). Overexpression of receptors can lead the tumor cell to become hyperresponsive to ambient levels of growth factors that normally would not trigger cell proliferation (Hanahan & Weinberg, 2000). The estrogen receptor (ER), a steroid hormone receptor in the nucleus and the HER-2 receptor, a transmembrane receptor tyrosine kinase, are both prominent examples of overexpressed receptors in breast cancer. Both of them have become important therapy targets.

Insensitivity to antigrowth signals:

To maintain the cells in a quiescent state, several anti-proliferative signals are necessary. These so-called negative controls fail to function properly in cancer. Many of these negative feedback loops involve proteins, which are key players of the cell-cycle protein machinery. This machinery consists of cyclins, which form complexes with their specific cyclin-dependent kinases (cdks) leading to the activation of their kinase function. In breast cancer, the importance of the insensitivity to antigrowth signals, either by overexpression of cyclins and/or by downregulation of the cdk-inhibitors, have been shown in several studies: Overexpression of Cyclin D1 occurs at a relatively early stage (Heffelfinger et al., 2000) and increased levels of Cyclin E and its related cdk (Keyomarsi et al., 2002) or downregulation of the cdk-inhibitor p27 are associated with worse outcome (Tan et al., 1997; Thor et

al., 2000; Tsihlias et al., 1999). The loss of cdk inhibitors may also be caused by loss of TGF-β signaling, which can be an early event in breast cancer (Donovan & Slingerland, 2000), either by loss of TGF-β production or through mutational inactivation of the TGF-β receptors (Baxter et al., 2002; Chen et al., 2006; Fynan & Reiss, 1993).

Evasion of apoptosis:

The ability of tumor cells to expand in number is not only limited by the rate of proliferation, but also by the rate of programmed cell death (apoptosis), which antagonizes the proliferative effect (Hanahan & Weinberg, 2000). Therefore, in order to increase its cell number, the tumor has to acquire resistance towards apoptosis. This can be achieved in different ways, for example by downregulation of bcl2, an anti-apoptotic gene with lower expression in the majority of the breast cancers or by mutation of p53, which in its normal state is able to induce apoptosis. Mutation of p53 occurs in 25-45% of primary breast cancers (Osborne et al., 1991).

Limitless replicative potential:

Mammalian cells carry an intrinsic program that limits the number of multiplications a cell is able to perform. This program is effectuated by the progressive loss of the telomeres, each time a cell divides. With each cell cycle, the telomeres become progressively shorter, resulting in senescence and cell death (Alberts et al., 2000; Coleman & Tsongalis, 2001; Hanahan & Weinberg, 2000). For a tumor cell to become immortal, it is necessary to disrupt this program. This can be achieved by telomere maintenance due to increased levels of telomerase, an enzyme that adds specific DNA sequence repeats to the 3' ends of the telomeres of the DNA strands. The human telomerase catalytic subunit gene (hTERT) has been found to be activated in over 90% of the breast tumors (Herbert et al., 2001), including pre-invasive tumors (Mueller et al., 2002; Shpitz et al., 1999). But no expression was observed in the adjacent normal tissue (Herbert et al., 2001).

Sustained angiogenesis:

The formation of new blood vessels is crucial for nourishing the tumor, especially if invasion and metastasis will occur (Folkman, 1971). In healthy adults, the normal vasculature is quiescent, with

each endothelial cell dividing once every 10 years (exceptions: wound healing, endometrial proliferation, postlactational mammary gland involution and pregnancy) (Sledge & Miller, 2003). In order to stimulate the vascular endothelial cells, i.e. to induce angiogenesis, the tumor cells have to activate the switch by changing the balance of angiogenic inducers and countervailing inhibitors (Hanahan & Weinberg, 2000). The most prominent angiogenic inducer is VEGF (vascular endothelial growth factor). In breast cancer, VEGF has been to be markedly upregulated in comparison to the surrounding normal tissue (Kawai et al., 2002).

Tissue invasion and metastasis:

A hallmark of the malignant phenotype of a cancer is the ability to invade through the basement membrane, i.e. the transition from a non-invasive to an invasive phenotype. Metastases, which are the cause of about 90% of human cancer deaths, can be seen as an extension of local invasion. To gain this capability, tumor cells have to alter several classes of proteins, which are responsible either to couple the cells to their environment (cell-cell adhesion molecules CAMs, integrins), or to degrade the basement membrane and the surrounding stroma, like matrix metalloproteinases (MMPs). MMP-2 and MMP-9 are two prominent examples, whose expression in breast cancer was found to be associated with stage and grade (Kossakowska et al., 1996; Zucker et al., 1993).

Finally, it has to be considered that the sequence how these hallmarks are acquired, i.e. how a tumor becomes physiologically a "successful" cancer, varies from patient to patient (Sledge & Miller, 2003): a mutation in a certain oncogene or in a tumor suppressor gene can occur early in some tumors, and late in others.

1.2.2.1. Biomarkers in breast cancer

Several of the above mentioned oncogenes and tumor suppressor genes are used as molecular markers, also called biomarkers. By analyzing biological tumor features in breast cancer, discrimination between diagnostic, prognostic and predictive markers is not always unproblematic. Diagnostic markers are routinely used by pathologists to facilitate the diagnosis. The most prominent ones are the cytokeratins and E-cadherin. Cytokeratins are also used as support for differentiation between benign and malignant breast lesions (Moriya et al., 2006). E-cadherin is a calcium regulated adhesion molecule, which is a central component of cell-cell adhesion junctions. Under normal circumstances, it is expressed in most epithelial tissues (Takeichi, 1990). E-cadherin is suggested to act as a tumor suppressor protein, since its loss can cause dedifferentiation and invasiveness in human carcinomas (Frixen et al., 1991; Larue et al., 1994). In the last few years, loss of E-cadherin has been consistently observed at sites of epithelial-mesenchymal transitions (EMT) during development and cancer (reviewed in (Kang & Massague, 2004)). In breast cancer diagnostics, E-cadherin is used as a marker to distinguish lobular from ductal carcinomas since its expression is almost undetectable in lobular tumors by immunohistochemistry (Moriya et al., 2006).

Many genes have been classified as prognostic factors in breast cancer. These factors provide information on outcome, independent of the chosen adjuvant therapy. The most prominent ones are summarized in Table 1. Most of them have already been mentioned when introducing the hallmarks of cancer: HER-2, ER, PR and EGFR are all receptors and can be categorized in to the "self-sufficiency in growth signals"; Cyclin A, B, D, E and p27 are cell-cycle components belonging to "insensitivity to anti-growth signals"; bcl-2 and survivin are involved in "evasion of apoptosis", and the telomerase hTERT has been introduced as an enzyme necessary for a tumor cell to get "limitless replicative potential".

Marker	Positive breast tumors (%)	Association with:	Reference
HER2	15-30	worse prognosis	Tsutsui et al.; Surgery, 2003
ER	70	good prognosis	Torhorst et al.; Am J Pathol., 2001
PR	60	good prognosis	Torhorst et al.; Am J Pathol., 2001
EGFR	36	worse post-relapse survival	Tsutsui et al.; Clin Cancer Res, 2002
Cyclin A	8 (0-51)	worse prognosis in node-negative patients	Kuhling et al.; 2003, J Pathol.
Cyclin B	5 (0-75)	worse prognosis in node-negative patients	Kuhling et al.; 2003, J Pathol.
Cyclin D	42	no real association, controversial results	Gradishar; Breast Cancer Res Treat., 2005
Cyclin E	10 (0-90)	worse prognosis	Keyomarsi et al.; NEJM, 2002
p27	80	good prognosis	Lloyd et al.; J. Pathol, 1999
p53	20-40	worse prognosis	Borresen-Dale et al.; Human Mutat, 2003
bcl-2	30-40	good prognosis	Chang J et al; Cancer, 2003
Survivin	80	cytoplasmic: worse prognosis	Sohn et al; Biomed Pharmacother, 2006
hTERT	59	short survival	Poremba et al; J Pathol, 2002

Table 1: Prognostic biomarkers in breast cancer. Numbers in parenthesis represent the range. All these markers have been analyzed by immunohistochemistry. Modified from (Coradini & Daidone, 2004).

The most informative markers for the patients are the predictive ones. Unfortunately, these markers are often confused with the above described prognostic markers. In contrast to prognostic markers, predictive markers are used to select responsiveness or resistance to a specific treatment (Duffy, 2005). Usually, predictive markers are prognostic too, but not vice versa. Currently, the American Society of Clinical Oncology recommends routine testing of three predictive markers for decision making on therapy in breast cancer: the estrogen receptor ER, the progesterone receptor PR and the human epidermal growth factor receptor-2 (HER-2) (Bast et al., 2001). Testing of ER and PR positivity in breast cancer is used for predicting response to hormone therapy. Initially, this therapy included oophorectomy (ovarian ablation). Current hormonal therapy is called "anti-estrogen therapy" and is more focused on the estrogen receptor, since only 5% of breast cancers are ER negative and PR positive, and only 10% of that subgroup will respond to hormonal therapy. Anti-ER treatment strategies include blocking by selective modulators (e.g., Tamoxifen), destabilization and degradation by selective downregulators (e.g., Fulvestrant) and disruption of estrogen synthesis (e.g., Anastozole). At the moment, it is unknown why only 70% of the ER+/PR+ and only 33% of the ER+/PR- do respond to these therapies.

The third predictive marker routinely used in breast cancer involves the HER-2 gene. It belongs to the EGF receptor family of receptor tyrosine kinases. Other members of this family are the HER-1 (EGFR), HER-3 and HER-4 receptors (Duffy, 2005). Ligand binding of these transmembrane receptors promotes dimerization leading to an activation and thus to an increased tyrosine kinase

activity (Burgess et al., 2003). In contrast to these members, no directly binding ligand has been identified for the HER-2 protein. It is suggested that the HER-2 receptor has a high level of constitutive activity and thus it is willing to form heterodimers with ligand activated forms of other HER receptors (Burgess et al., 2003). In human breast cancers, amplification and overexpression of the HER-2 gene are found in 15-30% of primary invasive tumors (Duffy, 2005), causing a very aggressive form of breast cancer. Several monoclonal antibodies against the HER-2 ectodomain were developed; one of these, also known as 4D5, was later modified for administration to patients and termed trastuzumab, the active component of the medicament Herceptin (Carter et al., 1992). The food and drug administration FDA has approved Herceptin for treating metastatic HER-2 positive breast cancers. Administration of Herceptin together with chemotherapy led to the shrinking of tumors of nearly half the women (45%) treated (Slamon et al., 2001). Currently, clinical trials with women carrying non-metastatic HER-2 positive breast cancer are in progress.

1.2.2.2. Molecular expression patterns in breast cancer

It was obvious that the two most prominent markers, the ER and the HER-2, would influence the gene expression patterns of breast cancer cells. Therefore, genetic classification of breast cancers based on RNA expression arrays has been performed several times. However, only the study of Perou and Sorlie (Perou et al., 2000; Sorlie et al., 2001) was reproduced on different data sets resulting in the same classification pattern (reviewed in (Sorlie, 2004)). Their genetic subtyping was based on the variation in expression patterns observed in 115 breast cancer samples. They selected an intrinsic set of 415 genes, followed by hierarchical clustering of all the samples based on this gene set. They distinguished two main classes of tumors, one with the characteristics of basal cells, the other of luminal cells. The so-called basal tumors expressed CK5 and CK17, but were ER negative, whereas the luminal phenotype was based on the expression of CK8/18, CK19 and a cluster of transcription factors including the estrogen receptor. This two-class model was further refined and several subclasses were created. One of these subclasses, which branched close to the basal-like cases, was characterized by the overexpression of the HER-2 gene, suggesting these samples to be the ones with

17q21 amplification. Survival analyses of these groups revealed worst survival for the HER-2 and basal-like group and best survival for the luminal group. This was not surprising, since ER positive patients of this study had been treated with Tamoxifen. Later, the luminal group was further subdivided into two subgroups, the luminal A group with high ER expression and the luminal B group with low to moderate ER expression. Results from different DNA microarray platforms obtained on different breast cancer sets (e.g. Van't Veer set, West set) with variations in technology and statistical analysis, provided highly consistent classification results. Recently, Calza et al (Calza et al., 2006) applied a similar analysis procedure to a public available Swedish breast cancer set (Pawitan et al., 2005), leading to the same discrimination of these subgroups. The existence of these subclasses has been reinforced at the protein level; using immunohistochemistry on tissue microarrays, the same or similar subclasses of tumors have been observed (Callagy et al., 2003; Korsching et al., 2002; van de Rijn et al., 2002).

1.3. Gene amplification in cancer

There are three types of genetic alterations that can transform a cancer-critical gene into an oncogene: a deletion or a mutation in the coding sequence (e.g. EGFR), a chromosome rearrangement, like the creation of a fusion protein (e.g. Bcr-Abl fusion gene) and the amplification of a specific DNA sequence.

Amplifications are mutations that result in multiple copies of genes in amplified chromosomal regions (so-called amplicons). In contrast to other organisms (e.g. oogenesis fruitfly) (Spradling & Mahowald, 1980), amplification in mammals is an unscheduled process. It can be seen as a common genetic mechanism for upregulating gene expression in cancer (Schwab, 1999). But the number of overexpressed genes in the amplified chromosomal regions varies depending on the cancer tissue and also on the study (Myllykangas et al., 2006; Myllykangas & Knuutila, 2006): In prostate cancer cell lines, 19.3% of the amplified genes were found to be overexpressed (Wolf et al., 2004). In breast tumors the fraction ranges between 44% (Hyman et al., 2002) and 62% (Pollack et al., 2002). Especially, advanced cancers, which have lost p53-mediated maintenance of genomic integrity or other genes involved in the apoptotic disruption of damaged cells, are affected by these chromosomal rearrangements (Livingstone et al., 1992; Yin et al., 1992). Due to variable sizes of the different amplicons, it remains a challenge to identify the driving genes, which are amplified and overexpressed, and give growth advantage to the tumor cell. Mostly, the driving genes of the amplicons are well-known oncogenes: for example the protein kinase AKT2 (14q32) in ovarian cancer; the v-myc viral oncogene homolog 1 (MYCL1; 1p34) in small cell lung cancer and glioma; MYCN (2p24) in about 20% of neuroblastoma; HER-1 (EGFR, 7p12) in glioma and non-small cell lung cancer or HER-2 (17q21) in 15-30% of breast tumors (Futreal et al., 2004; Myllykangas & Knuutila, 2006; Schwab, 1999).

1.3.1. Mechanism of gene amplification

Although the role of gene amplification in cellular transformation of human cancers is indisputable, the precise mechanisms how amplifications develop have not been entirely determined.

There are two types of cytogenetic manifestation of gene amplification: extrachromosomal double minutes (DMs) and intrachromosomal homogenously staining regions (HSRs). Double minutes are small circular DNA fragments, lacking a centromere and a telomere (Barker, 1982). It has been suggested, that they replicate autonomously, since they also contain replication origins (Carroll et al., 1993; Carroll et al., 1987). But the most popular model for the development of gene amplification is the breakage-fusion-bridge (BFB) model. Initially, this model leads to an intrachromosomal amplicon, which can excise itself and become an extrachromosomal double minute (Coleman & Tsongalis, 2001). The BFB-model was first described in 1942 (McClintock, 1942) and has been remodeled by several investigators leading to the actual model presented in figure 3 (Bailey & Murnane, 2006). It is suggested that an intrachromosomal gene amplification is the product of several rounds of BFB cycles.

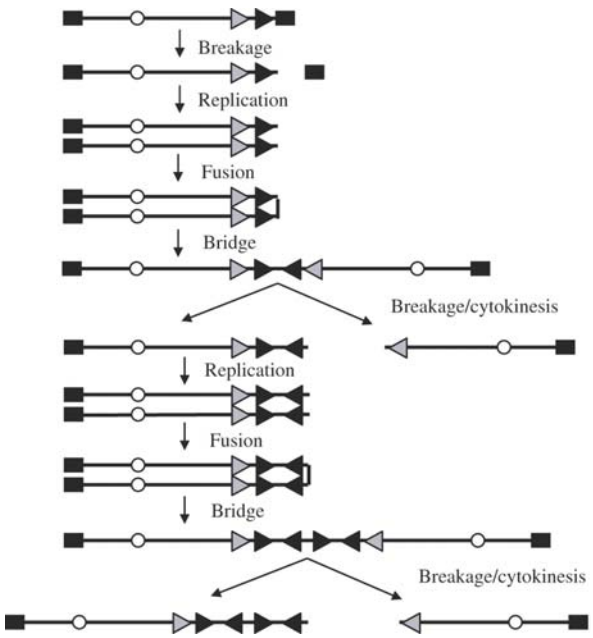


Figure 3. The bridge-fusion-model as actual model for the development of intrachromosomal gene amplification. Squares: telomeres, circles: centromeres, horizontal arrows: orientation of sequences.

The first step of the BFB model is the generation of a double-strand break (DSB). Next, the broken molecule is replicated and the sister chromatids fuse and form a bridge during anaphase. Asymmetrical breakage due to mechanical tension will lead to one daughter cell with a copy of the chromosome with an inverted repeat at the broken end, and one daughter cell owing a chromosome with a terminal deletion. Several rounds of BFB cycles would lead to the above mentioned HSR, which would be in line with the observation that HSRs are often organized as an inverted ladder and are often associated with a deletion from the amplicon towards a telomere (Bailey & Murnane, 2006; Haber & Debatisse, 2006; Narayanan et al., 2006). Nevertheless, it has to be considered that although mammalian cells have a robust non-homologous end joining machinery, the fusion of broken chromatids has never been directly demonstrated (Narayanan et al., 2006).

1.3.2. Detection of gene amplifications

The first method used for the detection of amplifications or other chromosomal changes was the use of conventional cytogenetics, in which chromosomal abnormalities are detected microscopically in cells arrested at methaphase. This procedure results in the typical black and white alternating banding pattern produced by the different staining methods (e.g. Q-banding, R-banding, G-banding). The use of this so-called chromosome banding for the detection of new amplifications in cancer is restricted by the limited cytogenetic information that is available from solid tumors, since it is difficult to culture neoplastic epithelial cells to bring them into metaphase, and by the low chromosome band-specific resolution of approximately 20 Mb. In 1992, the analysis of chromosomal changes in human malignancies was boosted by the development of the comparative genomic hybridization (CGH) (Kallioniemi et al., 1992). This molecular-cytogenetic method is based on the competitive hybridization of fluorescently labeled tumor DNA (e.g. Fluorescein, FITC) and normal DNA (e.g. Rhodamin-labeled) to normal human metaphase chromosome preparations. The fluorescence color ratio along the chromosome is used to evaluate regions of DNA loss or gains (Figure 4). However, the use of metaphase chromosomes limits detection of events involving small regions (less than 10 or 20 Mb) of the genome.

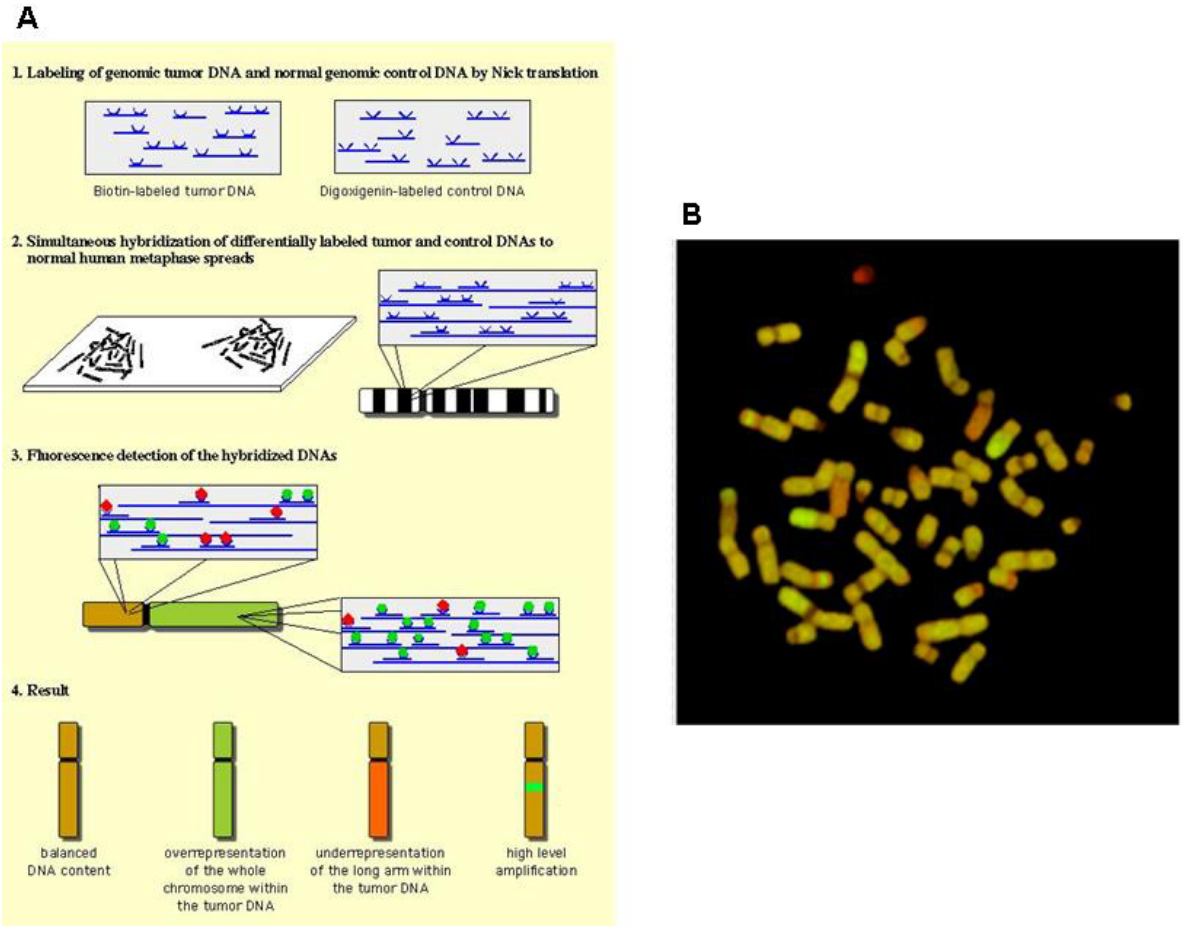


Figure 4. CGH: Comparative genomic hybridization. A) shows an overview of the steps involved in CGH (from Michael Baudis, www.progenetix.net). B) shows the result of a successful CGH (6p22 amplification).

Once an amplified region is identified, the suspected amplified genes need to be verified by Southern Blot or FISH (Fluorescence in Situ Hybridization). In both methods, the location of the gene or region of interest has to be known exactly since a specific probe has to be created. In 1997, Pinkel et al suggested that the hybridization of the fluorescently labeled probes to an array of mapped sequences instead of metaphase chromosomes could overcome the limitations of conventional CGH (Pinkel et al., 1998). This new technique was called array- or matrix CGH. The current sensitivity of this method is limited by the spacing of genomic clones used to construct the arrays (Mosse et al., 2005). Microarray platforms manufactured using BAC clones (Cai et al., 2002), cDNA clones (Pollack et al., 1999) and oligonucleotides (Carvalho et al., 2004) have been developed for aCGH. The theoretical resolution can be increased up to 200-300 kb, by choosing targets, which are positioned in the genome near to each other. Indeed, the minimal size of the amplicons detected has never been smaller than 1 or

2 Mb, due to the noise generated in the data when analyzing up to 40'000 probes at the same time (Myllykangas & Knuutila, 2006). Affymetrix, as leading manufacturer of commercial microarrays, designed a microarrays series for the analysis of single nucleotide polymorphisms (SNPs). Due to the distribution of the SNPs across the whole genome, their signals on these so-called SNP arrays can be used for the calculation of whole genome DNA copy numbers. As usual for Affymetrix, they do not apply competitive hybridization, since they use their single channel technology. Comparison of the acquired signals to a reference array or set of arrays must be done after the acquirement of the signals. As for aCGH technology, the resolution of this method depends on the number of probes available on the array. Affymetrix started with the 10K chip containing 11'555 SNPs, corresponding to an average distance between two SNPs of 210 kb, and promotes now its newest 500K chip with more than 500'000 SNPs and an average distance of 5.8 kb (Affymetrix, Product Sheet). Again, these numbers represent theoretical values, and the practical resolution that can be used for the final analysis, is lower and is dependent on the signal-to-noise ratio. This signal-to-noise ratio depends on several factors, such as heterogeneity of the starting material, noise generated in the different protocol steps and during the acquirement of the signals, and finally, the algorithm used for the analysis of the raw data.

1.4. Validation of candidate genes

The completion of the sequencing of the human genome has provided the research community with comprehensive information about the existence and location of nearly all human genes (Lander et al., 2001; Venter et al., 2001). Based on these data, several high-throughput analysis methods have become possible, such as cDNA microarrays for expression analysis or aCGH/SNP chips for detection of DNA copy number. Thus, analysis of tens of thousands of data points (i.e. genes) in one experiment has become possible, leading to the generation of several hundreds of candidate genes requiring only few experiments. For the validation of candidate genes, the reverse approach becomes appealing: the analysis of a candidate gene in hundreds or maybe thousands of tissues. This is particularly important in cancer classification, since there are several aberrant pathways that can lead to morphologically identical tumors. Thus, analysis of large numbers of tumors is necessary to obtain a full representation of the spectrum of relevance concerning the candidate gene. The validation of candidate genes with immunohistochemistry (IHC) for expression analysis or FISH for DNA copy number analysis can theoretically be performed on large tissue sections as usually used in pathology. However, the analysis of thousands of such slides for one candidate gene would lead to tremendous costs and exhaustion of valuable tissue archives. To overcome these shortcomings, the tissue microarray (TMA) technique was developed (Kononen et al., 1998).

1.4.1. Tissue microarrays

The tissue microarray technology allows the simultaneous analysis of up to 1'000 different tissue samples from routinely formalin-fixed paraffin blocks on a single microscope glass slide. The manufacture of these TMAs is done by taking minute tissue cylinders (typically 0.6mm in diameter) from different primary tumor blocks (the so-called donor blocks) and subsequently assembling them in an array-like format into an empty recipient block (Figure 5A). Regular microtomes can be used to cut sections from these TMA blocks to further perform in-situ analysis, like IHC and FISH (Figure 5B-E). Usually, only one spot with a diameter of 0.6mm of each tumor is sufficient for these kinds of

analyses. Several studies have shown that highly representative data about biomarkers in cancer can be obtained on the TMAs (Hoos et al., 2002; Nocito et al., 2001; Torhorst et al., 2001). However, it has been doubted if markers with heterogeneous expression patterns between different tumor areas, can reliably be determined using a single tissue core per tumor (Hoos et al., 2001). This may be a problem for the determination of the Ki67 labeling index (Ki67 LI), which shows considerable heterogeneity in breast tumors. The determination of the number of Ki67 expressing cells has become a standard procedure to assess the proliferative activity of tumor cells on tissue sections. The Ki67 LI marker has never been applied on a breast cancer TMA so far. But its successful application would not only help to determine the association of novel biomarkers with proliferation, but also prove the applicability of potential heterogeneous biomarkers on breast cancer TMAs.

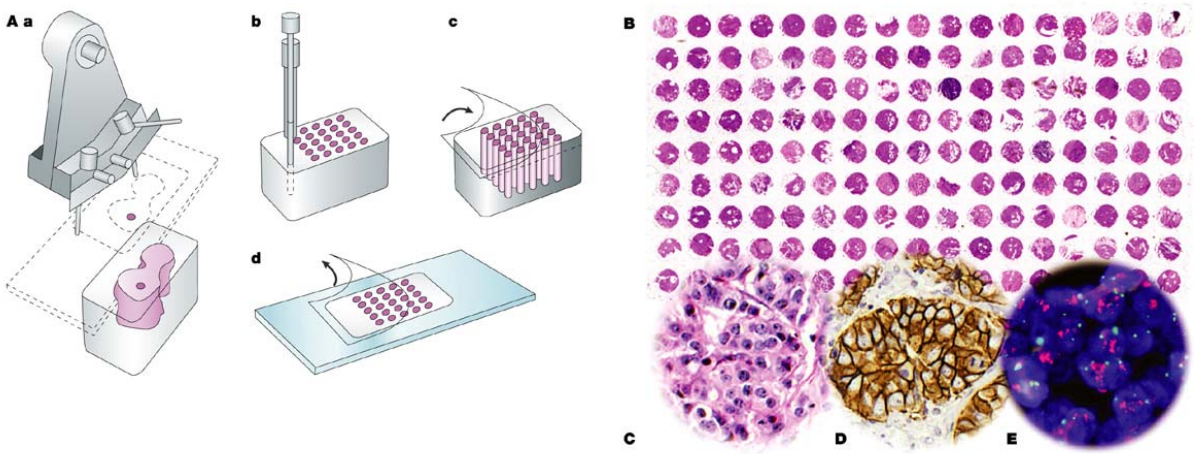


Figure 5. TMA manufacturing and applications. A) Cylindric tissue cores are removed from a donor block (a) and transferred to a recipient block (b). 4-8um sections are cut with a regular microtome using an adhesive tape (c) and placed on a microscopic slide (d). B) Overview of a haematoxylin-eosin (H&E) stained TMA section. C-E) Magnifications of spots from breast cancer TMA. C) H&E staining. D) IHC of Her-2. E) FISH of HER-2, showing red signals for HER-2 gene copies and green signals for the centromere 17. Adapted from (Sauter et al., 2003).

Nevertheless, the IHC analysis on TMAs is clearly limited by the availability of antibodies that are suited for formalin-fixed tissues and by the non-automated manual scoring of the intensity signal. The first limitation can be circumvented by the use of TMAs with frozen tissues (so-called frozen TMAs) (Schoenberg Fejzo & Slamon, 2001). However, the creation of frozen TMAs is limited by the availability of frozen tissues and by the fact that the blocks have to be kept frozen through and after

the arraying process. The second drawback is the limitation of the manual (and subjective) scoring of the intensity, i.e. the quantification of the signal. To overcome these limitations, the use of protein lysate arrays seemed to be a promising methodology worth developing.

1.4.2. Protein lysate microarrays

There are two major classes of protein lysate arrays currently under development: forward-phase protein arrays (FPAs) and reverse-phase protein arrays (RPAs). In FPAs, the analytes of interest are captured from the solution phase (lysate) by the spotted capture molecule, which is usually an antibody (Liotta et al., 2003; Pavlickova et al., 2004). One spot contains one type of immobilized antibody, thus an array can contain up to several hundreds of different antibodies. Since each array is incubated with one test sample, this approach is reminiscent of cDNA arrays, where one sample is analyzed for several "targets". More similar to the TMA approach, is the second major class of protein lysate arrays, the reverse-phase protein arrays. In this format, the test samples (lysates) are immobilized as spots to yield an array that consists of hundreds of different lysates. Each array is analyzed by one detection protein (i.e. antibody), leading to the measurement of one analyte end point across all samples. Due to the small volume necessary for a lysate spot (approximately 0.1-0.3 nl), it is possible to spot each sample in several replicates and in serial dilutions. The possibility of using a dual channel detection system (for example a Cy3-labeled antibody for the visualization of the protein of interest and Cy5 for a housekeeping protein) in combination with the availability of serial dilutions of the samples, allows the determination of the dynamic range for the detection of the specific antibody, and thus the quantification of the signal. The successful application of such a system would overcome the above mentioned limitations of traditional TMAs. It would allow the use of a large number of different antibodies. In combination with an automated analysis of the generated fluorescent signal, this would ideally not only give rise to a quantification of a specific protein amount, but also to a major increase in the sensitivity of detection.

Thus, it would become feasible to analyze the activation status of signal transduction pathways in cancer, by quantifying and comparing the phosphorylation status (using phosphorylation specific antibodies) of signal transduction molecules in hundreds or even thousands of tumors.

1.5. Aims of the thesis

This work has been divided into two major parts. This subdivision is also reflected in the structure of the results.

The aim of the first part (Part A) was to elucidate the influence of breast tumor tissue heterogeneity on marker evaluation using breast cancer TMAs. For this purpose, we decided to analyze the Ki67 LI marker as one of the most heterogeneous marker on the breast cancer prognosis TMA. The obtained data will be compared to ten different biomarkers (HER2, MDM2, Egfr, MYC, CCND1, ER, PR, p53, p16, bcl2), many of which have been previously related to proliferative activity in breast cancer. Its successful correlation would allow us to trust the impact of new biomarkers on proliferative activity using breast cancer TMAs.

The second part (Part B) comprised a screening of 30 breast cancer samples with the objective of finding new amplifications. For this screening, we decided to use the relatively new technology of the Affymetrix SNP 10k 2.0 microarrays. The aim of this study was to develop a method to detect unknown small-sized amplicons. Since there was no appropriate software available for this kind of analysis, we had to develop a new bioinformatics analysis procedure. Novel promising candidate gene amplifications would then be verified and validated using the FISH technique on breast cancer TMAs. The statistical analysis of these data would help us to elucidate the role of the newly detected amplifications in breast cancer. If the amplified candidate gene is unknown in breast cancer, functional analysis of this gene would help to determine its possible role in breast cancer development.

Finally, we intended to establish the relatively new technology of the reverse phase protein microarrays, in order to evaluate if its application would help us to circumvent the limitations of the tissue microarrays. This part of the work is shown in the Appendix.

2. Results

2.1. Part A: Tissue microarrays for comparing molecular features with proliferation activity in breast cancer (publication)

Published article: Int. J. Cancer: 118, 2190-2194 (2006)

2.2. Part B: Screening for new amplifications in breast cancer

2.2.1. Overview of the project

This study was based on 30 frozen breast cancer tissue samples from patients from the University Hospital in Basel. Extracted DNA was applied on the Affymetrix SNP 10k 2.0 arrays in order to detect new amplified chromosomal regions in breast cancer. In combination with the RNA expression profiles from the Affymetrix HG-U133A2 arrays, amplified genes, which also showed a high RNA expression, were chosen as candidate genes.

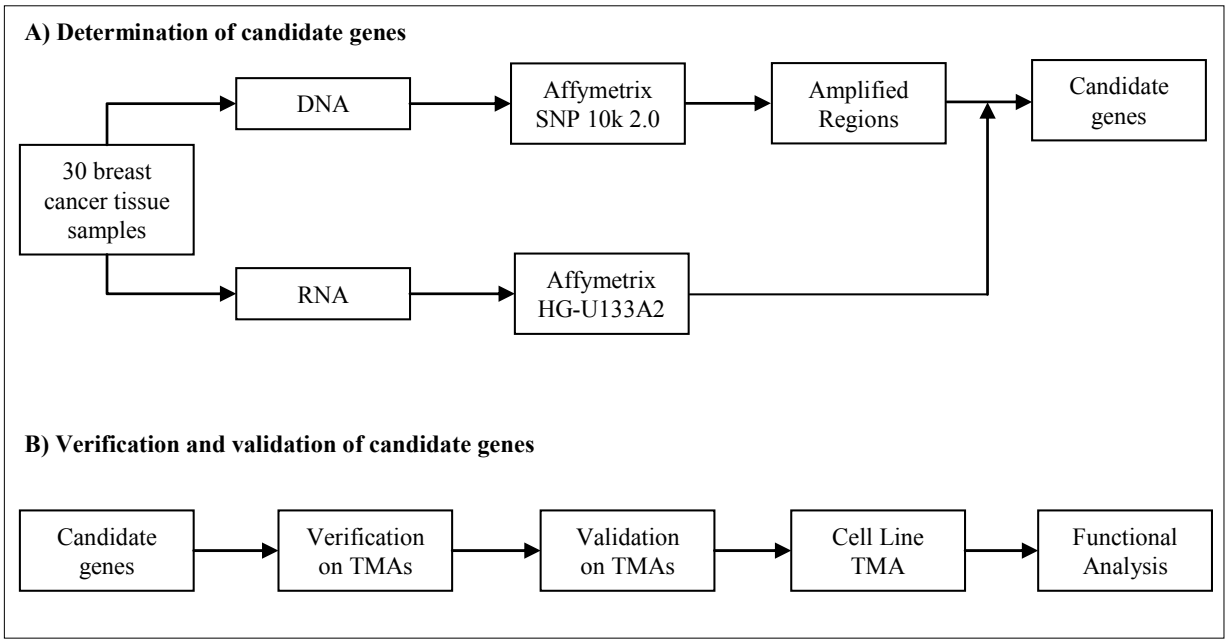


Figure 6. Overview of the project. A) shows the road to the candidate genes. B) shows an overview of the process of verification and validation.

For verification of the amplification status of the candidate genes, fluorescence in-situ hybridization was applied on two small-sized TMAs. These two TMAs (BreVer64 TMA and Breast AMPTest TMA) comprise the samples from the same breast cancer tissues as used for the hybridization of the Affymetrix arrays. The Breast AMPTest TMA contains additional 141 breast cancers and 57 normal breast samples. An overview of the composition of these two TMAs used for verification is shown in Figure 7. After successful verification, the FISH probes were applied on the Breast Prognosis TMA to investigate a potential correlation with survival and other clinicopathological features in breast cancer.

Additionally, screening of the cell line TMA, a tissue microarray containing 120 different tumor cell lines, was planned for the detection of cell lines, which carry the same amplification.

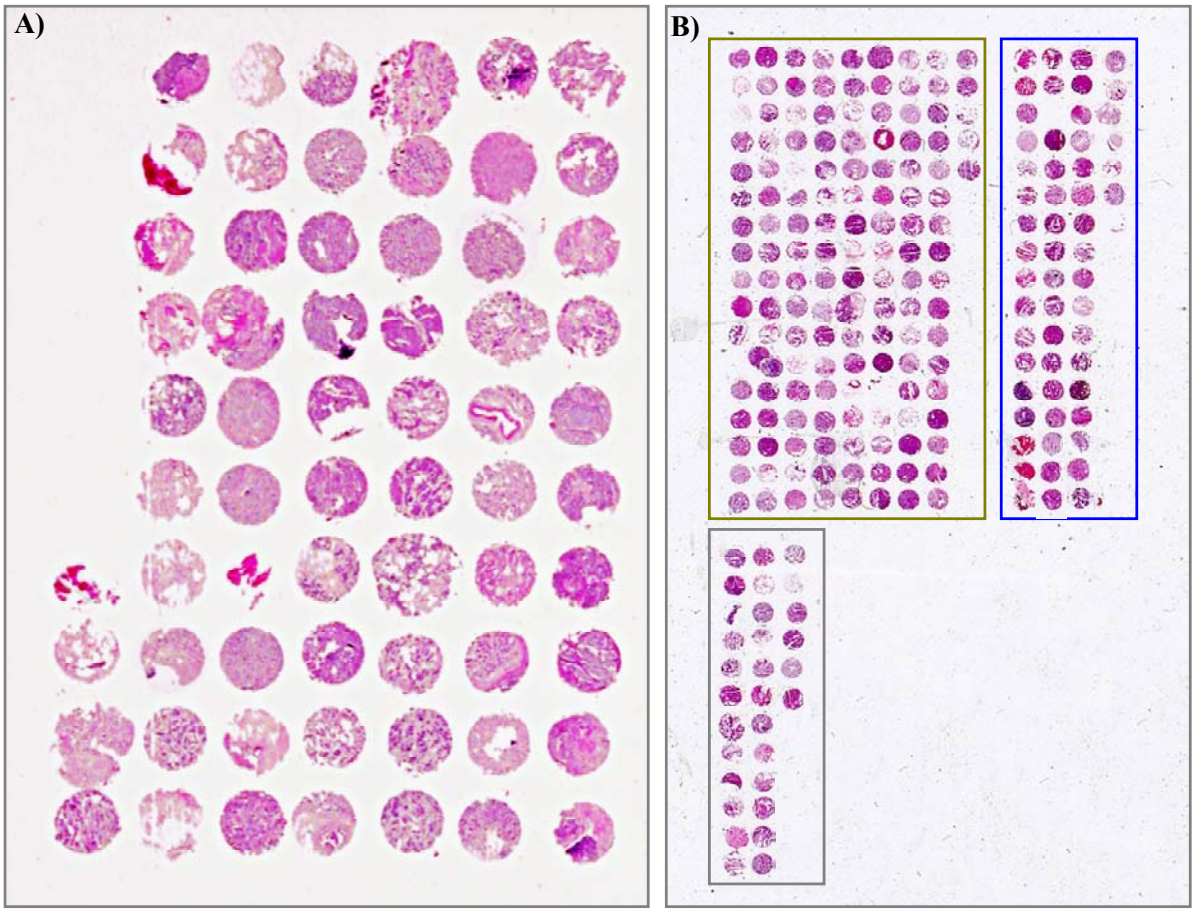


Figure 7. H&E stainings of verification-TMAs constructed in collaboration with the University Medical Center in Hamburg-Eppendorf. A) shows the BreVer64, a very small TMA containing the 30 breast cancer samples used for the hybridization of the Affymetrix arrays. Some breast cancers are represented by two or three cores from different regions of the same tumor. B) shows the Breast AMPTest TMA, comprising the 30 breast cancers from Basel (gray quadrant) and additional 141 breast cancer (green quadrant) and 57 normal breast samples (blue quadrant) from Hamburg.

2.2.2. Discovery of unrecognized amplified regions through development of a novel analysis procedure

2.2.2.1. Analysis of the breast cancer data set with known methods

The standard analysis method recommended by Affymetrix for the analysis of their SNP 10k 2.0 arrays in order to determine DNA copy numbers, is based on the use of their software Chromosome Copy Number Analysis Tool (CNAT) (Huang et al., 2004). This analysis method uses a set of 110 normal ethnically diverse individuals as reference. Each array to be analyzed is compared to this reference set leading to a copy number designation for each SNP on the array. When we performed this kind of analysis on the breast cancer data set described above, unfortunately, the generated results were not as hoped for. Neither known amplifications nor deletions were found using this method. As an example, the output of this analysis method (CNAT) for the chromosome 10 of the sample 02 is visualized in Figure 8A. In order to exclude the possibility that the weak results are due to the visualization method given by the CNAT, we used the DNAcopy package (Venkatraman & Olshen, 2004) from the Bioconductor suite of R to visualize the copy number data calculated by CNAT. The DNAcopy package had been developed once for the analysis of aCGH arrays. After importing copy number data, segments with similar DNA content were calculated and visualized (Figure 8B-C, red lines). The results achieved with this package were slightly better than the results from the CNAT visualization (Figure 8B). This became possible because of the incorporated smoothing function which led to a slight reduction of the noise. Having realized this, we suspected that not the visualization method, but the procedure of the calculation of the copy number data was the weak point of the analysis method used until now. This means, that the use of the Affymetrix reference set may not be the best choice. In order to circumvent this problem, we developed an analysis method which is not dependent on an external reference set. The method is explained in the next section, but as an outlook, Figure 8C shows the result of the above analyzed chromosome with the newly developed method.

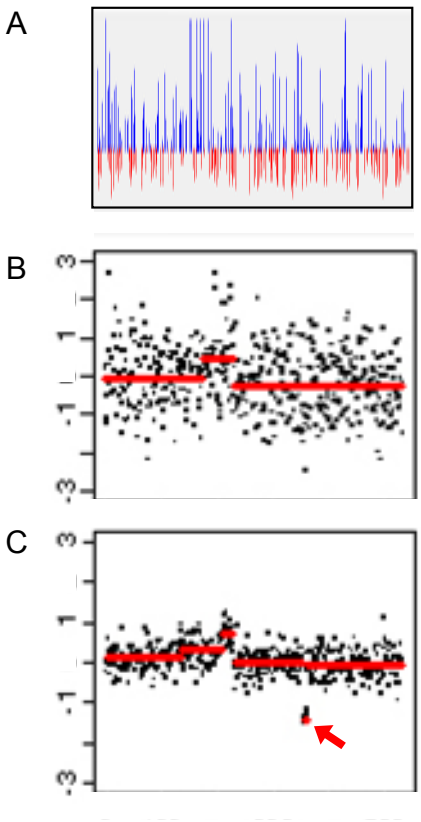


Figure 8. Overview of DNA copy number data of chromosome 10 of sample 02 using different analysis methods. From left to right: p-arm to q-arm. A) Calculation and visualization using Affymetrix' CNAT. B) Calculation using CNAT, visualization using the DNAcopy package. C) The analysis using our new analysis procedure (see next section) leads to the detection of a small deletion of 1.2 Mb, which comprises the PTEN gene (red arrow).

2.2.2.2. Our new analysis procedure

Due to our suspicion that a comparison with external data would bias the calculation of the DNA copy numbers, we decided not to use the reference set provided by Affymetrix. We imported our complete data set into the DChip software (Li & Hung Wong, 2001; Li & Wong, 2001) and used a normalization procedure as recommended for Affymetrix RNA expression arrays (e.g. HG-U133A). In order to reduce the noise, we smoothened our data by applying our own algorithm, as summarized in Figure 9B. In contrast to other methods (e.g. CNAT), only signals from arrays belonging to the same set were used for calculation of the final DNA copy number, but no signals from external array sets were involved. The visualization of the DNA copy numbers along the chromosome was still done with the DNAcopy package. An example of the improvement of our data analysis was seen in Figure 8. The analyzed chromosome 10 in Figure 8C did not only have less noise than the ones in Figure 8A and 8B, but a small deletion of 1.2Mb became detectable. This deletion was previously masked by the noise created by the involvement of the external reference data set.

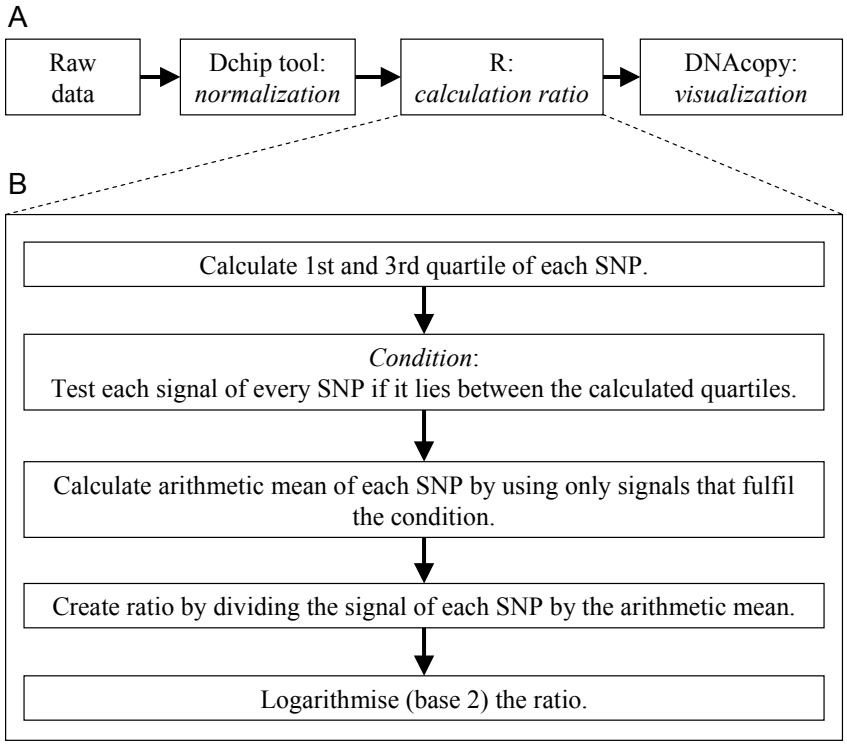


Figure 9. Newly developed analysis procedure. A) shows an overview of the procedure. B) shows the steps of the algorithm used for calculation of the ratio.

Evidence, that the karyograms detected by our analysis procedure are correct, is provided by the findings showing the detection of several already known amplifications and deletions in breast cancer. For example, the PTEN deletion (Kurose et al., 2001) in Figure 8C, the 13q13-14 deletion (Sabbir et al., 2006), which involves the Rb1 gene (Figure 10A), a small amplicon at 8p11-12 (Garcia et al., 2005) (Figure 10B and 10C), a known 17q22-23 amplification (Barlund et al., 1997) (Figure 10D and 10E) and the isochromosome 8 resulting in a 8q Loss / 8q Gain pattern (Figure 10F).

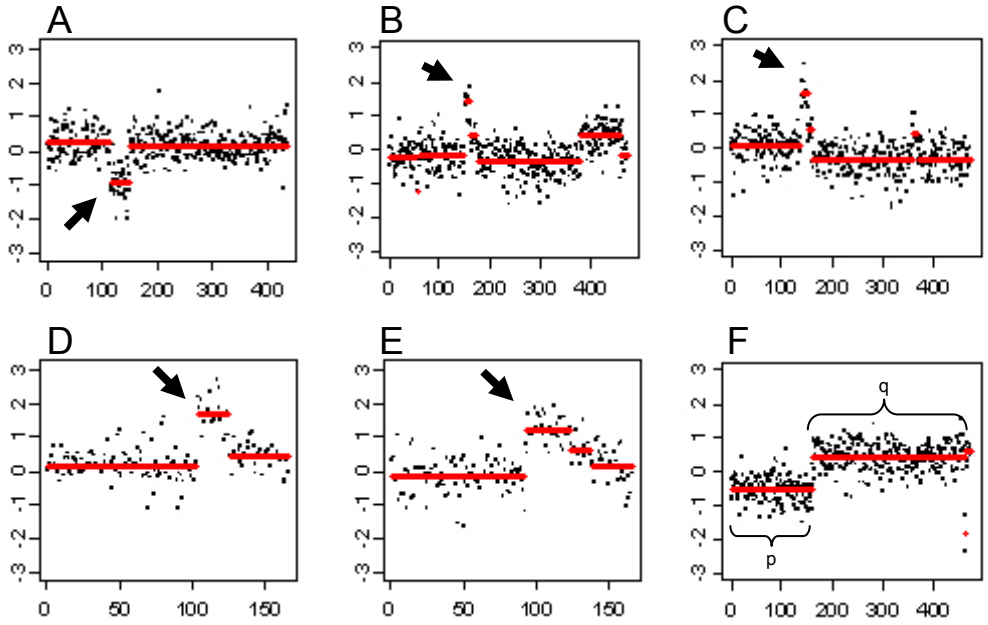


Figure 10. Known chromosomal aberrations found in our breast cancer data set. A)13q13-14 deletion in sample 27. B) 8p11-12 amplification in sample 09. C) 8p11-12 amplification in sample 17. D) 17q22-23 amplification in sample 05. E) 17q22-23 amplification in sample 16. F) 8p loss / 8q gain in sample 07.

2.2.2.3. New analysis method leads to the discovery of unrecognized amplifications in breast cancer

Several amplified regions were found in our breast cancer data set. Most of them were known ones (Figure 10) or were localized in regions devoid of genes. These kinds of amplifications were not further investigated. Our priority was clearly directed towards amplifications, which were found by our novel analysis method in at least two samples of our data set, but had never been described before in breast cancer (for example amplifications in 9p24, 10q22, 12p11; Figure 11A-C). Besides these, we also found amplified regions, which had been described once in the literature, but were never further investigated in breast cancer (6q25 amplification; Figure 11D).

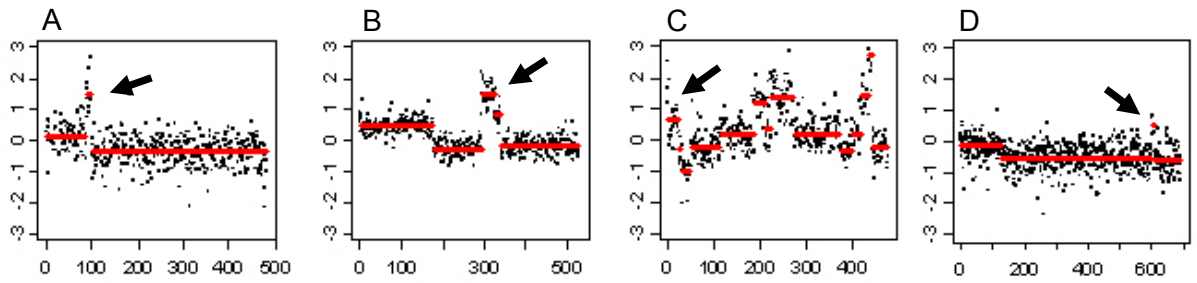


Figure 11. Interesting amplified regions found in our breast cancer data set. A) 9p24 amplification. B) 10q22 amplification. C) 12p11 amplification. D) 6q25 amplification. See text for further explanation.

9p24 (Figure 11A): Our analysis revealed two samples with an amplification of the 9p24 region. The only gene in the overlapping region was the nuclear factor 1 B gene (NFIB). FISH analysis on the BreVer64 TMA with a probe directed against this gene confirmed the amplifications. Further FISH analysis of the 141 Breast samples on the Breast AMPTest TMA from Hamburg revealed a 5% amplification rate. This gene was further investigated (see section 2.2.4).

10q22 (Figure 11B)**:** Amplification of this region has been reported several times in prostate cancer. This is the first time that this amplification was found in breast cancer. The amplification was verified by FISH against the KCNMA1 gene (potassium large conductance calcium-activated channel, subfamily M, alpha member 1) localized in the core of the amplified region. Further analysis of this amplification was not performed, since additional analysis of the 141 breast cancer samples on the Breast AMPTest TMA revealed no further case of amplification.

12p11 (Figure 11C)**:** Two samples from our set showed an amplification of this region. The region comprised the JARID1A gene (Jumonji, AT rich interactive domain 1A), which has been suggested to function as a repressor of differentiation, in the absence of normal pRb protein. FISH analysis confirmed the amplification in the two samples, but again, further analysis of this gene was not performed, since additional FISH analysis did not reveal any further amplified samples on the Breast AMPTest TMA.

6q25 (Figure 11D): Our analysis revealed amplification of the 6q25 region in two samples. The overlapping amplified regions comprised only the estrogen receptor alpha gene (ESR1). This data was verified using a FISH probe on the BreVer64 TMA. Analysis of the Breast AMPTest TMA showed an

overall amplification rate of more than 25%. Due to this high rate and the importance of this gene in breast cancer, it was further investigated (see next section).

2.2.3. Estrogen receptor alpha (ESR1) gene amplification is frequent in breast cancer and predicts response to tamoxifen (manuscript)

Article resubmitted to Nature Genetics.

ESTROGEN RECEPTOR ALPHA (ESR1) GENE AMPLIFICATION IS FREQUENT IN BREAST CANCER AND PREDICTS RESPONSE TO TAMOXIFEN

Frederik Holst[1]*, Phillip Stahl[1]*, Christian Ruiz[2]*, Zeenath Jehan[3], Annette Lebeau[1], Luigi Terracciano[2], Khawla Al-Kuraya[3], Fritz Jänicke[4], Guido Sauter[1], and Ronald Simon[1]

[1]Dept. of Pathology, University Medical Center Hamburg Eppendorf, Hamburg, Germany

[2]Dept. of Pathology, University Hospital Basel, Basel, Switzerland

[3]King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia

[4]Dept. of Gynecology, University Medical Center Hamburg Eppendorf, Hamburg, Germany

* These authors have equally contributed to the study

Please address all correspondence to:

Ronald Simon, Ph.D.

University Medical Center Hamburg-Eppendorf

Department of Pathology

Martinistrasse 52

D-20246 Hamburg / Germany

phone +49 (0)40 42803 7214

fax +49 (0)40 42803 5997

Running Title: Estrogen receptor gene amplification in breast cancer

Abstract

To search for previously undetected biologically important gene amplifications in breast cancers we used 10k Affymetrix DNA arrays. In one of 22 cancers we found a 6q25.1 amplification that involved only 8 SNPs, all located within one gene: Estrogen receptor alpha (ESR1). The subsequent analysis of more than 2000 clinical breast cancer samples revealed ESR1 amplification in 20.6% and lower level ESR1 copy number increases (ESR1 gains) in 15% of cancers. ESR1 amplified cancers showed estrogen receptor protein overexpression in 99% as compared to 66.6% of non-amplified cancers ($p<0.0001$). Within the group of estrogen receptor positive breast cancers, ESR1 amplification was an independent predictor of response to adjuvant anti-hormonal therapy. In a subset of 261 patients who had received adjuvant tamoxifen monotherapy, survival in women with ESR1 amplified cancers was significantly longer than in women with non-amplified estrogen receptor expressing cancers ($p=0.023$). Remarkably, ESR1 amplification was also found in a multitude of benign and precancerous breast diseases, suggesting that ESR1 amplification may be a very early – if not the first – genetic alteration in a large subset of breast cancers.

Introduction

Breast cancer is the leading malignancy in women, accounting for more than 350.000 deaths per year worldwide [1]. Several molecular pathways are known to play a role in breast cancer development and progression. Perhaps the most important pathway involves estrogen receptor alpha protein (ER). Binding of estrogen to ER causes phosphorylation and dimerization of the receptor that acts as a transcriptional promoter conferring a growth signal to breast epithelial cells [2]. This makes ER one of the most important therapeutic targets in breast cancer [3]. More than two thirds of breast cancers show expression of ER at the time of diagnosis [4], and immunohistochemical detection of ER expression is routinely used for decision making for hormonal (anti-ER) therapy of breast cancer [5]. Current anti ER treatment strategies include blocking by selective modulators (SERMs, e.g. Tamoxifen, Raloxifen), destabilization and degradation of ER by selective downregulators (SERDs, e.g. Fulvestrant) or disruption of estrogen synthesis (aromatase inhibitors, e.g. Anastozole, Letrozole, Exemestan) which results in a significant decrease of tumor growth in about 30-50% of patients [6].

Gene copy number increase (amplification) is a major mechanism for cancer cells to boost the expression of gene products that provide them with a growth or survival advantage. Numerous genes have been found amplified in a fraction of breast cancers including HER2 at 17q21 [7], CMYC at 8q24 [8], EGFR at 7p12 [9], AIB1 at 20q13 [10], or CCND1 at 11q13 [11]. The example of HER2 also shows that amplified genes may be particularly suited as therapeutic targets. Trastuzumab (Herceptin) is highly effective in the treatment of HER2 amplified/overexpressing breast cancers [12,13]. Several lines of evidence suggest that various genes that can give a growth advantage to breast epithelial cells through amplification are still not discovered. First, amplicons usually span several megabases of genomic distance and harbor numerous genes. This makes it difficult to determine which gene(s) drive amplification. Second, amplifications may be very small and therefore difficult to detect by classical amplification screening methods [14]. In this project we used a three-step approach to search for previously undetected, clinically relevant gene amplifications in breast cancer. First, we used a DNA

microarray approach (Affymetrix 10k mapping array) to screen for gene copy number changes in 22 high-grade breast cancers. Second, we specifically filtered these DNA copy number data for small size amplifications of potentially relevant gene classes. Then we used fluorescence in situ hybridization (FISH) to validate identified amplifications and to screen for their clinical relevance on a tissue microarray (TMA) containing samples from more than 2,000 breast cancer patients with clinical follow up data [15]. Stunningly, this approach exposed amplification of the ESR1 gene encoding estrogen receptor alpha (ER) as a key mechanism in breast cancer development.

Results

DNA Chip analysis. We used Affymetrix 10k SNP chips to search for genomic loci harboring gene amplifications in breast cancer. Because most gene amplifications are linked to high-grade cancers, we selected 22 grade 3 breast cancers (according to Bloom, Richardson and Ellis [16]) for analysis. Among several other amplifications (data not shown), we found 2 samples (9%) with amplification at chromosome 6q25.1. One of these tumors had a very narrow region of amplification at 6q25.1 that was detectable on 8 SNPs and spanned 514Kb (base positions 152.025.289- 152.539.833). The estrogen receptor alpha encoding gene ESR1 was the only gene inside this area (figure 1). This observation raised the possibility, that gene amplification might be a previously unrecognized key mechanism for overexpression of this critical breast cancer gene.

ESR1 amplification in breast cancer. To investigate the possible importance of our observation, we analyzed an existing breast cancer tissue microarray (TMA) containing more than 2,000 breast cancers [15]. For fluorescence in situ hybridization (FISH) analysis we prepared a probe from BAC RP11-450E24 located inside the ESR1 gene. FISH analysis of ESR1 was successful in 1739/2197 (79%) arrayed breast cancer tissues. No results could be obtained from 458 cases, either due to lack of tumor cells in the tissue spot, complete loss of tissue spots, or insufficient hybridization. We defined the threshold for ESR1 amplification according to the generally accepted scoring system used for HER2 amplification evaluation in FDA approved test kits (PathVysion, Abbott, IL). ESR1 amplification was assumed if at least two times more ESR1 signals than centromere 6 (cen 6) signals were present (ratio ESR1/cen6 $\geq$2.0). Tissue samples with an ESR1/cen 6 ratio larger than 1.0 but less than 2.0 (1.0 < ratio ESR1/cen6 < 2.0) were classified as „ESR1 gains“. All other tissues (ratio ESR1/cen6 $\leq$ 1.0) were considered normal. According to these criteria, we found ESR1 amplification in 358/1739 (20.6%) analyzable tissue samples. ESR1 copy number gains were present in another 266 (15.3%) of tumors. Examples of amplified and non-amplified cancer cells are shown in figure 2. Most amplified tumors showed a clustered arrangement of additional ESR1 copies indicating intrachromosomal

amplification type (homogeneously staining regions), and had moderate absolute copy numbers ranging between 6 and 9 copies. Only 18% of amplified cases had 10 or more ESR1 gene copies per cell. We next studied the relationship between ESR1 copy number changes and breast cancer phenotype (table 1). ESR1 amplifications and gains were significantly associated with low-grade tumors ($p<0.0001$) and absence of lymph node metastases ($p<0.05$). This finding was unexpected, because gene amplifications are usually linked to high grade and advanced cancers [9]. The frequency of ESR1 amplification varied considerably between cancers of different histological subtypes. For example, medullary cancers, which are characterized by a high-grade phenotype, had a particularly low fraction of amplified tumors (2.0%), whereas mucinous cancers that are mostly low-grade had particularly high rates of ESR1 amplifications (35.6%).

ESR1 amplification in non-malignant and premalignant breast. Because of the association of ESR1 amplification with low malignancy cancer phenotypes, we expanded our analysis to a series of non-malignant and preneoplastic tissues. Most remarkably, ESR1 amplification was also frequently found in several of these tissues (table 2). A particularly high frequency of ESR1 amplification was seen in benign papillomas (8/22, 36%), but amplification was also present in usual ductal hyperplasias (1/12, 8.3%). These findings suggested that ESR1 amplifications may play a key role in a distinct molecular "road to breast cancer" characterized by particularly high levels of estrogen receptor alpha expression.

Impact on estrogen receptor alpha protein expression levels. To investigate the impact of ESR1 amplification on ER protein levels we next compared ESR1 gene copy numbers to immunohistochemical ER protein expression levels. IHC analysis was successful in 2018/2197 (92%) breast cancers (table 1). As expected, ESR1 amplification was tightly linked to ER protein expression ($p<0.0001$, table 3). Virtually all cases with increased ESR1 gene copy numbers (amplifications and gains) had high-level expression. Among 341 breast cancers with ESR1 amplification, 339 (99%) had detectable ER expression. The vast majority of these tumors (94%) had the highest ER scores (7-8) according to Allred [17] . This was also true for tumors exhibiting ESR1 gains where 89% of samples

showed strong (score 7-8) ER expression. However, the data also showed that ESR1 amplifications or gains were not the sole reason for high-level ER expression. Almost half (46%) of strong (score 7-8) ER expressors had no ESR1 gene copy number alterations. Given the paramount role of gene amplification for response to other targeted therapies, this result raised the question of whether ESR1 amplified breast cancers might respond differently to anti-hormonal therapies than non-amplified overexpressors.

ESR1 amplification and response to tamoxifen. To determine the predicitive impact of ESR1 gene amplification on response to hormonal treatment, we analyzed a subgroup of 261 patients that had previously undergone adjuvant monotherapy with tamoxifen. For this analysis, we stratified the patients into three groups: I) ER IHC negative cancers (Allred scores 0-2), II) ER IHC positive cancers (Allred scores 3-8) lacking ESR1 amplification, and III) ESR1 amplified cancers. It showed that ESR1 amplified cancers are indeed highly responsive to tamoxifen (Figure 3). As expected, the combined DNA/protein analysis revealed the worst survival in IHC negative cancers (figure 3a). However, patient survival was significantly better in the ESR1 amplified population as compared to patients with non amplified ER positive cancers ($p=0.0233$; ESR1 amplification versus ER positive in figure 3a). To exclude a potential survival bias caused by classical prognostic factors of breast cancer we performed a multivariate analysis (Cox proportional hazards) including ER expression, tumor grade, pT and pN categories. We found that the prognostic value of ESR1 amplification in tamoxifen treated patients was independent of these factors (table 4). Stunningly, the prognostic impact of ESR1 amplification was even retained in the subgroup of cancers with strongest ER expression (figure 3b; ESR1 amplification versus ER 7-8, $p=0.0889$). Furthermore, response to tamoxifen was also dependent on the absolute ESR1 copy number (figure 3c).

Discussion

In this project, a "FISHing expedition" was necessary to discover that the most frequently analyzed and - on the protein level – best-characterized gene in breast cancer (more than 30,000 Medline entries so far) is a common subject of gene amplification. Although only one small tissue sample (0.6mm diameter) was analyzed per tumor, ESR1 amplification was detected in 20.6 % of breast cancers. Using the same breast cancer TMA and the same definition of amplification, we previously found amplifications of CCND1 in 20.1%, HER2 in 17.3%, MDM2 in 5.7%, CMYC in 5.3%, and EGFR in 0.8% of the tumors [9]. Given the critical role of estrogen receptor expression for breast cancer therapy and the well-known importance of gene amplification for drug target overexpression, it is remarkable that the critical relevance of ESR1 amplification has not been discovered earlier. To a large extent, the difficulties detecting the 6q25.1 amplicon appear to be caused by the generally small size of the amplicon. Preliminary mapping experiments suggest that the amplicon measures less than 1 megabase in the majority of cases (data not shown). In our DNA microarray analysis, amplification was detected only in two cases, including one tumor where the amplicon was limited to an approx. 514Kb sequence harboring only the ESR1 gene. Such single gene amplicons are difficult to identify even with high-resolution CGH arrays. Only one matrix CGH study had previously – together with 30 other amplicons - described a 6q25.1 amplification containing ESR1 in 3 of 31 analyzed breast cancers [18]. Because the experimental noise is often considerable in array hybridization experiments, single spot peaks are frequently seen and artifacts are difficult to distinguish from true amplification events.

The availability of large, well-characterized TMAs allowed us to extensively study the epidemiology of ESR1 amplifications. The striking association with positive ER IHC – 99% of our amplified cases were also found positive by IHC – and the ER expression levels not only validated our experimental approach but also demonstrated a strong functional importance of ESR1 gene amplification. Apparently, ESR1 amplification or gain is present in a subset of about 50% highly ER expressing breast cancers. This observation was interesting in the light of the striking association of other gene

amplifications with response to biological drugs targeting the specific gene products. Notably, amplification of HER2, EGFR and also TOP2A was always more predictive for response to therapy than protein overexpression detected by IHC [19-23]. The superior predictive role of gene amplification may partly be due to inherent technical issues of IHC as immunostaining is subject to tissue processing variability and not linearly related to the protein quantity. On the other hand, gene amplification is a result of complex selection processes, and it is tempting to speculate that genes that have undergone amplification in a particular cell clone may be critically important for cell growth and survival in this clone and thus represent the Achilles tendon for targeted therapy.

The availability of 261 breast cancers from patients who had undergone adjuvant anti-ER monotherapy with tamoxifen allowed us to investigate the possible impact of ESR1 amplification on response to anti-hormonal therapy. Despite of the small size of this cohort, the retrospective nature of data collection, and the potentially variable duration of tamoxifen therapy, a marked survival difference became apparent between ESR1 amplified and ESR1 non-amplified IHC positive tumors. Remarkably, this association was independent of grade, stage, and pN status in our multivariate analysis. Altogether, these data strongly suggest that ESR1 amplification may identify a subgroup of ER positive breast cancers with a particularly high likelihood to positively response to anti-estrogen therapy. This result could not completely be expected, because an opposite mechanism is well known to occur in prostate cancer. In this tumor type, amplification of the androgen receptor does not occur in untreated primary tumors but develops in about 20-30% of cases under anti hormonal therapy and causes resistance to further anti-hormonal therapy [24-26].

The high frequency of ESR1 amplification in low grade and early stage breast cancer together with the critical role of ER for proliferation control in breast epithelium would be well consistent with a very early – if not initializing – role of ER amplification for a subset of breast cancers. To investigate the role of ESR1 amplification in premalignant proliferative breast lesions another set of breast lesions including non-neoplastic and precancerous tissues was analyzed for ESR1 amplification. The observation of frequent and often high level ESR1 amplifications not only in carcinoma in situ but as well in papillomas (36%) and usual ductal hyperplasias (8%) was striking. Gene amplification has so

far never been demonstrated in benign breast lesions. These findings suggest that ESR1 amplification is not a hallmark for invasive breast cancer but may represent a key mechanism for development of various types of proliferative breast disease. As gene amplification is thought to be non-reversible, it is tempting to speculate that ER amplification may constitute a decisive mechanism for initiation of neoplastic (clonal) breast disease. It appears possible that presence of ESR1 amplification may therefore be indicative of proliferative breast disease with increased potential for malignant transformation. This hypothesis would be consistent with previous IHC studies suggesting a possible link between high-level ER expression in benign proliferative breast disease and an increased risk for breast cancer development [27,28]. It is thus possible, that ESR1 amplification status represents a useful prognostic marker in patients with benign appearing proliferative breast lesions.

In summary, our data suggest that ESR1 amplification is a frequent event in proliferating breast disease and breast cancer. ESR1 amplification defines a subtype of primary breast cancers that optimally benefits from anti hormonal therapy.

Materials and Methods

Tissues. For DNA Chip analysis experiments, 22 fresh frozen grade 3 breast cancer samples were selected from the frozen tissue archive of the Institute of Pathology, University Hospital Basel, Switzerland. For epidemiological evaluation of ESR1 amplifications, our existing Breast Prognosis TMA containing more than 2000 breast tissues was analyzed. The composition of this TMA has been described before [15]. The patho-histological and clinical data of these tissues are summarized in table 1. The type of adjuvant treatment was known for 420 patients. A subset of 261 patients received hormonal treatment with tamoxifen (5 years was standard, but detailed information was not available) but no adjuvant cytotoxic chemotherapy. The patients were treated and evaluated at the University Hospital in Basel (Switzerland), the Womens' Hospital Rheinfelden (Germany), and the Kreiskrankenhaus Lörrach (Germany) between 1985 and 2000. The patients' attending physicians were contacted to collect follow up data. The median observation time was 60 months (range 2-138). Staging variables were collected from the patient files. In addition, a second TMA containing 50 normal breast samples and 186 non-neoplastic and premalignant breast lesions was analyzed. This TMA included 62 DCIS, 10 LCIS (lobular neoplasias), 14 apocrine metaplasias, 27 usual ductal hyperplasias, 5 atypical ductal hyperplasias, 22 cases of fibrocystic disease, 31 papillomas, and 15 cases of sclerosing adenosis.

DNA Chip analysis. DNA was extracted from three punched tissue cylinders (diameter 0.6 mm) from each of the 30 fresh frozen tumor samples according to the manufacturer's instructions of the QIAmp DNA Mini Kit (Qiagen, Hilden, Germany). DNA was further processed as described in the GeneChip Mapping 10K 2.0 Assay Manual (Affymetrix, Santa Clara, CA). Briefly, 250ng of the DNA was digested with the XBA I restriction enzyme, ligated to an adaptor and amplified by PCR. The resulting PCR products were fragmented, end-labeled and hybridized to the GeneChip Human Mapping 10K Array Xba 142 2.0. After hybridization, the microarray chip was washed and stained on an Affymetrix fluidics station. The chips were scanned using the Affymetrix GeneChip scanner 3000.

Data Acquisition and Analysis. Raw data from the scanned arrays was acquired using the GeneChip Operating Software (Affymetrix). Quality of the data was checked as described in the GeneChip Mapping 10K 2.0 Assay Manual (Affymetrix). Only 22 of the samples passed these quality controls and were used for further data analysis. The data files were imported into the DChip software [29], and pre-processing and normalization were performed as described in the user manual. The resulting signal intensities were imported into the R suite [30]. Data were further processed with a home made analysis tool using all measurements between the 25th and 75th percentile as a reference for each DNA spot (code available upon request, C. Ruiz). Data were then modified to fit the input requirements of the DNAcopy package [31] of the Bioconductor suite [32]. The DNAcopy package was used to calculate and visualize segments with similar DNA content. Segments with higher DNA content than the above calculated reference were classified as candidate regions for gene amplification.

Fluorescence in situ hybridization (FISH). TMA sections were treated according to the Paraffin Pretreatment Reagent Kit protocol (Vysis, Downers Grove, IL) before hybridization. FISH was performed with a digoxigenated BAC probe (BAC RP11-450E24, RZPD, Germany) containing the ESR1 gene and a Spectrum-Orange labeled chromosome 6 centromeric probe (CEP6) as a reference (purchased from Vysis). Hybridization and posthybridization washes were according to the 'LSI procedure' (Vysis). Probe visualization using fluorescent isothiocyanate (FITC)-conjugated sheep anti-digoxigenin (Roche Diagnostics, Rotkreuz, Switzerland) was as described [33]. Slides were counterstained with 125 ng/ml 4',6-diamino-2-phenylindole in an antifade solution. Hybridization and posthybridization washes were according to the 'LSI procedure' (Vysis). Slides were then counterstained with 125 ng/ml 4',6-diamino-2-phenylindole in an antifade solution.

Immunohistochemistry. Immunohistochemical detection of ER alpha protein was performed using antibody NCL-L-ER-6F11 (Novocastra, Newcastle, UK). In brief, slides were deparaffinized and incubated in a pressure cooker at 120°C for 12 min in pH6 citrate buffer (Retrievit 6 #BS-1006-00, BioGenex, San Ramon, CA). After blocking of endogeneous peroxidase, prediluted (1:1000) primary antibody was applied and the slides were incubated overnight at 4°C. The Vectastain ABC Elite

system was used for detection of antibody binding. IHC scoring was performed according to the Allred score [17]. In brief, ER staining intensity was recorded in a 4-step scale (0-3) and the fraction of ER positive tumor cells in a 5-step 1-5) scale. Combination of both parameters results in an 8-step score, where all samples with score >2 are regarded as ER positive.

Statistics. Contingency table analysis and Chi-square tests were used to study the relationship between histological tumor type, grade, stage and gene amplification. Survival curves were plotted according to Kaplan-Meier. A log rank test was applied to examine the relationship between gene amplifications and patient survival. Cox regression analysis was performed to test for interdependencies between analyzed variables in their relation with patient survival.

Tables and Figures:

Table 1. Relationship between ESR1 copy number changes and invasive breast cancer phenotype

			ESR1 FISH result				ER IHC result		
		on array (n)	analyzable (n)	amplification	gain	p	analyzable (n)	positive*	p
Cancers	all	2197	1739	20.6%	15.3%		2018	76.6%	
Histology	ductal carcinoma	1552	1207	21.5%	16.0%		1429	77.1%	
	lobular carcinoma	312	207	19.3%	13.8%		275	87.6%	
	mucinous carcinoma	69	37	35.6%	24.4%	0.0337**	61	93.4%	<0.0001**
	medullary carcinoma	58	48	2.0%	2.0%	<0.0001**	52	17.3%	<0.0001**
	tubular carcinoma	56	42	18.6%	14.0%		48	89.6%	
	cribriform carcinoma	65	55	29.8%	12.3%		56	91.1%	
	papillary carcinoma	30	27	19.2%	15.4%		28	67.9%	
	others*	79	56	4.9%	14.8%		69	34.8%	
pT stage	pT1	820	578	22.3%	15.6%	0.7295	716	80.4%	0.0020
	pT2	1023	811	19.7%	14.8%		948	73.2%	
	pT3	124	92	18.8%	13.5%		114	72.8%	
	pT4	242	189	17.8%	16.8%		229	80.3%	
Nodal stage	pN0	950	711	22.3%	14.4%	0.0422	849	78.1%	0.1765
	pN1	793	608	16.5%	16.5%		726	75.6%	
	pN2	121	90	14.8%	12.5%		113	70.8%	
BRE grade	G1	545	421	25.6%	15.2%	<0.0001	522	92.9%	<0.0001
	G2	844	685	21.6%	18.5%		833	86.3%	
	G3	655	571	15.0%	11.7%		661	51.9%	

* according to Allred score;
** versus ductal carcinoma

Table 2. Prevalence of ESR1 copy number changes in normal and premalignant breast tissues

Histology	on array (n)	analyzable (n)	ESR1 FISH result	
			amplification (%)	gain (%)
normal breast tissue	50	21	0.0	0.0
fibrocystic disease	22	13	0.0	7.7
apocrine metaplasia	14	4	0.0	0.0
usual ductal hyperplasia	27	12	8.3	25.0
atypical ductal hyperplasia	5	1	0.0	100.0
sclerosing adenosis	15	8	0.0	0.0
papilloma	31	22	36.4	4.5
DCIS	62	40	35.0	7.5
LCIS (lobular neoplasia)	10	3	33.3	33.3

Table 3. Comparison of ER amplification and expression

ESR1 FISH result		ER IHC result (Allred score)			
	n	0-2 (%)	3-4 (%)	5-6 (%)	7-8 (%)
normal	1056	33.3	8.2	14.1	44.3
gain	255	2.4	1.6	7.5	88.6
amp	341	0.6	0.3	4.7	94.4

Table 4. Contribution of potential prognostic factors to tumor specific survival in breast cancer patients that received Tamoxifen monotherapy (multivariate COX regression model)

Parameter		**p-value**
pT stage	pT1-4	0.3867
BRE grade	G1-3	0.1568
pN stage	pN0-pN2	<0.0001
ER expression status	neg/pos (Allred score)	0.0003
ESR1 amplification status	Normal vs amplified/gained	0.0080

Figure 1. Position of the 6q25.1 amplicon spanning 1.2Mb between base positions 151.538.830 and 152.729.709. The blue and green bar indicates localization of the amplified regions identified in two breast cancer samples BD12 and BD25.

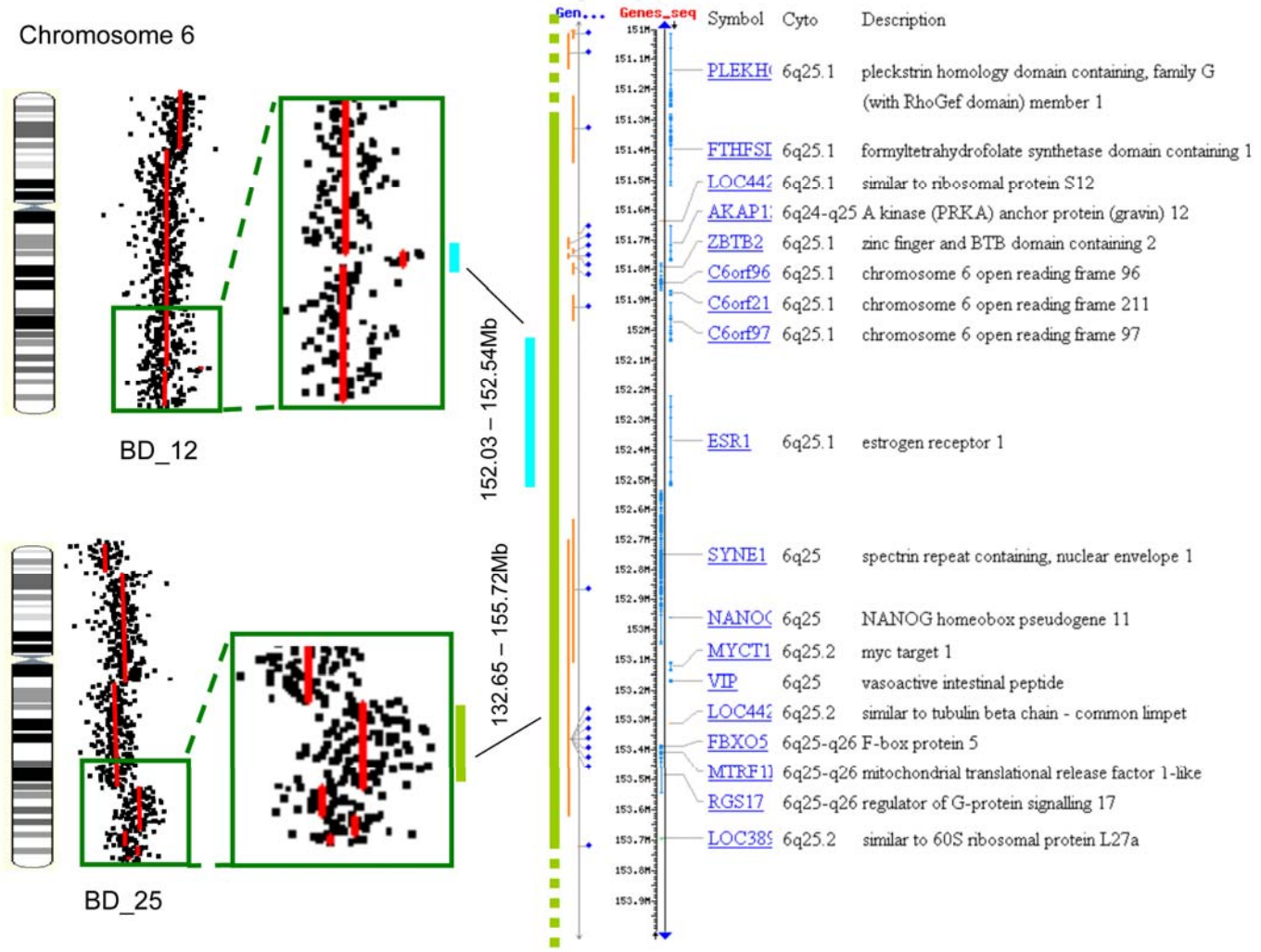


Figure 2. ESR1 FISH analysis in breast cancer.

a) Cell nuclei showing clouds of green ESR1 gene copies indicating intrachromosomal ESR1 amplification. Red signals represent centromere 6. Insert: Higher magnification of a cell nucleus showing 2 red centromere 6 signals and >10 green ESR1 signals.

b) Nuclei with normal ESR1 copy number (2 green ESR1 signals and 2 red centromere 6 signals per cell nuclei).

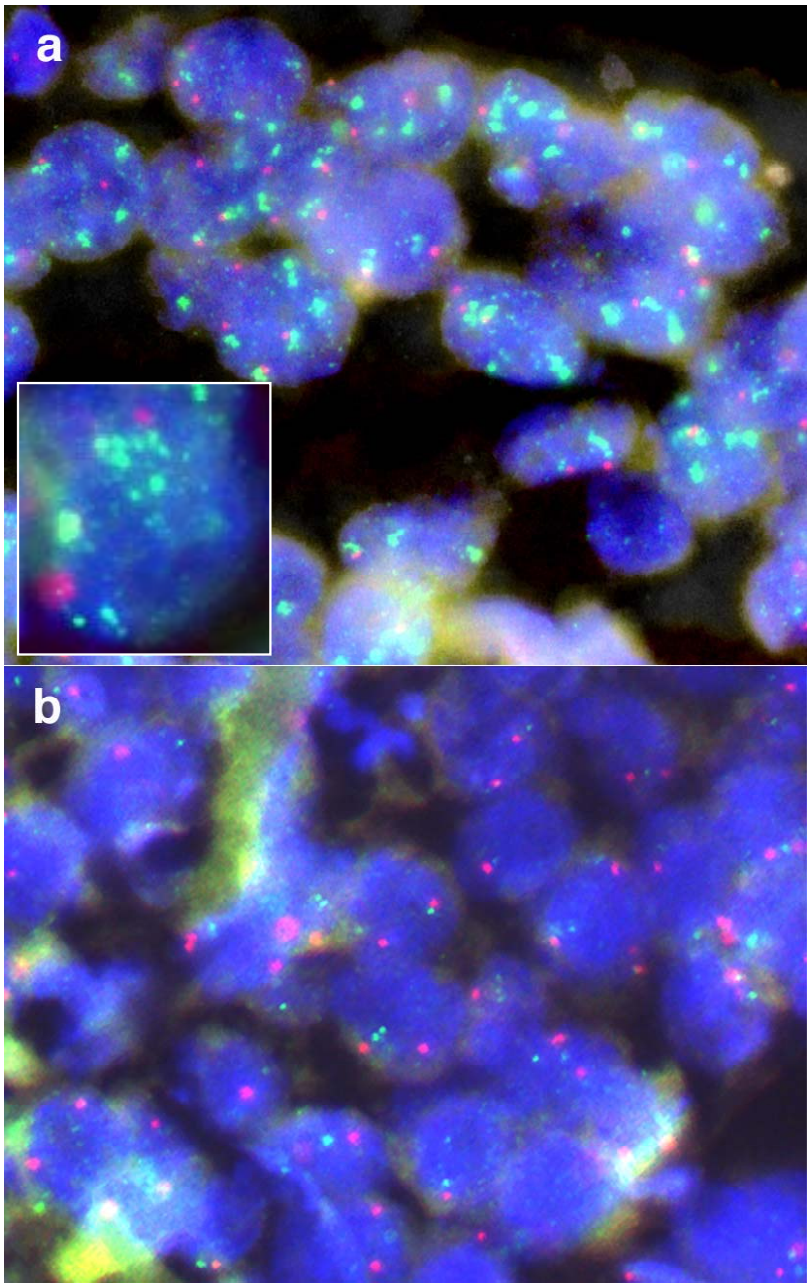
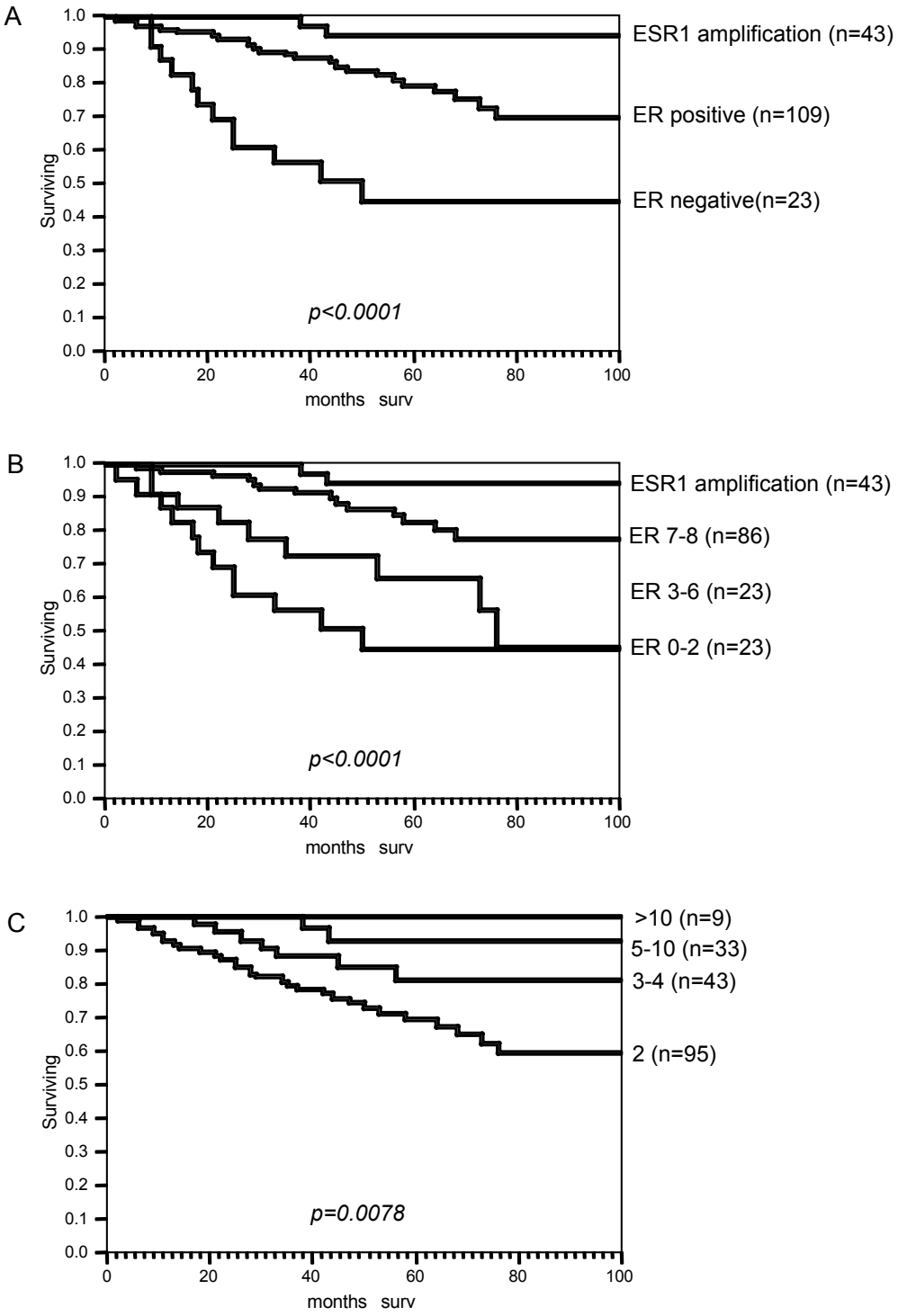


Figure 3. Impact of ESR1 amplification (defined as ESR1 gene to Cen6 copy number ratio ≥2.0) and expression on prognosis in patient that received Tamoxifen monotherapy. a) Immunohistochemistry results grouped into ER negative (scores 0-2) and ER positive (scores 3-8, no ESR1 amplification) according to Allred as compared to ESR1 amplification. b) Immunohistochemistry results by Allred scores versus ESR1 amplification. c) Impact of the ESR1 copy number on patient survival.



References

1. Ferlay, J., Bray, F., Pisani, P. & Parkin, D.M. *GLOBOCAN 2000: Cancer Incidence, Mortality and Prevalence Worldwide*, (IARCPress, Lyon, 2000).
2. Parker, M.G., Arbuckle, N., Dauvois, S., Danielian, P. & White, R. Structure and function of the estrogen receptor. *Ann N Y Acad Sci* **684**, 119-26 (1993).
3. Sunderland, M.C. & Osborne, C.K. Tamoxifen in premenopausal patients with metastatic breast cancer: a review. *J Clin Oncol* **9**, 1283-97 (1991).
4. Stierer, M. et al. Immunohistochemical and biochemical measurement of estrogen and progesterone receptors in primary breast cancer. Correlation of histopathology and prognostic factors. *Ann Surg* **218**, 13-21 (1993).
5. Andersen, J. & Poulsen, H.S. Immunohistochemical estrogen receptor determination in paraffin-embedded tissue. Prediction of response to hormonal treatment in advanced breast cancer. *Cancer* **64**, 1901-8 (1989).
6. Wolmark, N. & Dunn, B.K. The role of tamoxifen in breast cancer prevention: issues sparked by the NSABP Breast Cancer Prevention Trial (P-1). *Ann N Y Acad Sci* **949**, 99-108 (2001).
7. Pauletti, G., Godolphin, W., Press, M.F. & Slamon, D.J. Detection and quantitation of HER-2/neu gene amplification in human breast cancer archival material using fluorescence in situ hybridization. *Oncogene* **13**, 63-72 (1996).
8. Visscher, D.W., Wallis, T., Awussah, S., Mohamed, A. & Crissman, J.D. Evaluation of MYC and chromosome 8 copy number in breast carcinoma by interphase cytogenetics. *Genes Chromosomes Cancer* **18**, 1-7 (1997).
9. Al-Kuraya, K. et al. Prognostic relevance of gene amplifications and coamplifications in breast cancer. *Cancer Res* **64**, 8534-40 (2004).
10. Anzick, S.L. et al. AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. *Science* **277**, 965-8 (1997).
11. Simpson, J.F., Quan, D.E., O'Malley, F., Odom-Maryon, T. & Clarke, P.E. Amplification of CCND1 and expression of its protein product, cyclin D1, in ductal carcinoma in situ of the breast. *Am J Pathol* **151**, 161-8 (1997).
12. Albanell, J. & Baselga, J. Trastuzumab, a humanized anti-HER2 monoclonal antibody, for the treatment of breast cancer. *Drugs Today (Barc)* **35**, 931-46 (1999).
13. Pegram, M.D. et al. Phase II study of receptor-enhanced chemosensitivity using recombinant humanized anti-p185HER2/neu monoclonal antibody plus cisplatin in patients with HER2/neu-overexpressing metastatic breast cancer refractory to chemotherapy treatment. *J Clin Oncol* **16**, 2659-71 (1998).
14. du Manoir, S. et al. Quantitative analysis of comparative genomic hybridization. *Cytometry* **19**, 27-41 (1995).
15. Ruiz, C. et al. Tissue microarrays for comparing molecular features with proliferation activity in breast cancer. *Int J Cancer* **118**, 2190-4 (2006).
16. Elston, C.W. & Ellis, I.O. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* **19**, 403-10 (1991).
17. Harvey, J.M., Clark, G.M., Osborne, C.K. & Allred, D.C. Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. *J Clin Oncol* **17**, 1474-81 (1999).
18. Nessling, M. et al. Candidate genes in breast cancer revealed by microarray-based comparative genomic hybridization of archived tissue. *Cancer Res* **65**, 439-47 (2005).
19. Mass, R.D. et al. Evaluation of clinical outcomes according to HER2 detection by fluorescence in situ hybridization in women with metastatic breast cancer treated with trastuzumab. *Clin Breast Cancer* **6**, 240-6 (2005).

20. Cappuzzo, F. et al. Increased HER2 gene copy number is associated with response to gefitinib therapy in epidermal growth factor receptor-positive non-small-cell lung cancer patients. *J Clin Oncol* **23**, 5007-18 (2005).
21. Hirsch, F.R. et al. Increased epidermal growth factor receptor gene copy number detected by fluorescence in situ hybridization associates with increased sensitivity to gefitinib in patients with bronchioloalveolar carcinoma subtypes: a Southwest Oncology Group Study. *J Clin Oncol* **23**, 6838-45 (2005).
22. Takano, T. et al. Epidermal growth factor receptor gene mutations and increased copy numbers predict gefitinib sensitivity in patients with recurrent non-small-cell lung cancer. *J Clin Oncol* **23**, 6829-37 (2005).
23. Jarvinen, T.A. et al. Amplification and deletion of topoisomerase IIalpha associate with ErbB-2 amplification and affect sensitivity to topoisomerase II inhibitor doxorubicin in breast cancer. *Am J Pathol* **156**, 839-47 (2000).
24. Visakorpi, T. et al. In vivo amplification of the androgen receptor gene and progression of human prostate cancer. *Nat Genet* **9**, 401-6 (1995).
25. Koivisto, P. et al. Androgen receptor gene amplification: a possible molecular mechanism for androgen deprivation therapy failure in prostate cancer. *Cancer Res* **57**, 314-9 (1997).
26. Bubendorf, L. et al. Survey of gene amplifications during prostate cancer progression by high-throughout fluorescence in situ hybridization on tissue microarrays. *Cancer Res* **59**, 803-6 (1999).
27. Khan, S.A., Rogers, M.A., Khurana, K.K., Meguid, M.M. & Numann, P.J. Estrogen receptor expression in benign breast epithelium and breast cancer risk. *J Natl Cancer Inst* **90**, 37-42 (1998).
28. Shoker, B.S., Jarvis, C., Sibson, D.R., Walker, C. & Sloane, J.P. Oestrogen receptor expression in the normal and pre-cancerous breast. *J Pathol* **188**, 237-44 (1999).
29. Li, C. & Wong, W.H. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A* **98**, 31-6 (2001).
30. R Development Core Team. *R: A language and environment for statistical computing*, (2005).
31. Venkatraman, E.S. & Olshen, A. DNAcopy: DNA copy number data analysis. R package version 1.4.0. (2004).
32. Gentleman, R.C. et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80 (2004).
33. Wagner, U. et al. Chromosome 8p deletions are associated with invasive tumor growth in urinary bladder cancer. *Am J Pathol* **151**, 753-9 (1997).

2.2.4. Nuclear Factor 1 B (NFIB) as target gene of the 9p24 amplification in human breast cancer (manuscript)

Manuscript in preparation.

NUCLEAR FACTOR 1 B (NFIB) AS THE TARGET GENE OF THE 9P24 AMPLIFICATION IN HUMAN BREAST CANCER

Christian Ruiz[1], Martin Oeggerli[1], Frederik Holst[2], Philip Stahl[2], Barbara Stalder[1], Khawla Al-Kuraya[3], Luigi M. Terracciano[1], Michel P. Bihl[1], Guido Sauter[2], Ronald Simon[2]

[1]Institute of Pathology, University Hospital Basel, Basel, Switzerland

[2]Dept. of Pathology, University Medical Center Hamburg Eppendorf, Hamburg, Germany

[3]King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia

Running title:

NFIB gene amplification in breast cancer

Abstract

Gene amplification is a common mechanism in cancer to overexpress potential oncogenes. Several amplifications with impact on prognosis have already been described in breast cancer. Using Affymetrix SNP Chip technology, we screened 30 human high grade breast cancers for new small-sized amplifications. We identified two cancers with a 9p24 amplification. One of these two amplicons spanned a small region of 0.8 Mb, comprising only the Nuclear Factor 1 B gene (NFIB). Fluorescence in situ hybridization (FISH) analysis of this gene was applied on a tissue microarray (TMA) containing more than 2000 breast tumors. NFIB gene amplification was found in 5% of all breast cancers analyzed, but with an increased amplification rate of 19% in medullary carcinomas ($p = 0.0007$). Amplification was correlated with high grade ($p < 0.0001$) and high Ki67 proliferation index ($p < 0.0001$). NFIB gene silencing using RNA interference technology in three cell lines expressing this protein led to a reduction in proliferation. This suggests that NFIB stimulates proliferation. Analysis of a breast cancer data set from a public repository (GSE 1456) revealed higher levels of NFIB expression in breast tumors belonging to the subgroup of breast cancers with a basal-like expression. Taken together these findings suggest an important role of NFIB in breast cancer development.

Introduction

The nuclear factor I (NFI) proteins belong to a family of sequence specific DNA-binding proteins. They all bind as homo- and heterodimers via an amino terminal DNA-binding and dimerization domain to the palindromic sequence TGGC/A(N5)GCCA and with lower affinity to the half palindrome (Kruse & Sippel, 1994). Originally, this protein family was identified to be required for replication of adenovirus DNA, but in the last two decades, it has been recognized as a potent transcriptional regulator of many viral (Apt et al., 1993; Kumar et al., 1993; Mink et al., 1992; Plumb et al., 1991) and cellular genes (Cereghini et al., 1987; Rossi et al., 1988). The NFI family comprises four members: NFIA, NFIB, NFIC and NFIX, and expression of several isoforms through alternative splicing have been reported for each of these genes (reviewed in (Gronostajski, 2000)). The expression pattern of the four NFI genes has mainly been studied in mice. They are most highly expressed in lung, liver and heart, and weakly in spleen and testis (Chaudhry et al., 1997). NFIB has been suggested to be essential for normal lung development, since NFIB null mutant mice die early postnatally and display severe lung defects (Grunder et al., 2002). In man, a correlation to cancer has been suggested, since NFIB was found as translocation partner of the high mobility group AT-hook 2 gene (HMGA2) in two pleomorphic adenomas (Geurts et al., 1998) and in two lipomas (Nilsson et al., 2005) resulting in a fusion protein containing the last five amino acids of NFIB. In addition, the NFIB gene was also reported to be part of a 9p23-24 amplicon found in esophageal cancer cell lines (Yang et al., 2000). However, the authors suggest that GASC1, which is localized more than 7 Mb away from the NFIB gene locus, as the major tumor promoting candidate gene of that amplicon. Recently, Zhou et al (Zhou et al., 2006) reported gains on DNA level of the NFIB gene when inactivating p53 and Rb in the prostate epithelium of the mouse.

Gene amplification is a common mechanism for cancer cells to increase the expression levels of proteins that can provide them with a growth or survival advantage. Breast cancer as the leading cancer cause in women is also influenced by gene amplification. Several genes have been found to be

amplified in a fraction of breast cancers including ERBB2 at 17q21, CMYC at 8q24, EGFR at 7p12, AIB1 at 20q13, or CCND1 at 11q13 (Al-Kuraya et al., 2004; Anzick et al., 1997; Pauletti et al., 1996; Simpson et al., 1997; Visscher et al., 1997). Recently, we were able to show amplification of the estrogen receptor alpha gene (ESR1) in breast cancer and its potential impact on breast cancer drug response (manuscript under review, Nature Genetics). For this study, we used the same approach to screen 30 high grad breast cancers for gene copy number changes using DNA microarrays (Affymetrix SNP 10k array). We focused on small-sized amplifications harboring potentially relevant gene classes. This procedure led to the discovery of the NFIB gene amplifications in certain breast cancers. In order to evaluate its relevance in breast cancer, we applied fluorescence in situ hybridization (FISH) on a tissue microarray (TMA) containing samples from more than 2000 breast cancer patients, and determined a proliferation supporting role of this protein by using RNA interference in three cell lines expressing NFIB.

Materials and Methods

Tissues. For DNA and RNA chip analysis experiments, 30 fresh frozen grade 3 breast cancer samples were selected from the frozen tissue archive of the Institute of Pathology, University Hospital Basel, Switzerland. For epidemiological evaluation of NFIB amplifications, our existing Breast Prognosis TMA containing more than 2000 breast tissues was analyzed. The composition of this TMA has been described before (Ruiz et al., 2006).

Affymetrix DNA SNP and Affymetrix RNA chip analysis. DNA and RNA were extracted from three punched tissue cylinders (diameter 0.6mm) from each of the 30 fresh frozen tumor samples according to the manufacturer's instructions of the QIAmp DNA Mini Kit (for DNA) and RNeasy Kit (for RNA) (Qiagen, Hilden Germany). DNA and RNA were further processed as described in the Affymetrix GeneChip Assay manuals (Affymetrix, Santa Clara, CA). After hybridization to the GeneChip Human Mapping 10k Array Xba 142 2.0, respectively to the GeneChip HG U133A 2.0, the microarray chips were washed and stained on an Affymetrix fluidics station. The chips were scanned using the Affymetrix GeneChip scanner 3000.

Data acquisition and analysis of the Affymetrix SNP arrays. Raw data from the scanned SNP arrays was acquired using the GeneChip Operating Software (Affymetrix). Quality of the data was checked as described in the GeneChip Mapping 10K 2.0 Assay Manual (Affymetrix). Only 22 of the samples passed these quality controls and were used for further data analysis. The data files were imported into the DChip software (Li & Wong, 2001), and pre-processing and normalization were performed as described in the user manual. The resulting signal intensities were imported into the R suite (R Development Core Team, 2005). Data were further processed with a home made analysis tool using all measurements between the 25^{th} and 75^{th} percentile as a reference for each DNA spot (code available upon request, C. Ruiz). Data were then modified to fit the input requirements of the DNAcopy package (Venkatraman & Olshen, 2004) of the Bioconductor suite (Gentleman et al., 2004).

The DNAcopy package was used to calculate and visualize segments with similar DNA content. Segments with higher DNA content than the above calculated reference were classified as candidate regions for gene amplification.

Data acquisition and analysis of the Affymetrix RNA arrays. Raw data from the scanned expression arrays was acquired using the GeneChip Operating Software (Affymetrix). In the case of the public available Swedish Breast cancer set (GSE 1456), the corresponding files were downloaded from the Gene Expression Omnibus (GEO) repository (Pubmed, NCBI). Data were imported into the Bioconductor suite and normalized by usage of the RMA algorithm (Irizarry et al., 2003) using the standard settings. The normalized values were filtered using the Genefilter package (Gentleman et al., 2006) of Bioconductor. The filtered data-set was subjected to a t-test to find differential expressed genes. Resulting p-values were manually corrected by using Benjamini and Hochberg false discovery rate multiple testing (Hochberg & Benjamini, 1990). Correspondance analysis and Between Group Analysis were performed using the made4 package (Culhane, 2006).

Fluorescence in situ hybridization (FISH). NFIB FISH analysis of the TMAs was performed with digoxigenated BAC probes (RP11-280O24; RZPD, Berlin, Germany) and the centromeric probe of the chromosome 9 (Vysis, Downers Grove, IL). Hybridization and posthybridization washes were according to the 'LSI procedure' (Vysis). Amplification was defined as presence of at least 2 times more gene copies than centromere signals.

Cell culture. Cells from the different cell lines were cultured in Optimem (Invitrogen, Carlsbad, CA) containing 10% FCS (Amimed, Basel, Switzerland) and 1% Penicillin/Streptamycin (Amimed) in a humidified incubator at 37°C and 5% CO_2.

RNA extraction from cell lines. RNA extraction from cell lines was performed as described in the manufacturer's specifications of the RNeasy kit (Qiagen, Hilden, Germany) including a DNAse I incubation (Qiagen). RNA concentration was determined using a spectrophotometer.

RNA interference. NFIB specific, as well as unspecific RNAi (nsRNA, control) were purchased from Qiagen. Lipofectamine (Invitrogen) was used as transfection reagent according to the manufacturer's protocol. RNAi was applied as described in the Qiagen protocol. Briefly, siRNA transfection was performed in serum- and antibiotic free medium for 24 hours. After transfection, medium was replaced with 10% FCS containing medium. Cells were harvested at regular time points to determine NFIB knockdown and proliferation rate.

Proliferation assay. Cells were harvested using Trypsin/EDTA (1:250) (Amimed) four days after addition of the siRNA. Counting of the cells was performed using the CASY cell counter (Schärfesystem GmbH, Reutlingen, Germany) as described in the manufacturer's handbook. Cell numbers were normalized to cells treated with control nsRNA.

Reverse transcription and polymerase chain reaction (PCR). 2µg of total RNA was reverse transcribed into cDNA using the Transcriptor First Strand cDNA Synthesis kit (Roche Diagnostics, Rotkreuz, Switzerland) following the manufacturer's protocol. RNA extracted from cell lines was reverse transcribed using Oligo-p(dT)15 primers (Roche Diagnostics). For RNA from frozen tissue, random hexamer primers (Roche Diagnostics) were chosen. The cDNA from the reverse transcription was amplified using the AmpliTaq Gold DNA polymerase (Applied Biosystems, Foster City, CA) as described in the AmpliTaq protocol in the presence of 1.0µM of the appropriate primers (MWG Biotech AG, Ebersberg, Germany) listed below (5' → 3'). Separation of the amplification products was obtained by agarose gel electrophoresis (Ambion Inc, Austin, TX).

NFIB: Exon 8 forward:

5' – CCT CCA CCT TCA CCG TTG CCA TTT – 3'

NFIB: Exon 9 reverse:

5' – CCA AGC TAG CCC AGG TAC CAG GA - 3'

NFIB: full length cDNA forward:

5' – GCG GCT CAT GAA ATC CCC ACA – 3'

Real-time PCR. Different amounts of cDNA and different primer concentrations were used for quantitative real-time PCR using the LightCycler FastStart DNA Master^{PLUS} HybProbe Kit (Roche Diagnostics) on the LightCycler instrument (Roche Diagnostics). Normalization was done by using GAPDH or β–Actin concentration as control. Primers and hybridization probes used are listed below (TIB MOLBIOL, Berlin, Germany):

Forward primers:

GAPDH: 5' – GAA GGT GAA GGT CGG AGT C – 3'

NFIB: 5' – TCA CTC AGG ATG AAT TTC ACC – 3'

ACTB: 5' – AGC CTC GCC TTT GCC GA – 3'

Reverse primers:

GAPDH: 5' – GAA GAT GGT GAT GGG ATT TC – 3'

NFIB: 5' – TGC CCA CTT CTG TTT GAT T– 3'

ACTB: 5' – CTG GTG CCT GGG GCG – 3'

Hybridization probes Fluorescein part:

GAPDH: 5' – AGG GGT CAT TGA TGG CAA CAA TAT CCA –FL –3'

NFIB: 5' – AAG CTC ATC TTT CAG TGC TCT TTC TTC –FL –3'

ACTB: 5' – TTG CAC ATG CCG GAG CCG TTG –FL –3'

Hybridization probes LightCycler part:

GAPDH: 5' – LC Red640-TTT ACC AGA GTT AAA AGC AGC CCT GGT G –3'

NFIB: 5' – LC Red640-TCC TTT GAC ATT CGC TTC TCA TGC T –3'

ACTB: 5' – LC Red640-CGA CGA CGA GCG CGG CGA TAT C –3'

Whole cell protein lysates and subcellular fractionation. Whole cell protein lysates were created by lysing the cells in SDS lysis buffer (50mM Tris pH 6.8, 2% SDS, 5% Glycerol, 5mM EDTA) including protease inhibitors (Complete Mini, Roche Diagnostics) directly on the plate. For subcellular fractionation two distinct buffers were used, first the cytosol fraction was extracted using a low salt buffer (20mM Hepes pH 7.9, 10mM KCl, 0.1mM $NaVO_4$, 1mM EDTA, 1mM EGTA, 0.2% NP-40, 10% Glycerol). Then, the nuclear fraction was extracted using a high salt buffer (420mM NaCl, 20mM

Hepes pH 7.9, 10mM KCl, 0.1mM $NaVO_4$, 1mM EDTA, 1mM EGTA, 20% Glycerol). Both buffers were supplemented with protease inhibitors (Complete Mini, Roche Diagnostics). After removing the insoluble material by centrifugation, the lysates were shock-frozen in liquid nitrogen and stored at -80°C.

Western blot analysis. Concentration of protein lysates was determined using the RC DC Protein Assay (Bio-Rad Laboratories, Hercules, CA) and equal amounts of proteins (10ug or 15µg) were resolved by usage of 4-20% precast Tris-glycine gels (Invitrogen) and transferred to Immobilon-P (Millipore) membranes. Membranes were blocked with 5% blocking solution (5% dry, non-fat skim milk powder in PBS) overnight at 4°C. Incubation with the rabbit polyclonal anti-NFIB antibody (ABCAM, Cambridge, UK) was done for 2 hours at room temperature; Incubation with Anti-GAPDH antibody (ABCAM) for one hour. Detection was performed with an anti-rabbit or an anti-mouse horseradish peroxidise (HRP)-conjugated secondary antibody (Jackson Immunoresearch, West Baltimore, PA) by enhanced chemiluminescence (ECL and ECL Plus, Amersham).

Cell line TMA. Cells were cultured as described until ten 75cm^2 plates were 70-80% confluent. Cells were harvested, washed and centrifuged. The cell pellet was fixed in formalin for 24 hours and then placed into a Cytoblock cassette (Thermo Shandon GmbH, Frankfurt, Germany) and further processed as described in the Cytoblock Cell Preparation System Kit. Briefly, the cassette was incubated in 70% ethanol, dewatered and embedded into liquid paraffin. One day later, the paraffin block was punched. The cell line TMA was created as described (Simon et al., 2004). Briefly, punched cylinders were brought into a recipient paraffin block with a semiautomatic robotic precision instrument. Four micrometer sections of the resulting TMA block were transferred to an adhesive coated slide system (Instrumedics, Hackensack, NJ).

Statistics. Data from the TMA evaluation was collected in Excel (Microsoft Corporation, Redmond, WA) and transferred to JMP 5.1 software (SAS, Cary, NC). T-tests, analysis of variance tests, Chi-Square tests and survival analysis (Kaplan-Meier) were all performed using the JMP software.

Results

DNA copy number analysis using SNP arrays. Twenty-two of the thirty SNP Affymetrix arrays passed the rigorous quality control according to the criteria described in the Affymetrix and in the DChip manual. After thorough analysis of all of the samples (data not shown), two samples (sample 26 and sample 27) showed a 9p24 amplification (Figure 1). Whereas the amplicon of sample 27 spanned a region of about 3.5Mb comprising 13 genes, the clear-cut amplified region of sample 26 was restricted to a small region of 0.7Mb involving only the NFIB gene. This could be observed using the DChip analysis method (Figure 1A) as well as using the DNAcopy visualization (Figure 1B). Successful verification of these two samples was done by applying a FISH probe against the NFIB gene (one example shown in Figure 1C). RNA expression analysis of these samples on the Affymetrix RNA arrays revealed high concordance between amplification on DNA level and RNA overexpression (Table 1). Differential expression of 9p24 amplified samples against 9p24 normal samples revealed several statistically significant differences (data not shown), but the NFIB gene was the only gene of that list localized in the above mentioned region (Table 1). Both, the DNA data as well as the RNA data concerning this region strongly suggest that the NFIB gene is the so-called candidate gene of the 9p24 amplicon.

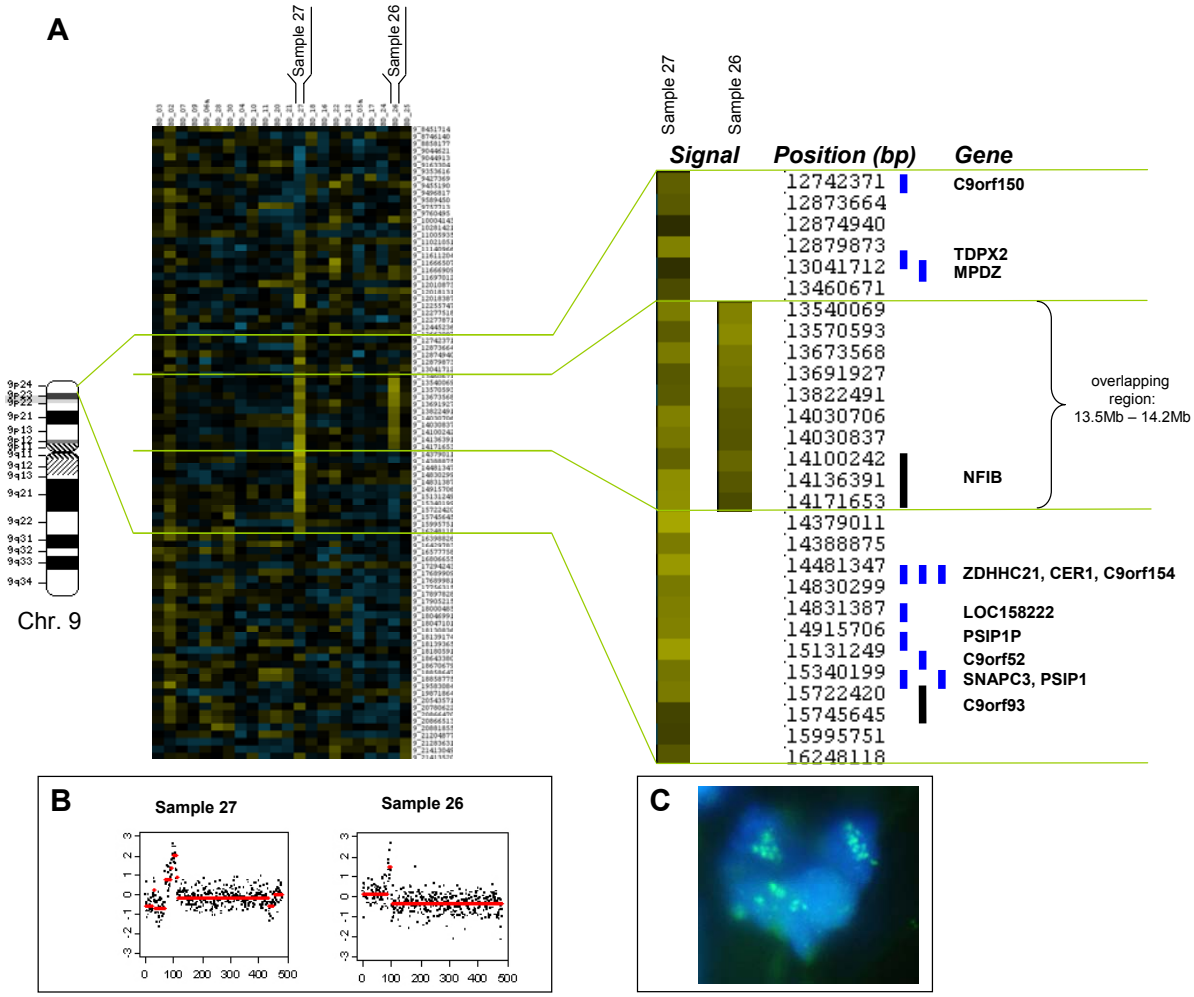


Figure 1. SNP array DNA copy number analysis. A) shows a heatmap representing the DNA copy numbers in colors (yellow: high DNA copy number). On the right, the amplified region is shown in detail with all the genes that are localized in that region. B) shows DNA copy number visualization along the whole chromosome 9. C) shows the verification of an amplified sample by FISH analysis. Green fluorescent points show NFIB gene copies.

	C9orf150	MPDZ	NFIB	CER1	SNAPC3	PSIP1
Normal samples	-	-	1.0	-	-	1.0
Sample 26	-	-	**+3.6**	-	-	**-1.7**
Sample 27	-	**+3.3**	**+4.9**	-	**+1.7**	**+4.7**
DE: Normal vs Amplified	*No*	*No*	*p < 0.0001*	*No*	*No*	*No*

Table 1. Statistical analysis of the Affymetrix RNA expression arrays. Only genes are listed that are localized in the region shown in Figure 1. Present values are normalized to 1. Absent genes are marked with a hyphen. Only the NFIB transcript is statistically significantly elevated in amplified samples. DE: Differential expression between amplified and normal samples.

NFIB gene amplification analysis on 2200 breast cancers. In order to identify the role of NFIB gene amplification in a breast cancer population, our Breast Prognosis TMA containing 2200 breast tumors was subjected to NFIB gene FISH analysis. FISH analysis was successful in 1608 of 2197 (73.2%) arrayed breast cancer samples (Table 2). FISH analysis failed in 589 cases either due to lack of tumor cells in the spot or because of FISH-related problems, like weak hybridization and high background. Increased copy number of NFIB (at least twofold more NFIB gene copies than centromere 9 signals) was found in 5% of all breast cancers analyzed. Most of these amplified cases had moderate absolute copy numbers ranging between 4 and 10 NFIB gene copies. The associations with tumor phenotype are summarized in Table 2. Amplification rate of NFIB was not evenly distributed among the different histological subtypes in breast cancer. Medullary carcinomas showed a particularly high rate of NFIB amplification (19%, $p = 0.0007$). A similar rate was achieved by the apocrine tumors (22%, $p = 0.0908$), but the low number of analyzable samples of this group ($n = 9$) diminished the significance of this result. NFIB gene amplification was strongly associated with tumor grade 3 ($p < 0.0001$), but no association was found with tumor or nodal stages. NFIB gene amplification was also significantly associated with tumor cell proliferation ($p < 0.0001$), as determined by the use of the Ki67 labeling index (LI) (Table 2B). This was also true by analyzing only grade 3 breast cancers ($p = 0.0082$). Higher proliferation rates were also found in the grade 1 and grade 2 subgroups, but due to the low numbers of amplified cases in those groups ($n = 4$ for grade 1, $n = 13$ for grade 2), no significant association was found. The impact of NFIB gene amplification on patient survival was also analyzed (Figure 2). Patients carrying a NFIB amplification showed a worse prognosis than normal NFIB patients ($p = 0.0278$), but this association seemed dependent of the grade, since the same analysis with only grade 3 patients did not show an effect ($p = 0.4764$).

A

			NFIB amplification			
		on array (n)	analyzable (n)	Increased copy number (n)	Increased copy number	p
Breast Cancers	all	2197	1608	76	5%	
Histology	ductal carcinoma	1552	1099	60	5%	
	lobular carcinoma	312	212	2	1%	0.0070*
	mucinous carcinoma	69	45	0	0%	
	medullary carcinoma	58	48	9	19%	0.0007*
	tubular carcinoma	56	33	0	0%	
	cribriform carcinoma	65	49	0	0%	
	papillary carcinoma	30	23	1	4%	
	apocrine carcinoma	15	9	2	22%	0.0908*
	others	79	38	1	3%	
pT stage	pT1	820	541	22	4%	0.6521
	pT2	1023	767	44	6%	
	pT3	124	100	3	3%	
	pT4	242	170	6	4%	
Nodal stage	pN0	950	689	30	4%	0.4239
	pN1	793	578	33	6%	
	pN2	121	83	2	2%	
BRE grade	G1	545	378	4	1%	<0.0001
	G2	844	565	13	2%	
	G3	655	513	51	10%	

*versus ductal carcinoma

B

NFIB and Ki67 LI

	number	mean Ki67 LI	Lower 95%	Upper 95%	p-value
NFIB amplified	63	41.35	37.66	45.04	< 0.0001
NFIB normal	1228	28.02	27.18	28.85	

NFIB and Ki67 LI on Grade 3 tumors

	number	mean Ki67 LI	Lower 95%	Upper 95%	p-value
NFIB amplified	41	45.02	40.44	49.61	0.0082
NFIB normal	378	38.5	36.99	40.01	

Table 2. Statistical analysis of the Breast Prognosis TMA comprising more than 2000 breast tumors. A) Statistical summaries. B) Influence of the NFIB amplification on the Ki67 proliferation index.

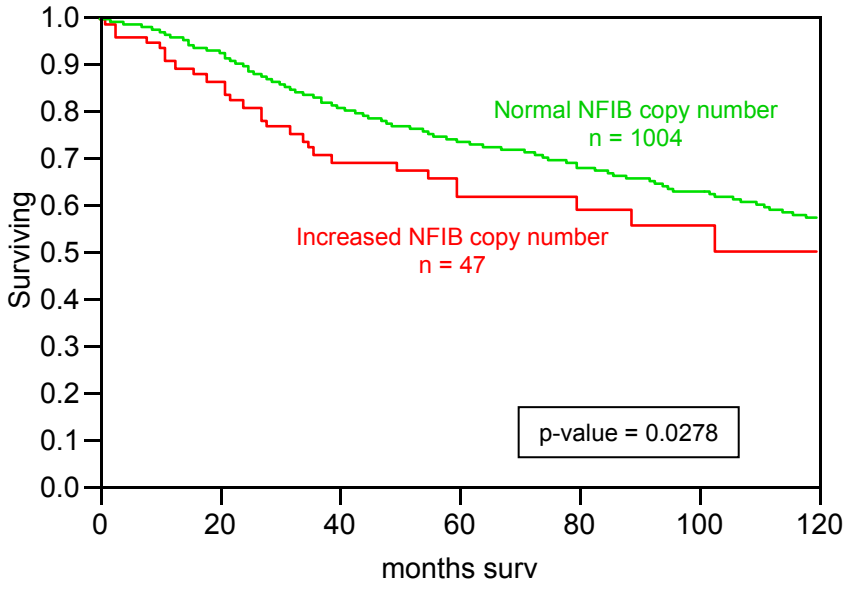


Figure 2. Survival analysis of the breast samples on the breast prognosis TMA. Breast tumors with elevated number of NFIB gene copies, show a worse survival prognosis during a 10-year-interval.

NFIB gene amplification analysis on 120 cell lines. In order to find cell lines carrying NFIB gene amplifications, our cell line TMA comprising more than 120 different human cell lines was subjected to FISH analysis with a probe against the NFIB gene. An overview of the cell lines of this TMA is listed as supplementary information. The human thyroid sarcoma cell line named S-117 was the only cell line found carrying such an amplification (i.e. more than two times more NFIB gene copies than chromosome 9 centromeres). None of the analyzed breast cell lines (SKBR3, T47D, MCF7, MDA-MB-435, BT-549, BT-474, ZR-75-1) showed a NFIB gene amplification.

NFIB gene knockdown leads to reduced proliferation levels. In order to further investigate the previously mentioned association between the NFIB gene amplification and the Ki67 LI, the effect of NFIB gene knockdown on cell proliferation was studied (Figure 3). First of all, two breast cell lines expressing the NFIB gene (T47D and MCF7, Figure 3A) and the cell line with the NFIB gene amplification (S-117, Figure 3A) were chosen for this analysis. After gene knockdown using siRNA, RNA expression levels dropped down to 20% (Figure 3B). The effect of this knockdown could also be observed three and four days later on protein levels (Figure 3C). The knockdown of NFIB on RNA and on protein level led to reduced cell proliferation (Figure 3D). The largest effect could be detected in the cell line MCF7, where a reduction of proliferation by 64% was achieved.

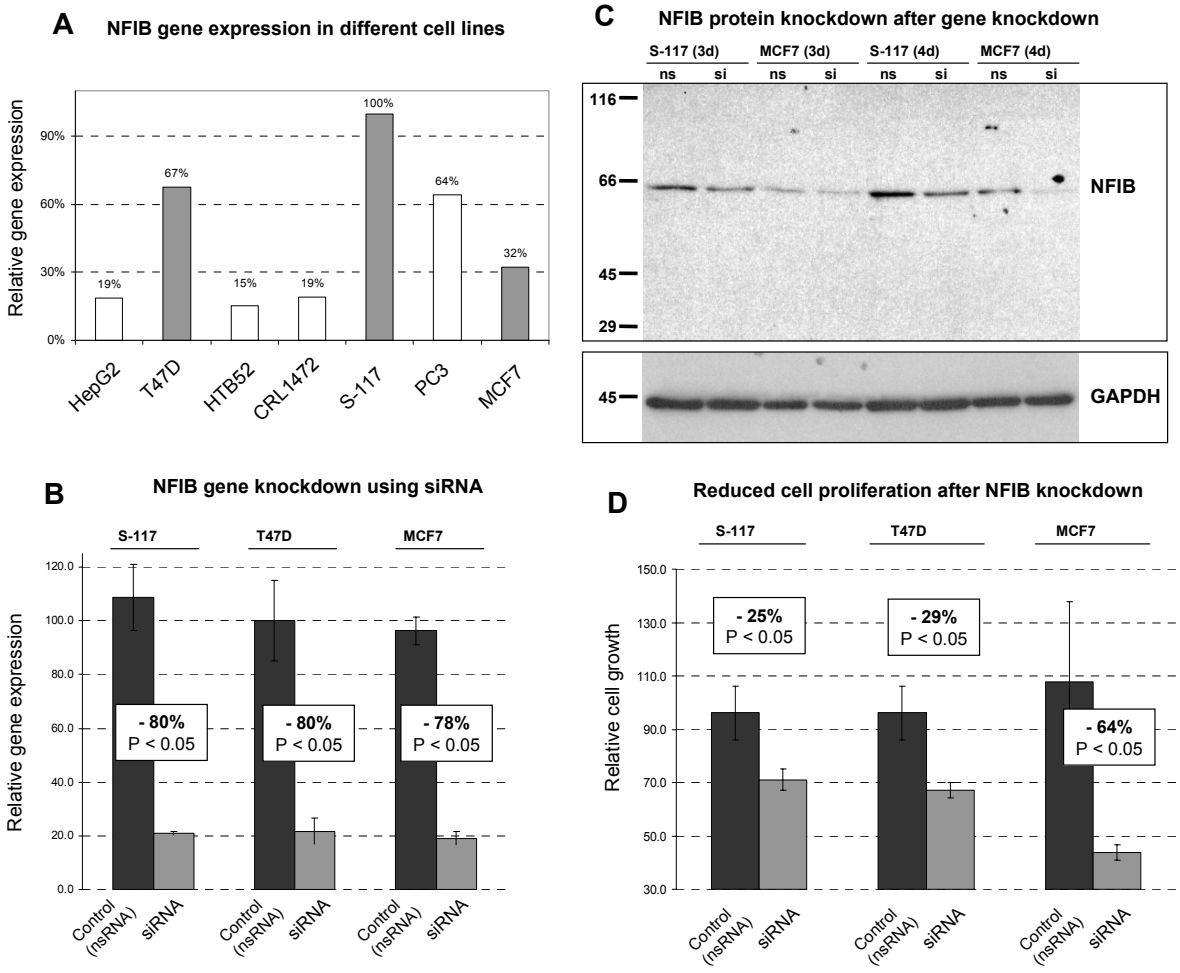


Figure 3. NFIB gene expression. A) shows NFIB RNA levels of different tumor cell lines. Grey ones were chosen for further analysis. B) shows successful NFIB gene expression knockdown using siRNA. C) shows NFIB protein knockdown on whole cell lysates three and four days after transfection. D) shows attenuated cell growth of cells with reduced NFIB expression.

NFIB protein expression. Whole cell lysate western blot analysis with anti-NFIB revealed only one band at 60-65kDa (Figure 3C). In the database Swiss-Prot three isoforms for the human NFIB protein are described. A suggested "normal" isoform with a molecular weight (MW) of 47 kDa (Swiss-Prot: NFIB_HUMAN), a smaller one with a MW of 22 kDa, which was exclusively described in 1997 (Liu et al., 1997) and a third one without any further description or reference are listed. In order to analyze the different NFIB isoforms in our cell lines, we performed subcellular fractionation and western blot analysis with the only commercially available antibody (see materials and methods for description). The main isoform detected in all of the cell lines analyzed had a MW of around 60-65 kDa and was present in both, the cytosolic and in the nuclear fraction (Figure 4A-C). The 47 kDa isoform could only be detected by massive overloading of the gel with nuclear material (Figure 4C), suggesting that this isoform is only present in low amounts. Interestingly, the cell line S-117 showed an additional highly expressed protein isoform in the nuclear fraction at 55 kDa (Figure 4B and C). Using database searches (NCBI-Entrez, Expasy/TrEMBL and Enseml), two cDNA-isoforms for NFIB whose translational products have calculated MWs of 54.9kDa (TrEMBL: Q5VW27) or 55.2kDa (Q5VW29), were found. The exon structures of these two isoforms and of the "normal" isoform are shown in Figure 4D. Reverse-transcription PCR followed by nucleic acid sequencing proved the existence of Q5VW29 (55.2kDa) and excluded the presence of the Q5VW27 (54.9kDa) isoform (Figure 4E).

The highly expressed protein isoform at 60-65kDa (Figure 4A) belongs either to the predicted protein product Q5VW30 (62.3kDa, TrEMBL) or to the isoform Q5VW26 (63.4kDa, TrEMBL). Using full length cDNA cloning, followed by nucleic acid sequencing, we were able to show that the Q5VW30 isoform is present in all of the cell lines analyzed, but were not able to exclude the existence of the isoform Q5VW26 (data not shown).

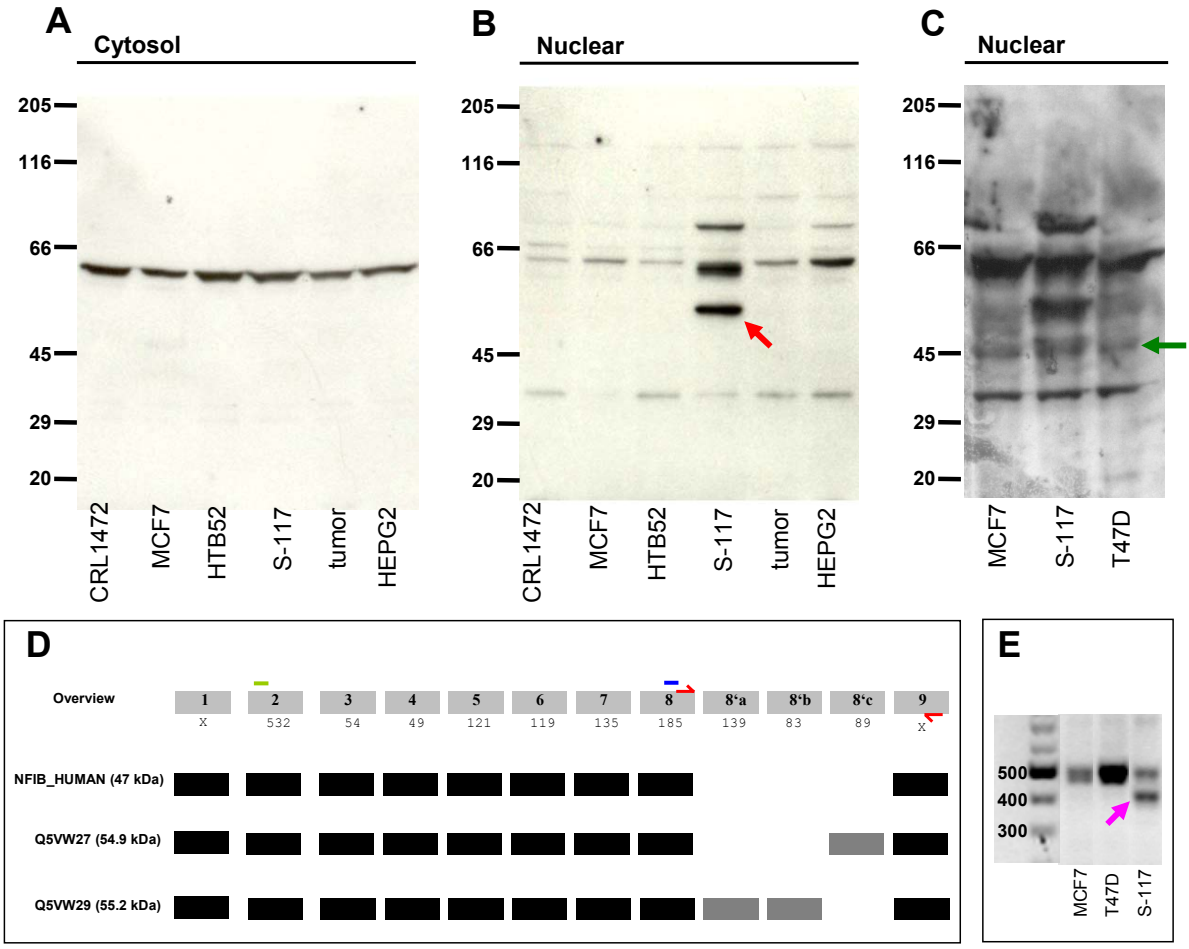
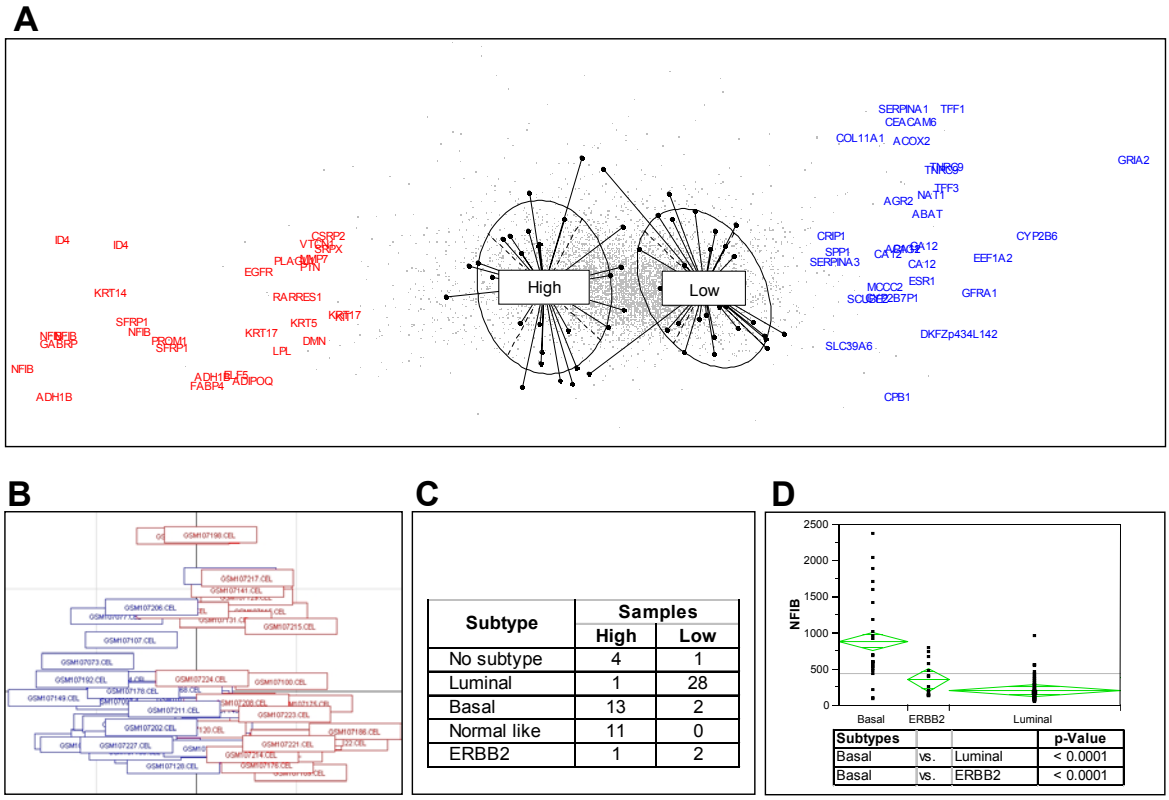


Figure 4. NFIB protein expression pattern. A) shows NFIB protein expression in the cytosol. B) The amplified cell line S-117 shows a distinct nuclear NFIB expression pattern (red arrow). C) Only by overloading the gel with nuclear extracts, the normal isoform at 47kDa becomes detectable (green arrow). D) Overview of the exon structure of the isoforms mentioned above: NFIB_HUMAN as the "normal" isoform and Q5VW27 and Q5VW29 as the two isoforms whose translational products have a calculated MW of around 55 kDa. Primers are shown as red bars. The blue bar shows the position of the epitope of the antibody used. The green bar shows the DNA binding site of NFIB. E) Gel analysis of rt-PCR products generated using the primers from D. Pink arrow shows that only the S-117 expresses the NFIB isoform whose translational product has a calculated MW of around 55 kDa. Nucleic acid sequencing of this band showed that this NFIB isoform is the Q5VW29 (55.2 kDa).

RNA expression analysis of NFIB expressing breast tumors. In order to find a role of the NFIB gene in breast cancer, we used a public Affymetrix RNA expression data set (GSE1456, GEO public repository) comprising expression arrays of 159 breast cancer patients. From this data set, we chose the 30 samples with the highest NFIB expression values and 33 samples with none or very low NFIB expression and subjected these two groups to differential expression analysis. Surprisingly, a huge number of differentially expressed genes (approximately 2500; 1600 with an absolute fold change of more than 1.4) after rigorous statistical analysis were detected. (The 500 most significant differentially expressed are listed in the supplemental section.) In order to visualize the distribution of the samples, an unsupervised correspondence analysis was performed. Colouring of the samples according to their NFIB expression status revealed the existence of two groups with only few crossing points (Figure 5B). Additionally, a supervised Between Group Analysis (BGA) using the NFIB expression status as classifier was performed, leading to a similar result (Figure 5A). The so-called topgenes, i.e. the genes that best separated the two groups, were selected and compared to the 500 gene list mentioned above (supplemental data, Table S2 and S3). Almost all of the top ranked genes were present in both lists (92% overlapping in the top 100 genes). Several cancer-relevant genes were found to be positively (e.g. ID4, EGFR, PDGFRA, KIT) or negatively (e.g. ESR1) correlated with NFIB.

Additional information provided with the public breast cancer data set used is the so-called genetic subtype first introduced by Perou and Sorlie (Perou et al., 2000; Sorlie et al., 2001). This information has been recently assigned to this data set (Calza et al., 2006). Application of this sample information to our two groups, revealed that 28 of the 33 samples (85%) with no or low NFIB expression had been characterized as luminal. But only 1 of 30 samples with a high NFIB gene expression belonged to that subtype (Figure 5C).

This suggested a major role for NFIB in assisting the separation of the breast cancer samples into genetic subclasses. We analyzed the impact of the NFIB gene expression on the different subtypes using the full data set (159 breast cancer samples). The result was striking: NFIB gene expression was significantly different between the three groups (Figure 5D), showing highest expression levels in the basal-like and lowest levels in the luminal group.



Subtype	Samples	
	High	Low
No subtype	4	1
Luminal	1	28
Basal	13	2
Normal like	11	0
ERBB2	1	2

Subtypes			p-Value
Basal	vs.	Luminal	< 0.0001
Basal	vs.	ERBB2	< 0.0001

Figure 5. NFIB gene expression analysis on the public breast cancer data set GSE1456. A) Between Group Analysis of the 63 selected samples using NFIB expression status as classifier leads to several genes which helps to separate the two groups. Programming code: courtesy of F. Baty. B) Unsupervised correspondence analysis of the 63 selected samples. Colouring according to the NFIB expression (red: high NFIB expression, blue: low NFIB expression) revealed two groups. C) Genetic subtypes of the 63 samples selected. Only one luminal sample shows high NFIB gene expression. D) NFIB gene expression analysis on all of the 159 breast cancer samples revealed distinct expression levels of the different genetic subtypes. All Pairs Tukey HSD was used.

Discussion

In this study, we have demonstrated the existence of NFIB gene amplifications in human breast cancer. This 9p24 amplification has never been described in breast cancer before. Similar to our recently published work about the ESR1 gene amplification, the NFIB gene amplification was first detected using Affymetrix SNP microarrays in only two cases, including one sample where the amplicon was limited to a size of less than 1 Mb, only comprising the NFIB gene.

The use of a FISH probe for the NFIB gene allowed us to validate our findings on a large Breast Prognosis TMA comprising more than 2000 breast cancers. The overall NFIB gene amplification rate of 5% was similar to the MDM2 (5.7%) and CMYC (5%) amplification rates, previously defined using the same TMA. The correlation of the NFIB amplified samples with a higher proliferation index (Ki67 LI) suggested a potential role for NFIB as a proliferation supporting protein in breast cancer. This finding was confirmed by the results obtained by the RNA interference experiments. Knockdown of NFIB RNA, followed by NFIB protein reduction was accompanied by reduced levels of proliferation. The most impressive effect was obtained in the cell line MCF7, which showed a reduced proliferation of more than 60%. This may be explained by the observation that this cell line had the lowest NFIB expression levels under normal culture conditions. RNA knockdown in MCF7 cells may cause NFIB levels to fall below a threshold level required for continued proliferation. Since our RNA interference approach was designed to affect all of the described NFIB isoforms, it is not clear whether all of the isoforms influenced cell proliferation to the same extent. Western blot analysis revealed that the most prominent NFIB isoform in the cell lines tested, was by far not the so-called "normal" isoform with a MW of 47 kDa (SwissProt: NFIB_HUMAN), but an isoform with a MW around 63 kDa. We were able to show the existence of the "predicted" NFIB isoform Q5VW30 with a calculated MW of 62.3 kDa, but further experiments are necessary to exclude the presence of other NFIB isoforms with similar MW. In addition, we were able to confirm that the NFIB protein product with a size of approximately 53 kDa in the cell line S-117 belongs to the "predicted" isoform Q5VW29. We

suggest an important role for this variant since it was found in the cell line with an amplified NFIB gene and the protein product could only be detected in the nuclear fraction of the cells, in contrast to the other main isoform which was prominent in the cytosol.

The availability of a public Affymetrix expression data set comprising 159 breast cancers allowed us to study the influence of NFIB gene expression in breast cancer. The result was intriguing: data analysis of samples with highest, respectively lowest levels of NFIB revealed a clear cut separation of the breast tumors. Inclusion of additional public information describing the tumors analyzed revealed that NFIB was differentially expressed between the different molecular expression groups (luminal, basal and ERBB2) introduced by Perou and Sorlie (Perou et al., 2000; Sorlie et al., 2001). Breast cancer samples belonging to the basal group were characterized by the authors as ER and ERBB2 negative, luminal samples as ER positive and samples belonging to the ERBB2 group as Her-2 positive. Here we show that NFIB is highly expressed in the samples belonging to the basal group and lowest in the luminal breast cancers ($p < 0.0001$). An equivalent result was also obtained by investigating the top genes which best divided the NFIB high group from the NFIB low group (Figure 5A and Supplemental tables S2 and S3). Several genes which had previously been defined to be characteristic for the basal-like group appeared as top ranked genes in the NFIB high group (e.g. KRT5 and KRT17). On the other hand, several genes which were said to be responsible for the characterization of the luminal group were top ranked in the NFIB low group, as for example the ESR1 gene. This finding was not surprising since we could show that most of the selected NFIB low samples (28 of 33, Figure 5C) were previously classified as luminal and were therefore ER positive. Similar observations were also obtained on DNA amplification levels on the Breast Prognosis TMA. NFIB gene amplification showed its highest rate on medullary carcinomas (19%, $p = 0.0007$). Since it has been shown that amplified genes tend to be overexpressed, these results are consistent with a recent publication which classified the medullary carcinomas as basal breast cancers (Bertucci et al., 2006). These findings are further supported by the ESR1 FISH analysis on the same array, which showed by far its lowest rate on medullary carcinomas (2%, $p < 0.0001$) (manuscript under review,

Nature Genetics). This is again in accordance with previous data, since ER positive samples are classified as luminal and these show lowest NFIB expression levels.

Since the NFIB gene is not only overexpressed in the basal group of breast cancers, but also amplified in a considerable fraction of them, we suggest that this gene is relevant for the development of these breast cancers. Our data from the NFIB gene knockdown indicate a role for NFIB to support tumor cell growth.

References

Al-Kuraya K, Schraml P, Torhorst J, Tapia C, Zaharieva B, Novotny H, Spichtin H, Maurer R, Mirlacher M, Kochli O, Zuber M, Dieterich H, Mross F, Wilber K, Simon R and Sauter G. *Prognostic relevance of gene amplifications and coamplifications in breast cancer* (2004). *Cancer Res*, 64**,** 8534-40.

Anzick SL, Kononen J, Walker RL, Azorsa DO, Tanner MM, Guan XY, Sauter G, Kallioniemi OP, Trent JM and Meltzer PS. *AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer* (1997). *Science*, 277**,** 965-8.

Apt D, Chong T, Liu Y and Bernard HU. *Nuclear factor I and epithelial cell-specific transcription of human papillomavirus type 16* (1993). *J Virol*, 67**,** 4455-63.

Bertucci F, Finetti P, Cervera N, Charafe-Jauffret E, Mamessier E, Adelaide J, Debono S, Houvenaeghel G, Maraninchi D, Viens P, Charpin C, Jacquemier J and Birnbaum D. *Gene expression profiling shows medullary breast cancer is a subgroup of basal breast cancers* (2006). *Cancer Res*, 66**,** 4636-44.

Calza S, Hall P, Auer G, Bjohle J, Klaar S, Kronenwett U, Liu ET, Miller L, Ploner A, Smeds J, Bergh J and Pawitan Y. *Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients* (2006). *Breast Cancer Res*, 8**,** R34.

Cereghini S, Raymondjean M, Carranca AG, Herbomel P and Yaniv M. *Factors involved in control of tissue-specific expression of albumin gene* (1987). *Cell*, 50**,** 627-38.

Chaudhry AZ, Lyons GE and Gronostajski RM. *Expression patterns of the four nuclear factor I genes during mouse embryogenesis indicate a potential role in development* (1997). *Dev Dyn*, 208**,** 313-25.

Culhane AC. (2006).

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, et al. *Bioconductor: open software development for computational biology and bioinformatics* (2004). *Genome Biol*, 5**,** R80.

Gentleman RC, Carey VJ and Huber W. (2006).

Geurts JM, Schoenmakers EF, Roijer E, Astrom AK, Stenman G and van de Ven WJ. *Identification of NFIB as recurrent translocation partner gene of HMGIC in pleomorphic adenomas* (1998). *Oncogene*, 16**,** 865-72.

Gronostajski RM. *Roles of the NFI/CTF gene family in transcription and development* (2000). *Gene*, 249**,** 31-45.

Grunder A, Ebel TT, Mallo M, Schwarzkopf G, Shimizu T, Sippel AE and Schrewe H. *Nuclear factor I-B (Nfib) deficient mice have severe lung hypoplasia* (2002). *Mech Dev*, 112**,** 69-77.

Hochberg Y and Benjamini Y. *More powerful procedures for multiple significance testing* (1990). *Stat Med*, 9**,** 811-8.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U and Speed TP. *Exploration, normalization, and summaries of high density oligonucleotide array probe level data* (2003). *Biostatistics*, 4**,** 249-64.

Kruse U and Sippel AE. *Transcription factor nuclear factor I proteins form stable homo- and heterodimers* (1994). *FEBS Lett*, 348**,** 46-50.

Kumar KU, Pater A and Pater MM. *Human JC virus perfect palindromic nuclear factor 1-binding sequences important for glial cell-specific expression in differentiating embryonal carcinoma cells* (1993). *J Virol*, 67**,** 572-6.

Li C and Wong WH. *Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection* (2001). *Proc Natl Acad Sci U S A*, 98**,** 31-6.

Liu Y, Bernard HU and Apt D. *NFI-B3, a novel transcriptional repressor of the nuclear factor I family, is generated by alternative RNA processing* (1997). *J Biol Chem*, 272**,** 10739-45.

Mink S, Hartig E, Jennewein P, Doppler W and Cato AC. *A mammary cell-specific enhancer in mouse mammary tumor virus DNA is composed of multiple regulatory elements including binding*

sites for CTF/NFI and a novel transcription factor, mammary cell-activating factor (1992). *Mol Cell Biol*, 12**,** 4906-18.

Nilsson M, Panagopoulos I, Mertens F and Mandahl N. *Fusion of the HMGA2 and NFIB genes in lipoma* (2005). *Virchows Arch*, 447**,** 855-8.

Pauletti G, Godolphin W, Press MF and Slamon DJ. *Detection and quantitation of HER-2/neu gene amplification in human breast cancer archival material using fluorescence in situ hybridization* (1996). *Oncogene*, 13**,** 63-72.

Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO and Botstein D. *Molecular portraits of human breast tumours* (2000). *Nature*, 406**,** 747-52.

Plumb M, Fulton R, Breimer L, Stewart M, Willison K and Neil JC. *Nuclear factor 1 activates the feline leukemia virus long terminal repeat but is posttranscriptionally down-regulated in leukemia cell lines* (1991). *J Virol*, 65**,** 1991-9.

R Development Core Team. (2005).

Rossi P, Karsenty G, Roberts AB, Roche NS, Sporn MB and de Crombrugghe B. *A nuclear factor 1 binding site mediates the transcriptional activation of a type I collagen promoter by transforming growth factor-beta* (1988). *Cell*, 52**,** 405-14.

Ruiz C, Seibt S, Al Kuraya K, Siraj AK, Mirlacher M, Schraml P, Maurer R, Spichtin H, Torhorst J, Popovska S, Simon R and Sauter G. *Tissue microarrays for comparing molecular features with proliferation activity in breast cancer* (2006). *Int J Cancer*, 118**,** 2190-4.

Simon R, Mirlacher M and Sauter G. (2004). *Molecular diagnosis of cancer*. Humana Press: Tatowa.

Simpson JF, Quan DE, O'Malley F, Odom-Maryon T and Clarke PE. *Amplification of CCND1 and expression of its protein product, cyclin D1, in ductal carcinoma in situ of the breast* (1997). *Am J Pathol*, 151**,** 161-8.

Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lonning P and Borresen-Dale AL. *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications* (2001). *Proc Natl Acad Sci U S A*, 98**,** 10869-74.

Venkatraman ES and Olshen A. (2004).

Visscher DW, Wallis T, Awussah S, Mohamed A and Crissman JD. *Evaluation of MYC and chromosome 8 copy number in breast carcinoma by interphase cytogenetics* (1997). *Genes Chromosomes Cancer*, 18**,** 1-7.

Yang ZQ, Imoto I, Fukuda Y, Pimkhaokham A, Shimada Y, Imamura M, Sugano S, Nakamura Y and Inazawa J. *Identification of a novel gene, GASC1, within an amplicon at 9p23-24 frequently detected in esophageal cancer cell lines* (2000). *Cancer Res*, 60**,** 4735-9.

Zhou Z, Flesken-Nikitin A, Corney DC, Wang W, Goodrich DW, Roy-Burman P and Nikitin AY. *Synergy of p53 and Rb Deficiency in a Conditional Mouse Model for Metastatic Prostate Cancer* (2006). *Cancer Res*, 66**,** 7889-98.

Supplemental Data

Cell lines on the cell line tissue microarray

Cell line name	ATCC-Nr.	Organ	Cell line name	ATCC-Nr.	Organ
SKBR-3	HTB-30	breast	A 549	CCL-185	lung
T47D	HTB-133	breast	NCI-H460	HTB-177	lung
MCF 7	HTB-22	breast	NCI-H226	CRL-5826	lung
MDA-MB-435	HTB-129	breast	NCI-H522	CRL-5810	lung
BT-549	HTB-122	breast	HOP-62		lung
BT-474	HTB 20	breast	EKVX		lung
MBA-MD-415		breast	NCI-H322M		lung
HBL100	HTB-124	breast	NCI-H23	CRL-5800	lung
ZR-75-1	CRL-1500	breast	MRC-5	CCL-171	lung
MBA-MB-453		breast	NCI-H82	HTB-175	lung
BT-474	HTB-20	breast	NCI-H51OA		lung
SK-OV-3	HTB-77	ovary	MALME-3M		lung metastasis
OVCAR-3		ovary	8505C 219		thyroid carcinoma
OVCAR-8		ovary	CAL_62		thyroid carcinoma
OVCAR-4		ovary	BHT-101		thyroid carcinoma
OVAR-5		ovary	ONCO-DG-1		thyroid carcinoma
HELA	CCL-2	cervix	B-C PAP		thyroid carcinoma
786-O	CRL-1932	kidney	ML-1		thyroid carcinoma
CAK I		kidney	8305 c		thyroid carcinoma
A 498	HTB-44	kidney	S-117		thyroid gland
ACHN	CRL-1611	kidney	SK-MEL-2	HTB-68	skin
786-O	CRL-1932	kidney	SK-MEL-5	HTB-70	skin
Caki-2		kidney	A 375	CRL-1619	skin
293	CRL-1573	kidney	A 431	CRL-2592	skin
COS 1	CRL-1650	kidney	UACC 257		malignant melanoma
WS-1	CRL-2029	kidney	UACC-62		malignant melanoma
TK10		renal	MEL-HO		melanoma
UO-31		renal	SM		melanoma
J82	HTB-1	urinary	M14		amelanotic melanoma
SCaBER	HTB-3	urinary	T98G	CRL-1690	brain
T24	HTB-4	urinary	U-138 MG	HTB-16	brain
HT-1376	CRL-1472	urinary	A172	CRL-1620	brain
5637	HTB-9	urinary	U87-MB	HTB-14	brain
TCCSUP	HTB-5	urinary	SNB-19		cns
RT-112		urinary	SF 268		cns
KU-19-19		urinary	SF-539		cns
CRL 7930		urinary	SF-295		cns
DLD-1	CCL-221	colon	U 251		cns/ glioma
COLO 320 DM	CCL-220	colon	LN 229		glioblastoma
COLO 201	CCL-224	colon	LN 401		glioblastoma
RKO	CRL-2577	colon	DBTRG		glioblastoma
SW-480	CCL-228	colon	GAMG p6		glioma
HAT-29	HTB-38	colon	LN 405 p 9		astrocytom
SK-CO-1	HTB-39	colon	SJCRH-30WCB		neuroblastom
SW-403	CCI-230	colon	OCI-LY19		bone
HCT 116	CCL-247	colon	MOLT-4	CRL-1582	peripheral blood
HCT 15	CCL-225	colon	IM-9	CCL-159	peripheral blood
SW-620	CCL-227	colon	CRO-AP 3		b-cell-lymphoma
HT29	HTB-38	colon	IGR1		groin lymphnode
COLO 205	CCL-222	colon	SR	CRL-2262	immunoblastic large cell lymphoma
HCL-2998		colon	COLO 849		right axillary lymphnode
CCL-244		colon	HUVEC		endothelial
HCT 8	CCL-244	colon	HUT 12		fibrosarcoma
SW-948	CCL-237	colon	HT	CRL-2260	ascites
Hep-G2	HB-8065	liver	SU-DHL-4		peritoneal effussion
HS-766 T		pancreatic	SU-DHL-6		peritoneal effussion
PC-3		prostata			

Supplemental table S1. Listing of the cell lines present on the cell line tissue microarray.

Differential expressed genes based on the NFIB expression status.

Upregulated genes:

Affy_ProbeID	GeneSymbol	Description	Fold Change	p-Value
209351_at	KRT14	keratin 14 (epidermolysis bullosa simplex, Dowling-Meara, Koebner)	**15.3**	0.000030
205044_at	GABRP	gamma-aminobutyric acid (GABA) A receptor, pi	**15.0**	0.000000
209289_at	NFIB	nuclear factor I/B	**11.1**	0.000000
209613_s_at	ADH1B	alcohol dehydrogenase IB (class I), beta polypeptide	**10.7**	0.000019
209292_at	ID4	inhibitor of DNA binding 4, dominant negative helix-loop-helix protein	**9.5**	0.000001
202037_s_at	SFRP1	secreted frizzled-related protein 1	**8.0**	0.000000
207175_at	ADIPOQ	adiponectin, C1Q and collagen domain containing	**7.9**	0.001936
203980_at	FABP4	fatty acid binding protein 4, adipocyte	**7.9**	0.000282
220625_s_at	ELF5	E74-like factor 5 (ets domain transcription factor)	**7.5**	0.005698
205157_s_at	KRT17	keratin 17	**7.4**	0.000258
203548_s_at	LPL	lipoprotein lipase	**7.3**	0.001069
201820_at	KRT5	keratin 5 (epidermolysis bullosa simplex, Dowling-Meara/Kobner/Webe	**6.5**	0.000012
204455_at	DST	dystonin	**5.7**	0.009781
211737_x_at	PTN	pleiotrophin (heparin binding growth factor 8, neurite growth-promoting	**5.7**	0.015979
204304_s_at	PROM1	prominin 1	**5.6**	0.000109
212730_at	DMN	desmuslin	**5.3**	0.000000
209894_at	LEPR	leptin receptor	**4.9**	0.000169
201983_s_at	EGFR	epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b)	**4.7**	0.000000
203400_s_at	TF	transferrin	**4.2**	0.006400
204259_at	MMP7	matrix metallopeptidase 7 (matrilysin, uterine)	**4.0**	0.002927
211726_s_at	FMO2	flavin containing monooxygenase 2	**3.9**	0.000008
209763_at	CHRDL1	chordin-like 1	**3.9**	0.000000
213524_s_at	G0S2	G0/G1switch 2	**3.8**	0.035328
209283_at	CRYAB	crystallin, alpha B	**3.8**	0.000001
204855_at	SERPINB5	serpin peptidase inhibitor, clade B (ovalbumin), member 5	**3.7**	0.014040
221872_at	RARRES1	retinoic acid receptor responder (tazarotene induced) 1	**3.5**	0.000020
209318_x_at	PLAGL1	pleiomorphic adenoma gene-like 1	**3.5**	0.000000
204719_at	ABCA8	ATP-binding cassette, sub-family A (ABC1), member 8	**3.5**	0.000451
208335_s_at	DARC	Duffy blood group, chemokine receptor	**3.4**	0.000226
204151_x_at	AKR1C1	aldo-keto reductase family 1, member C1 (dihydrodiol dehydrogenase 1	**3.4**	0.018950
202746_at	ITM2A	integral membrane protein 2A	**3.4**	0.003419
205051_s_at	KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog	**3.4**	0.000046
209540_at	IGF1	insulin-like growth factor 1 (somatomedin C)	**3.3**	0.010939
202274_at	ACTG2	actin, gamma 2, smooth muscle, enteric	**3.3**	0.044845
209392_at	ENPP2	ectonucleotide pyrophosphatase/phosphodiesterase 2 (autotaxin)	**3.3**	0.000736
207030_s_at	CSRP2	cysteine and glycine-rich protein 2	**3.2**	0.000000
219497_s_at	BCL11A	B-cell CLL/lymphoma 11A (zinc finger protein)	**3.1**	0.001525
204955_at	SRPX	sushi-repeat-containing protein, X-linked	**3.1**	0.000027
209047_at	AQP1	aquaporin 1 (Colton blood group)	**3.1**	0.000000
201540_at	FHL1	four and a half LIM domains 1	**3.0**	0.019119
203131_at	PDGFRA	platelet-derived growth factor receptor, alpha polypeptide	**3.0**	0.001425
203706_s_at	FZD7	frizzled homolog 7 (Drosophila)	**2.9**	0.000000
201348_at	GPX3	glutathione peroxidase 3 (plasma)	**2.9**	0.034153
202555_s_at	MYLK	myosin, light polypeptide kinase	**2.9**	0.002204
212771_at	C10orf38	chromosome 10 open reading frame 38	**2.9**	0.000000
209774_x_at	CXCL2	chemokine (C-X-C motif) ligand 2	**2.9**	0.032566
202342_s_at	TRIM2	tripartite motif-containing 2	**2.9**	0.000008
202760_s_at	PALM2-AKAP2		**2.8**	0.003870
214456_x_at	SAA2	serum amyloid A2	**2.8**	0.000150
219768_at	VTCN1	V-set domain containing T cell activation inhibitor 1	**2.8**	0.000012
209205_s_at	LMO4	LIM domain only 4	**2.7**	0.000003
201497_x_at	MYH11	myosin, heavy polypeptide 11, smooth muscle	**2.7**	0.000004
207808_s_at	PROS1	protein S (alpha)	**2.7**	0.000006
207057_at	SLC16A7	solute carrier family 16 (monocarboxylic acid transporters), member 7	**2.6**	0.033920
201579_at	FAT	FAT tumor suppressor homolog 1 (Drosophila)	**2.6**	0.000000
207480_s_at	MEIS2	Meis1, myeloid ecotropic viral integration site 1 homolog 2 (mouse)	**2.6**	0.002431
823_at	CX3CL1	chemokine (C-X3-C motif) ligand 1	**2.6**	0.027118
213564_x_at	HSPA8	heat shock 70kDa protein 8	**2.6**	0.000000
202965_s_at	CAPN6	calpain 6	**2.6**	0.046439
202350_s_at	MATN2	matrilin 2	**2.6**	0.000029
213451_x_at	TNXB	tenascin XB	**2.6**	0.001687
219563_at	C14orf139	chromosome 14 open reading frame 139	**2.5**	0.000512
203637_s_at	MID1	midline 1 (Opitz/BBB syndrome)	**2.4**	0.000000
218736_s_at	PALMD	palmdelphin	**2.4**	0.000220
201030_x_at	LDHB	lactate dehydrogenase B	**2.4**	0.000000
214051_at	MGC39900		**2.4**	0.019306
205392_s_at	CCL14	chemokine (C-C motif) ligand 14	**2.4**	0.006287
204750_s_at	DSC2	desmocollin 2	**2.4**	0.006075
202973_x_at	FAM13A1	family with sequence similarity 13, member A1	**2.4**	0.000625

Affy_ProbeID	GeneSymbol	Description	Fold Change	p-Value
206560_s_at	MIA	melanoma inhibitory activity	**2.4**	0.042449
204359_at	FLRT2	fibronectin leucine rich transmembrane protein 2	**2.4**	0.002051
203323_at	CAV2	caveolin 2	**2.4**	0.000208
203951_at	CNN1	calponin 1, basic, smooth muscle	**2.3**	0.011330
203632_s_at	GPRC5B	G protein-coupled receptor, family C, group 5, member B	**2.3**	0.000000
209170_s_at	GPM6B	glycoprotein M6B	**2.3**	0.014103
221748_s_at	TNS1	tensin 1	**2.2**	0.000635
213348_at	CDKN1C	cyclin-dependent kinase inhibitor 1C (p57, Kip2)	**2.2**	0.000116
203256_at	CDH3	cadherin 3, type 1, P-cadherin (placental)	**2.2**	0.036454
203510_at	MET	met proto-oncogene (hepatocyte growth factor receptor)	**2.2**	0.000011
203574_at	NFIL3	nuclear factor, interleukin 3 regulated	**2.2**	0.000007
208798_x_at	GOLGA8A	golgi autoantigen, golgin subfamily a, 8A	**2.2**	0.017342
206953_s_at	LPHN2	latrophilin 2	**2.2**	0.000003
203373_at	SOCS2	suppressor of cytokine signaling 2	**2.2**	0.044004
212762_s_at	TCF7L2	transcription factor 7-like 2 (T-cell specific, HMG-box)	**2.2**	0.000001
209183_s_at	C10orf10	chromosome 10 open reading frame 10	**2.2**	0.000035
213260_at	FOXC1	forkhead box C1	**2.2**	0.000081
209355_s_at	PPAP2B	phosphatidic acid phosphatase type 2B	**2.2**	0.028322
210024_s_at	UBE2E3	ubiquitin-conjugating enzyme E2E 3 (UBC4/5 homolog, yeast)	**2.2**	0.000000
204897_at	PTGER4	prostaglandin E receptor 4 (subtype EP4)	**2.2**	0.009698
202992_at	C7	complement component 7	**2.1**	0.009689
201300_s_at	PRNP	prion protein (p27-30) (Creutzfeld-Jakob disease, Gerstmann-Strausler	**2.1**	0.000012
208370_s_at	DSCR1	Down syndrome critical region gene 1	**2.1**	0.000795
209074_s_at	FAM107A	family with sequence similarity 107, member A	**2.1**	0.010565
202133_at	WWTR1	WW domain containing transcription regulator 1	**2.1**	0.000015
208131_s_at	PTGIS	prostaglandin I2 (prostacyclin) synthase	**2.1**	0.000016
218236_s_at	PRKD3	protein kinase D3	**2.1**	0.000015
203180_at	ALDH1A3	aldehyde dehydrogenase 1 family, member A3	**2.1**	0.035185
204326_x_at	MT1L	metallothionein 1L	**2.0**	0.004895
203408_s_at	SATB1	special AT-rich sequence binding protein 1 (binds to nuclear matrix/sca	**2.0**	0.002413
212558_at	SPRY1	sprouty homolog 1, antagonist of FGF signaling (Drosophila)	**2.0**	0.003633
209369_at	ANXA3	annexin A3	**2.0**	0.003498
212724_at	RND3	Rho family GTPase 3	**2.0**	0.003076
219134_at	ELTD1	EGF, latrophilin and seven transmembrane domain containing 1	**2.0**	0.035824
209108_at	TSPAN6	tetraspanin 6	**2.0**	0.002146
213844_at	HOXA5	homeobox A5	**2.0**	0.000018
201160_s_at	CSDA	cold shock domain protein A	**2.0**	0.000109
210473_s_at	CDC2L2	cell division cycle 2-like 2 (PITSLRE proteins)	**2.0**	0.000658
218319_at	PELI1	pellino homolog 1 (Drosophila)	**2.0**	0.000000
206453_s_at	NDRG2	NDRG family member 2	**2.0**	0.001007
204400_at	EFS	embryonal Fyn-associated substrate	**2.0**	0.012968
217833_at	SYNCRIP	synaptotagmin binding, cytoplasmic RNA interacting protein	**1.9**	0.000000
219093_at	FLJ20701		**1.9**	0.040636
209212_s_at	KLF5	Kruppel-like factor 5 (intestinal)	**1.9**	0.040609
221016_s_at	TCF7L1	transcription factor 7-like 1 (T-cell specific, HMG-box)	**1.9**	0.000000
204159_at	CDKN2C	cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4)	**1.9**	0.000024
204042_at	WASF3	WAS protein family, member 3	**1.9**	0.000024
210946_at	PPAP2A	phosphatidic acid phosphatase type 2A	**1.9**	0.000674
219064_at	ITIH5	inter-alpha (globulin) inhibitor H5	**1.9**	0.019792
204011_at	SPRY2	sprouty homolog 2 (Drosophila)	**1.9**	0.011280
218031_s_at			**1.8**	0.013477
201034_at	ADD3	adducin 3 (gamma)	**1.8**	0.040335
212263_at	QKI	quaking homolog, KH domain RNA binding (mouse)	**1.8**	0.000019
218486_at	KLF11	Kruppel-like factor 11	**1.8**	0.000000
202146_at	IFRD1	interferon-related developmental regulator 1	**1.8**	0.003469
203038_at	PTPRK	protein tyrosine phosphatase, receptor type, K	**1.8**	0.001707
201328_at	ETS2	v-ets erythroblastosis virus E26 oncogene homolog 2 (avian)	**1.8**	0.000043
204872_at	TLE4	transducin-like enhancer of split 4 (E(sp1) homolog, Drosophila)	**1.8**	0.000003
218019_s_at			**1.8**	0.000000
203088_at	FBLN5	fibulin 5	**1.8**	0.018248
202933_s_at	YES1	v-yes-1 Yamaguchi sarcoma viral oncogene homolog 1	**1.8**	0.008264
204223_at	PRELP	proline/arginine-rich end leucine-rich repeat protein	**1.8**	0.001106
200762_at	DPYSL2	dihydropyrimidinase-like 2	**1.8**	0.018297
209373_at	MALL	mal, T-cell differentiation protein-like	**1.8**	0.042302
212589_at	RRAS2	related RAS viral (r-ras) oncogene homolog 2	**1.8**	0.033426
209897_s_at	SLIT2	slit homolog 2 (Drosophila)	**1.7**	0.022056
212345_s_at	CREB3L2	cAMP responsive element binding protein 3-like 2	**1.7**	0.000493
215039_at	LOC339524		**1.7**	0.010515
201811_x_at	SH3BP5	SH3-domain binding protein 5 (BTK-associated)	**1.7**	0.001735
203895_at	PLCB4	phospholipase C, beta 4	**1.7**	0.002575
212614_at	ARID5B	AT rich interactive domain 5B (MRF1-like)	**1.7**	0.001779
213541_s_at	ERG	v-ets erythroblastosis virus E26 oncogene like (avian)	**1.7**	0.000039
221556_at	CDC14B	CDC14 cell division cycle 14 homolog B (S. cerevisiae)	**1.7**	0.000423

Supplemental table S2. Genes that are upregulated on the NFIB high expressing samples. Genes marked in grey were also present on the topgenes of the Between Group Analysis. P-values have been adjusted according to Benjamini and Hochberg. Duplicate gene entries were removed.

Downregulated genes:

Affy_ProbeID	GeneSymbol	Description	Fold Change	p-Value
205358_at	GRIA2	glutamate receptor, ionotropic, AMPA 2	**-14.6**	0.019359
206754_s_at	CYP2B6	cytochrome P450, family 2, subfamily B, polypeptide 6	**-7.1**	0.000001
204623_at	TFF3	trefoil factor 3 (intestinal)	**-5.0**	0.000028
205364_at	ACOX2	acyl-Coenzyme A oxidase 2, branched chain	**-4.7**	0.000026
205009_at	TFF1	trefoil factor 1 (breast cancer, estrogen-inducible sequence expressed i	**-4.7**	0.026084
221577_x_at	GDF15	growth differentiation factor 15	**-4.3**	0.017850
210272_at	CYP2B7P1	cytochrome P450, family 2, subfamily B, polypeptide 7 pseudogene 1	**-4.1**	0.000002
219872_at	C4orf18	chromosome 4 open reading frame 18	**-3.8**	0.000197
201884_at	CEACAM5	carcinoembryonic antigen-related cell adhesion molecule 5	**-3.7**	0.009543
204540_at	EEF1A2	eukaryotic translation elongation factor 1 alpha 2	**-3.7**	0.000282
205696_s_at	GFRA1	GDNF family receptor alpha 1	**-3.6**	0.000127
209460_at	ABAT	4-aminobutyrate aminotransferase	**-3.5**	0.009234
204597_x_at	STC1	stanniocalcin 1	**-3.3**	0.003454
210735_s_at	CA12	carbonic anhydrase XII	**-3.3**	0.000000
216623_x_at	TNRC9	trinucleotide repeat containing 9	**-3.2**	0.017680
218280_x_at	HIST2H2AA	histone 2, H2aa	**-3.1**	0.000516
209624_s_at	MCCC2	methylcrotonoyl-Coenzyme A carboxylase 2 (beta)	**-3.0**	0.000000
202376_at	SERPINA3	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin),	**-2.8**	0.003104
215867_x_at	AP1G1	adaptor-related protein complex 1, gamma 1 subunit	**-2.7**	0.000740
204070_at	RARRES3	retinoic acid receptor responder (tazarotene induced) 3	**-2.7**	0.031627
202089_s_at	SLC39A6	solute carrier family 39 (zinc transporter), member 6	**-2.7**	0.017770
203876_s_at	MMP11	matrix metallopeptidase 11 (stromelysin 3)	**-2.6**	0.001742
201130_s_at	CDH1	cadherin 1, type 1, E-cadherin (epithelial)	**-2.5**	0.012849
218532_s_at	FLJ20152		**-2.4**	0.003561
201841_s_at	HSPB1	heat shock 27kDa protein 1	**-2.4**	0.000305
203771_s_at	BLVRA	biliverdin reductase A	**-2.4**	0.000003
205645_at	REPS2	RALBP1 associated Eps domain containing 2	**-2.3**	0.012407
203108_at	GPRC5A	G protein-coupled receptor, family C, group 5, member A	**-2.3**	0.004445
215552_s_at	ESR1	estrogen receptor 1	**-2.3**	0.000001
205081_at	CRIP1	cysteine-rich protein 1 (intestinal)	**-2.2**	0.005826
208682_s_at	MAGED2	melanoma antigen family D, 2	**-2.2**	0.000064
201169_s_at	BHLHB2	basic helix-loop-helix domain containing, class B, 2	**-2.2**	0.000250
218510_x_at	FLJ20152		**-2.2**	0.000027
219438_at	FAM77C	family with sequence similarity 77, member C	**-2.2**	0.003006
217202_s_at	GLUL	glutamate-ammonia ligase (glutamine synthetase)	**-2.1**	0.007699
202489_s_at	FXYD3	FXYD domain containing ion transport regulator 3	**-2.1**	0.037247
219051_x_at	METRN	meteorin, glial cell differentiation regulator	**-2.1**	0.000001
209276_s_at	GLRX	glutaredoxin (thioltransferase)	**-2.1**	0.000002
201107_s_at	THBS1	thrombospondin 1	**-2.1**	0.000005
201596_x_at	KRT18	keratin 18	**-2.1**	0.033621
201754_at	COX6C	cytochrome c oxidase subunit VIc	**-2.1**	0.000030
201563_at	SORD	sorbitol dehydrogenase	**-2.0**	0.031669
201461_s_at	MAPKAPK2	mitogen-activated protein kinase-activated protein kinase 2	**-2.0**	0.001783
218409_s_at	DNAJC1	DnaJ (Hsp40) homolog, subfamily C, member 1	**-2.0**	0.000029
200923_at	LGALS3BP	lectin, galactoside-binding, soluble, 3 binding protein	**-2.0**	0.011551
204934_s_at	HPN	hepsin (transmembrane protease, serine 1)	**-1.9**	0.000001
202121_s_at	CHMP2A	chromatin modifying protein 2A	**-1.9**	0.000006
202308_at	SREBF1	sterol regulatory element binding transcription factor 1	**-1.9**	0.012844
202996_at	POLD4	polymerase (DNA-directed), delta 4	**-1.9**	0.000000
204006_s_at	FCGR3B	Fc fragment of IgG, low affinity IIIb, receptor (CD16b)	**-1.9**	0.021015
210085_s_at	ANXA9	annexin A9	**-1.9**	0.000296
202545_at	PRKCD	protein kinase C, delta	**-1.9**	0.000000
208677_s_at	BSG	basigin (Ok blood group)	**-1.9**	0.027673
207549_x_at	CD46	CD46 antigen, complement regulatory protein	**-1.9**	0.010264
213419_at	APBB2	amyloid beta (A4) precursor protein-binding, family B, member 2 (Fe65-	**-1.9**	0.000001
200862_at	DHCR24	24-dehydrocholesterol reductase	**-1.8**	0.043355
222067_x_at	HIST1H2BD	histone 1, H2bd	**-1.8**	0.005305
214269_at	MFSD7	major facilitator superfamily domain containing 7	**-1.8**	0.013211
209417_s_at	IFI35	interferon-induced protein 35	**-1.8**	0.046785
204981_at	SLC22A18	solute carrier family 22 (organic cation transporter), member 18	**-1.8**	0.000002
208621_s_at	VIL2	villin 2 (ezrin)	**-1.8**	0.000533
201349_at	SLC9A3R1	solute carrier family 9 (sodium/hydrogen exchanger), member 3 regulat	**-1.8**	0.021231
217234_s_at	VIL2	villin 2 (ezrin)	**-1.8**	0.013253
208934_s_at	LGALS8	lectin, galactoside-binding, soluble, 8 (galectin 8)	**-1.8**	0.000101
220240_s_at	TMCO3	transmembrane and coiled-coil domains 3	**-1.8**	0.008691
205354_at	GAMT	guanidinoacetate N-methyltransferase	**-1.7**	0.000008
218813_s_at	SH3GLB2	SH3-domain GRB2-like endophilin B2	**-1.7**	0.001294
203071_at	SEMA3B	sema domain, immunoglobulin domain (Ig), short basic domain, secrete	**-1.7**	0.033097
220744_s_at	IFT122	intraflagellar transport 122 homolog (Chlamydomonas)	**-1.7**	0.017281
202201_at	BLVRB	biliverdin reductase B (flavin reductase (NADPH))	**-1.7**	0.001106
207782_s_at	PSEN1	presenilin 1 (Alzheimer disease 3)	**-1.7**	0.000437

Affy_ProbeID	GeneSymbol	Description	Fold Change	p-Value
206472_s_at	TLE3	transducin-like enhancer of split 3 (E(sp1) homolog, Drosophila)	**-1.7**	0.000000
211015_s_at	HSPA4	heat shock 70kDa protein 4	**-1.7**	0.000012
220540_at	KCNK15	potassium channel, subfamily K, member 15	**-1.7**	0.035033
212637_s_at	WWP1	WW domain containing E3 ubiquitin protein ligase 1	**-1.7**	0.026216
221759_at	G6PC3	glucose 6 phosphatase, catalytic, 3	**-1.7**	0.000025
210246_s_at	ABCC8	ATP-binding cassette, sub-family C (CFTR/MRP), member 8	**-1.7**	0.000000
203228_at	PAFAH1B3	platelet-activating factor acetylhydrolase, isoform Ib, gamma subunit 29	**-1.7**	0.000017
203252_at	CDK2AP2	CDK2-associated protein 2	**-1.7**	0.005450
203143_s_at	LOC147299		**-1.7**	0.017562
220588_at	BCAS4	breast carcinoma amplified sequence 4	**-1.7**	0.000001
201284_s_at	APEH	N-acylaminoacyl-peptide hydrolase	**-1.7**	0.000003
212446_s_at	LASS6	LAG1 longevity assurance homolog 6 (S. cerevisiae)	**-1.6**	0.008539
214196_s_at	TPP1	tripeptidyl peptidase I	**-1.6**	0.029728
210904_s_at	IL13RA1	interleukin 13 receptor, alpha 1	**-1.6**	0.000273
214446_at	ELL2	elongation factor, RNA polymerase II, 2	**-1.6**	0.001428
203880_at	COX17	COX17 homolog, cytochrome c oxidase assembly protein (yeast)	**-1.6**	0.002081
218394_at	ROGDI	rogdi homolog (Drosophila)	**-1.6**	0.001589
208583_x_at	HIST1H2AJ	histone 1, H2aj	**-1.6**	0.000001
209367_at	STXBP2	syntaxin binding protein 2	**-1.6**	0.000000
219709_x_at	C16orf24	chromosome 16 open reading frame 24	**-1.6**	0.000480
209149_s_at	TM9SF1	transmembrane 9 superfamily member 1	**-1.6**	0.000001
216202_s_at	SPTLC2	serine palmitoyltransferase, long chain base subunit 2	**-1.6**	0.000027
200846_s_at	PPP1CA	protein phosphatase 1, catalytic subunit, alpha isoform	**-1.6**	0.000730
216988_s_at	PTP4A2	protein tyrosine phosphatase type IVA, member 2	**-1.6**	0.004742
200776_s_at	LOC151579		**-1.6**	0.001281
202641_at	ARL3	ADP-ribosylation factor-like 3	**-1.6**	0.001660
214522_x_at	HIST1H2AD	histone 1, H2ad	**-1.6**	0.000460
218556_at	ORMDL2	ORM1-like 2 (S. cerevisiae)	**-1.6**	0.000008
202740_at	ACY1	aminoacylase 1	**-1.6**	0.002389
210534_s_at	EPPB9		**-1.6**	0.008091
207809_s_at	ATP6AP1	ATPase, H+ transporting, lysosomal accessory protein 1	**-1.6**	0.000935
218436_at	SIL1	SIL1 homolog, endoplasmic reticulum chaperone (S. cerevisiae)	**-1.6**	0.004556
215891_s_at	GM2A	GM2 ganglioside activator	**-1.6**	0.026880
210720_s_at	APBA2BP	amyloid beta (A4) precursor protein-binding, family A, member 2 bindin	**-1.6**	0.000077
220613_s_at	SYTL2	synaptotagmin-like 2	**-1.6**	0.000017
202528_at	GALE	UDP-galactose-4-epimerase	**-1.6**	0.017502
218921_at	SIGIRR	single immunoglobulin and toll-interleukin 1 receptor (TIR) domain	**-1.6**	0.000282
204862_s_at	NME3	non-metastatic cells 3, protein expressed in	**-1.6**	0.004424
221823_at	LOC90355		**-1.6**	0.003597
201096_s_at	ARF4	ADP-ribosylation factor 4	**-1.6**	0.000005
200803_s_at	TEGT	testis enhanced gene transcript (BAX inhibitor 1)	**-1.6**	0.001915
219562_at	RAB26	RAB26, member RAS oncogene family	**-1.6**	0.000016
202587_s_at	AK1	adenylate kinase 1	**-1.6**	0.000541
201457_x_at	BUB3	BUB3 budding uninhibited by benzimidazoles 3 homolog (yeast)	**-1.5**	0.000045
200078_s_at	ATP6V0B	ATPase, H+ transporting, lysosomal 21kDa, V0 subunit b	**-1.5**	0.002715
221882_s_at	TMEM8	transmembrane protein 8 (five membrane-spanning domains)	**-1.5**	0.013758
203458_at	SPR	sepiapterin reductase (7,8-dihydrobiopterin:NADP+ oxidoreductase)	**-1.5**	0.000044
218866_s_at	POLR3K	polymerase (RNA) III (DNA directed) polypeptide K, 12.3 kDa	**-1.5**	0.002110
201583_s_at	SEC23B	Sec23 homolog B (S. cerevisiae)	**-1.5**	0.031603
212495_at	JMJD2B	jumonji domain containing 2B	**-1.5**	0.019902
212400_at	FAM102A	family with sequence similarity 102, member A	**-1.5**	0.000001
218189_s_at	NANS	N-acetylneuraminic acid synthase (sialic acid synthase)	**-1.5**	0.026942
211495_x_at	TNFSF13	tumor necrosis factor (ligand) superfamily, member 13	**-1.5**	0.000018
202180_s_at	MVP	major vault protein	**-1.5**	0.001008
208872_s_at	REEP5	receptor accessory protein 5	**-1.5**	0.002465
220937_s_at	ST6GALNAC4	ST6 (alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3)-N-acetylgalact	**-1.5**	0.033984
209177_at	C3orf60	chromosome 3 open reading frame 60	**-1.5**	0.000351
205455_at	MST1R	macrophage stimulating 1 receptor (c-met-related tyrosine kinase)	**-1.5**	0.000841
217770_at	PIGT	phosphatidylinositol glycan, class T	**-1.5**	0.000021
212574_x_at	C19orf6	chromosome 19 open reading frame 6	**-1.5**	0.020814
218302_at	PSENEN	presenilin enhancer 2 homolog (C. elegans)	**-1.5**	0.025941
200769_s_at	MAT2A	methionine adenosyltransferase II, alpha	**-1.5**	0.003115
208270_s_at	RNPEP	arginyl aminopeptidase (aminopeptidase B)	**-1.5**	0.000647
210974_s_at	AP3D1	adaptor-related protein complex 3, delta 1 subunit	**-1.5**	0.000073
209872_s_at	PKP3	plakophilin 3	**-1.5**	0.016472
221754_s_at	CORO1B	coronin, actin binding protein, 1B	**-1.5**	0.007073
206491_s_at	NAPA	N-ethylmaleimide-sensitive factor attachment protein, alpha	**-1.5**	0.003131
203258_at	DRAP1	DR1-associated protein 1 (negative cofactor 2 alpha)	**-1.5**	0.032251
218291_at	MAPBPIP		**-1.5**	0.000058

Affy_ProbeID	GeneSymbol	Description	Fold Change	p-Value
220477_s_at	C20orf30	chromosome 20 open reading frame 30	**-1.5**	0.023263
209665_at	CYB561D2	cytochrome b-561 domain containing 2	**-1.5**	0.000005
214992_s_at	DNASE2	deoxyribonuclease II, lysosomal	**-1.5**	0.008132
211113_s_at	ABCG1	ATP-binding cassette, sub-family G (WHITE), member 1	**-1.5**	0.010944
218548_x_at	TEX264	testis expressed sequence 264	**-1.5**	0.006374
200085_s_at	TCEB2	transcription elongation factor B (SIII), polypeptide 2 (18kDa, elongin B)	**-1.5**	0.000026
216483_s_at	C19orf10	chromosome 19 open reading frame 10	**-1.5**	0.000020
217927_at	SPCS1	signal peptidase complex subunit 1 homolog (S. cerevisiae)	**-1.5**	0.000002
218262_at	FLJ22318		**-1.5**	0.000192
218010_x_at	C20orf149	chromosome 20 open reading frame 149	**-1.5**	0.014961
204088_at	P2RX4	purinergic receptor P2X, ligand-gated ion channel, 4	**-1.5**	0.015753
200096_s_at	ATP6V0E	ATPase, H+ transporting, lysosomal 9kDa, V0 subunit e	**-1.5**	0.034068
201471_s_at	SQSTM1	sequestosome 1	**-1.4**	0.001966
218688_at	DAK	dihydroxyacetone kinase 2 homolog (yeast)	**-1.4**	0.002899
201003_x_at	Kua-UEV		**-1.4**	0.000259
209731_at	NTHL1	nth endonuclease III-like 1 (E. coli)	**-1.4**	0.002775
218931_at	RAB17	RAB17, member RAS oncogene family	**-1.4**	0.002628
221946_at	C9orf116	chromosome 9 open reading frame 116	**-1.4**	0.000001
204398_s_at	EML2	echinoderm microtubule associated protein like 2	**-1.4**	0.000082
200929_at	TMED10	transmembrane emp24-like trafficking protein 10 (yeast)	**-1.4**	0.000077
203054_s_at	TCTA	T-cell leukemia translocation altered gene	**-1.4**	0.001556
221585_at	CACNG4	calcium channel, voltage-dependent, gamma subunit 4	**-1.4**	0.003035
214687_x_at	ALDOA	aldolase A, fructose-bisphosphate	**-1.4**	0.009821
203890_s_at	DAPK3	death-associated protein kinase 3	**-1.4**	0.001814
200622_x_at	CALM3	calmodulin 3 (phosphorylase kinase, delta)	**-1.4**	0.000811
217785_s_at	YKT6		**-1.4**	0.013313
204125_at	NDUFAF1	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, assembly fac	**-1.4**	0.024381
208911_s_at	PDHB	pyruvate dehydrogenase (lipoamide) beta	**-1.4**	0.000004
203246_s_at	TUSC4	tumor suppressor candidate 4	**-1.4**	0.000554
211622_s_at	ARF3	ADP-ribosylation factor 3	**-1.4**	0.000302
217748_at	ADIPOR1	adiponectin receptor 1	**-1.4**	0.000865
203272_s_at	TUSC2	tumor suppressor candidate 2	**-1.4**	0.000210
202852_s_at	FLJ11506		**-1.4**	0.032850
209853_s_at	PSME3	proteasome (prosome, macropain) activator subunit 3 (PA28 gamma, I	**-1.4**	0.000138
212995_x_at	FLJ14346		**-1.4**	0.034196
208689_s_at	RPN2	ribophorin II	**-1.4**	0.003273
219203_at	C14orf122	chromosome 14 open reading frame 122	**-1.4**	0.000004
217977_at	SEPX1	selenoprotein X, 1	**-1.4**	0.000299
219725_at	TREM2	triggering receptor expressed on myeloid cells 2	**-1.4**	0.000289
204824_at	ENDOG	endonuclease G	**-1.4**	0.017578
202296_s_at	RER1	RER1 retention in endoplasmic reticulum 1 homolog (S. cerevisiae)	**-1.4**	0.021011
57540_at	RBKS	ribokinase	**-1.4**	0.016288
202143_s_at	COPS8	COP9 constitutive photomorphogenic homolog subunit 8 (Arabidopsis)	**-1.4**	0.016403
200736_s_at	GPX1	glutathione peroxidase 1	**-1.4**	0.012940
218567_x_at	DPP3	dipeptidyl-peptidase 3	**-1.4**	0.016852
221856_s_at	FAM63A	family with sequence similarity 63, member A	**-1.4**	0.000575
36994_at	ATP6V0C	ATPase, H+ transporting, lysosomal 16kDa, V0 subunit c	**-1.4**	0.034177
218112_at	MRPS34	mitochondrial ribosomal protein S34	**-1.4**	0.000110
221512_at	C1orf160	chromosome 1 open reading frame 160	**-1.4**	0.022190
204262_s_at	PSEN2	presenilin 2 (Alzheimer disease 4)	**-1.4**	0.011113
212729_at	DLG3	discs, large homolog 3 (neuroendocrine-dlg, Drosophila)	**-1.4**	0.000439
213057_at	ATPAF2	ATP synthase mitochondrial F1 complex assembly factor 2	**-1.4**	0.000443

Supplemental Table S3. Genes that are downregulated on the NFIB high expressing samples. Genes marked in grey were also present on the topgenes of the Between Group Analysis. P-values have been adjusted according to Benjamini and Hochberg. Duplicate gene entries were removed.

3. Discussion

Breast cancer is the most common form of cancer and the most common cause of cancer death in females, worldwide (Sasco et al., 2003). It will affect approximately one out of twelve women at some stage in their life. Although the incidence as well as the prevalence of this disease in western countries increases, the mortality is decreasing due to improved treatment of breast cancer patients. Nevertheless, the decision making for treatment is delicate, since breast tumors are heterogeneous and consist of several pathological subtypes with different clinical presentations and outcomes. In addition, patients show a diverse range of responses to a given treatment. The decision making on adjuvant therapy is more and more influenced or even determined by molecular biological findings. Estrogen receptor and progesterone receptor positive patients are treated with anti-hormonal therapy. Diagnostics can even go further to the genomic level, as in the case of HER-2 gene amplification detection. Patients with a positive finding will be administered Herceptin. In the last decades the role of amplifications in cancer has become indisputable. Besides HER-2, several other oncogenes could be localized on amplified regions, for example CCND1 at 11q13 (Karlseder et al., 1994) or MDM2 at 12q13 (Courjal et al., 1996).

In our study we were able to identify several novel amplified regions in breast cancer. Two of them were further investigated: the 6q25 and the 9p24 gene amplifications.

The **6q25 amplification** comprising the ESR1 gene was the most astonishing finding in our study. It is remarkable, that the importance of the ESR1 gene amplification, which is of one the most or maybe even the most frequent amplification in breast cancer (21%), has not been discovered earlier. This is particularly surprising since ER expression levels are routinely measured for diagnosis and represent a major factor for treatment decision. The difficulties detecting this amplification are likely due to the small size of the amplicon; in one case the amplicon only measured 514 Kb. Although ESR1 amplification frequency is relatively high, it is only present in a subset of about 50% of the breast cancers expressing high ER levels (on IHC). Further statistical analysis of these results in combination with additional data about the treatment these patients had received, revealed ESR1 gene amplified

patients as a subgroup of ER (IHC) positive patients with a much higher likelihood to positively respond to anti-hormonal therapy. We could also show that ESR1 gene amplification can be present in premalignant breast lesions which may have an increased potential for malignant transformation. All these findings advocate a major impact of ESR1 gene amplification on decision making regarding anti-hormonal therapy in breast cancer patients. Our findings are supported by the data available from public resources. As mentioned before (Introduction), Sorlie et al showed that the luminal genetic subgroups were associated with best survival in comparison to ER negative groups (i.e. basal-like and HER-2 group) (Figure 12). They reproduced these findings on another breast cancer data set and showed that the luminal A group, which is the ER high-expressing group, has a markedly better survival than the luminal B group (low to moderate ER expression levels) (Figure 12). The authors denote this finding as an intriguing result, since both luminal groups are ER positive. Further, they suggest that the luminal B subtype may reflect a group of patients who will not benefit from adjuvant Tamoxifen. This suggestion completely reflects our findings on the effect of ESR1 gene amplification on Tamoxifen drug response. This may lead to the solution of the long-lasting mystery of the two luminal subgroups, suggesting that the luminal A group is the ER positive ESR1 amplified group, and the luminal B group is not ESR1 amplified, but still ER positive. Successful demonstration of these findings would make it reasonable to analyze genes differentially expressed between these two luminal groups. This would help us to find genes which are responsible for the different behaviour of the tumors that are ER positive, but not ESR1 amplified, in order to find a specific drug target for this subgroup.

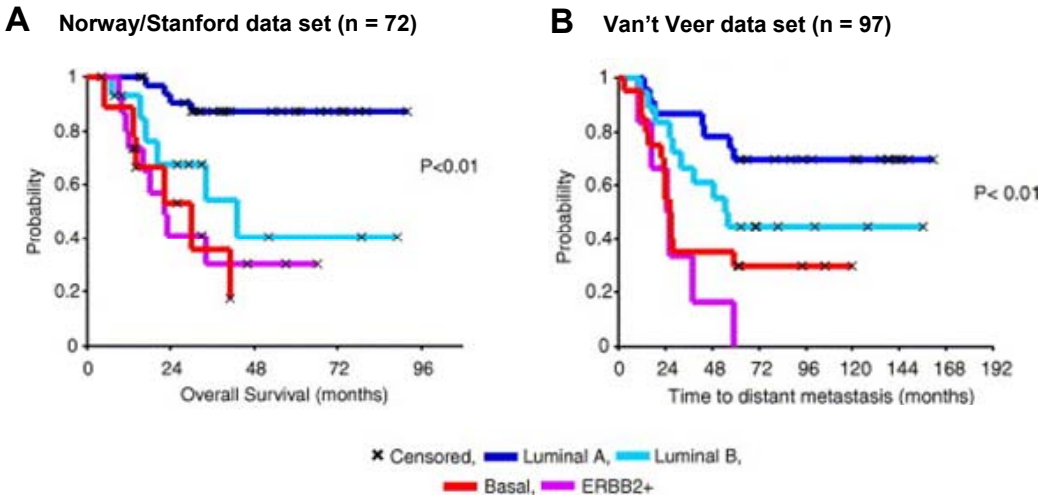


Figure 12. Kaplan-Meier survival analysis of the Perou/Sorlie subgroups applied on two different breast cancer data sets: A) Norway/Stanford data set. B) Van't Veer data set. Modified from Sorlie et al (Sorlie et al., 2003).

The other amplification we have focused our work on, was the **9p24 gene amplification**. Similar to the ESR1 gene amplification, the amplicon revealed in one case a very small size of 0.7 Mb, only comprising the NFIB gene. This amplification has never been described in breast cancer, so far. One publication described once a 9p23-24 amplification in squamous cell carcinoma of the esophagus, but the authors reported that the most prominent amplified region harbors the GASC1 gene, which is 7 Mb away from the NFIB gene (Yang et al., 2001).

Our findings from the FISH analysis on the Breast Prognosis TMA in combination with the RNA expression analysis strongly suggest the NFIB gene as the functionally important gene of the 9p24 amplicon in breast cancer. Further, since NFIB gene amplified samples showed a higher proliferation index (Ki67 LI) and the NFIB RNA interference experiments resulted in reduced levels of proliferation, we advocate a proliferation supporting role for the NFIB protein. These findings are consistent with the already described properties of the members of the NFI family. The members of this family were first described as being required for the replication of Adenovirus DNA, later shown to regulate the transcription of a large variety of cellular and viral genes (reviewed in (Gronostajski, 2000)). NFIB does also contain the DNA binding domain specific for the same palindromic sequence as the other members of the NFI family. Nevertheless, less is known about the NFIB protein. So far, the NFIB protein isoform with a MW of 47 kDa (SwissProt: NFIB_HUMAN) was handled as the so-

called "normal" isoform. Our analyses revealed that the most prominent NFIB isoform in the cell lines tested was an isoform with a MW around 63 kDa. Further analyses of this isoform, especially regarding its potential target genes are still necessary. The analysis of an Affymetrix expression microarray public set (GSE 1456) comprising 159 breast cancers revealed an intriguing finding: samples belonging to the basal-like group of breast cancers showed a statistically significantly higher NFIB expression than samples belonging to the other molecular expression groups, previously described by Perou and Sorlie (Perou et al., 2000; Sorlie et al., 2001). An important role concerning NFIB in the basal-like breast cancers was strongly enforced by the results obtained from the evaluation of the DNA amplification levels on the Breast Prognosis TMA. NFIB gene amplification showed its highest rate on medullary carcinomas (19%, $p = 0.0007$). Interestingly, this subtype of breast carcinomas has recently been classified as breast cancers belonging to the basal-like expression group (Bertucci et al., 2006). In summary, we have shown that the NFIB gene is not only overexpressed in the basal group of breast cancers, but also amplified in a considerable fraction of them. These findings suggest that the NFIB gene is relevant for the development of these breast cancers. Our data from the NFIB gene knockdown experiments point towards a proliferation supporting role of NFIB in these cancers.

Since we have found several new amplified regions in our limited study, we are convinced that the majority of the amplifications in breast cancer are still unknown. Our findings suggest the existence of a higher number of different amplifications than anticipated, but with a decreasing size. The two amplicons described in this work (the 9p24 and the 6q25) were smaller than 1 Mb. The most significant prerequisite for the detection of small gene amplifications is the resolution of the arrays used. We became aware of this when we investigated the most prominent gene amplification in breast cancer, the HER-2 gene amplification at 17q21. FISH analysis of the used breast cancer samples revealed four HER-2 amplified samples. Surprisingly, our bioinformatics approach was only able to confirm this result in two cases. Thorough informatic analysis of this region on the SNP array revealed an insufficient resolution of barely one SNP per Mb around the HER-2 gene locus. In addition, there was also a two Mb SNP-less region nearby this gene. Hence, only the detection of larger

amplifications on the 17q21 region was possible. Fortunately, most of the chromosomes were better resolved and could be used for the detection of small amplicons.

In addition, we have to consider that our set comprised only 30 breast cancer samples and that most of the amplifications occur at a low rate. In order to enhance our chances, we selected only grade 3 breast cancers for the analysis, since these are known to harbor more amplifications. However, this is not necessarily the best selection since the ESR1 gene amplification was predominantly present in low grade tumors.

For this study, we used the Affymetrix SNP 10k 2.0 arrays, since at the time this project started (spring 2004), these arrays were up-to-date. They consisted of 10'204 SNPs with a mean inter-SNP distance of 258 kb. In the mean time the actual high-end SNP arrays consist of 500'000 SNPs. In order to achieve a satisfactory resolution with the old type of microarrays, it was key to keep the noise level as low as possible. This was especially ambitious because we used breast cancer samples and these are known to be heterogeneous in cellular composition. Tissue pieces punched from the tumor will always be contaminated with a (small) fraction of non-tumorigenic cells. Since this noise is intrinsic, i.e. it has not been artificially added by a technical procedure, it is not possible to completely remove it. For this reason, the use of the reference set provided by Affymetrix comprising 110 normal human samples was not acceptable for our analysis. The bias generated by using a reference set, whose nature and processing was not performed in the same laboratory, was also criticized by other authors (Nannya et al., 2005). They even argued that the use of different PCR machines in the same laboratory may lead to significantly higher noise. Hence, we decided to develop our own analysis procedure which does not involve the use of an external reference set. We used a combination of freely available analysis tools, starting with the DChip software for normalization of the arrays. Between normalization and calculation of the DNA copy numbers, we implemented an additional smoothening step (Figure 9B). The appliance of the normalization, followed by this smoothening step, led to a tremendous increase in the signal-to-noise ratio (Figure 8). This was the major breakthrough which allowed us the detection of small-sized amplicons, like the NFIB (0.7 Mb) and the ESR1 (0.5 Mb) gene amplifications. These new amplifications are the smallest in size ever described in the literature. Many screenings for

amplifications in breast cancer have been performed with high resolution array based CGH, but have never led to the new identification of such small single gene amplicons.

The amplifications discovered in this study may represent only the tip of the iceberg, and not a complete profile. The findings of this study strongly indicate that there must be a huge number of undiscovered small-sized gene amplifications. Our novel methodology, combined with the improvements in array resolution and complexity, is expected to yield many more potential targets for cancer therapy in the future.

3.1. Conclusion

In this work, we presented two new amplifications in breast cancer, the 9p24 amplicon comprising the NFIB gene, and the 6q25 amplicon with the ESR1 gene. The impact of the latter on breast cancer is indisputable, as we have demonstrated the relevance of its existence on decision making for therapy. In contrast, the NFIB gene is relatively unknown and additional work has to be performed in order to confirm its role in breast cancer, such as overexpression of the different NFIB isoforms. FISH analysis on a multi-tumor TMA would reveal if this gene amplification is present in other tumors than breast cancer. We already presented the first indication by showing its amplification in the non-breast tumor cell line S-117 (thyroid sarcoma).

Importantly, this work shows that unconventional approaches, like the screening of 30 breast cancer samples, can be very successful. During the last 30 years, thousands of reports about the estrogen receptor and its role in breast cancer have been published, but until our discovery, the importance of its amplification was unknown. Hence, we suggest to repeat this study with at least 100 breast cancer samples comprising all kind of grades, but using a higher resolution, like the Affymetrix 1-Million-SNP array which will be introduced in the near future. We are convinced that such a study would allow the detection of several very small gene amplifications and deletions with high impact on breast cancer development or even therapy.

4. References

Alberts B, Johnson A, Lewis J, Raff M, Roberts K and Walter P. (2000). *Molecular Biology of the Cell*. Garland Science.

Bailey SM and Murnane JP. *Telomeres, chromosome instability and cancer* (2006). *Nucleic Acids Res*, 34**,** 2408-17.

Barker PE. *Double minutes in human tumor cells* (1982). *Cancer Genet Cytogenet*, 5**,** 81-94.

Barlund M, Tirkkonen M, Forozan F, Tanner MM, Kallioniemi O and Kallioniemi A. *Increased copy number at 17q22-q24 by CGH in breast cancer is due to high-level amplification of two separate regions* (1997). *Genes Chromosomes Cancer*, 20**,** 372-6.

Bast RC, Jr., Ravdin P, Hayes DF, Bates S, Fritsche H, Jr., Jessup JM, Kemeny N, Locker GY, Mennel RG and Somerfield MR. *2000 update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology* (2001). *J Clin Oncol*, 19**,** 1865-78.

Baxter SW, Choong DY, Eccles DM and Campbell IG. *Transforming growth factor beta receptor 1 polyalanine polymorphism and exon 5 mutation analysis in breast and ovarian cancer* (2002). *Cancer Epidemiol Biomarkers Prev*, 11**,** 211-4.

Bertucci F, Finetti P, Cervera N, Charafe-Jauffret E, Mamessier E, Adelaide J, Debono S, Houvenaeghel G, Maraninchi D, Viens P, Charpin C, Jacquemier J and Birnbaum D. *Gene expression profiling shows medullary breast cancer is a subgroup of basal breast cancers* (2006). *Cancer Res*, 66**,** 4636-44.

Burgess AW, Cho HS, Eigenbrot C, Ferguson KM, Garrett TP, Leahy DJ, Lemmon MA, Sliwkowski MX, Ward CW and Yokoyama S. *An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors* (2003). *Mol Cell*, 12**,** 541-52.

Cahill DP, Kinzler KW, Vogelstein B and Lengauer C. *Genetic instability and darwinian selection in tumours* (1999). *Trends Cell Biol*, 9**,** M57-60.

Cai WW, Mao JH, Chow CW, Damani S, Balmain A and Bradley A. *Genome-wide detection of chromosomal imbalances in tumors using BAC microarrays* (2002). *Nat Biotechnol*, 20**,** 393-6.

Callagy G, Cattaneo E, Daigo Y, Happerfield L, Bobrow LG, Pharoah PD and Caldas C. *Molecular classification of breast carcinomas using tissue microarrays* (2003). *Diagn Mol Pathol*, 12**,** 27-34.

Calza S, Hall P, Auer G, Bjohle J, Klaar S, Kronenwett U, Liu ET, Miller L, Ploner A, Smeds J, Bergh J and Pawitan Y. *Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients* (2006). *Breast Cancer Res*, 8**,** R34.

Carroll SM, DeRose ML, Kolman JL, Nonet GH, Kelly RE and Wahl GM. *Localization of a bidirectional DNA replication origin in the native locus and in episomally amplified murine adenosine deaminase loci* (1993). *Mol Cell Biol*, 13**,** 2971-81.

Carroll SM, Gaudray P, De Rose ML, Emery JF, Meinkoth JL, Nakkim E, Subler M, Von Hoff DD and Wahl GM. *Characterization of an episome produced in hamster cells that amplify a transfected CAD gene at high frequency: functional evidence for a mammalian replication origin* (1987). *Mol Cell Biol*, 7**,** 1740-50.

Carter P, Presta L, Gorman CM, Ridgway JB, Henner D, Wong WL, Rowland AM, Kotts C, Carver ME and Shepard HM. *Humanization of an anti-p185HER2 antibody for human cancer therapy* (1992). *Proc Natl Acad Sci U S A*, 89**,** 4285-9.

Carvalho B, Ouwerkerk E, Meijer GA and Ylstra B. *High resolution microarray comparative genomic hybridisation analysis using spotted oligonucleotides* (2004). *J Clin Pathol*, 57**,** 644-6.

Chan SM, Ermann J, Su L, Fathman CG and Utz PJ. *Protein microarrays for multiplex analysis of signal transduction pathways* (2004). *Nat Med*, 10**,** 1390-6.

Chen T, Jackson CR, Link A, Markey MP, Colligan BM, Douglass LE, Pemberton JO, Deddens JA, Graff JR and Carter JH. *Int7G24A variant of transforming growth factor-beta receptor type I is associated with invasive breast cancer* (2006). *Clin Cancer Res*, 12**,** 392-7.

Coleman WB and Tsongalis GJ. (2001). *The Molecular Basis of Human Cancer*. Humana Press Inc.

Coradini D and Daidone MG. *Biomolecular prognostic factors in breast cancer* (2004). *Curr Opin Obstet Gynecol*, 16**,** 49-55.

Courjal F, Cuny M, Rodriguez C, Louason G, Speiser P, Katsaros D, Tanner MM, Zeillinger R and Theillet C. *DNA amplifications at 20q13 and MDM2 define distinct subsets of evolved breast and ovarian tumours* (1996). *Br J Cancer*, 74**,** 1984-9.

Donovan J and Slingerland J. *Transforming growth factor-beta and breast cancer: Cell cycle arrest by transforming growth factor-beta and its disruption in cancer* (2000). *Breast Cancer Res*, 2**,** 116-24.

Duffy MJ. *Predictive markers in breast and other cancers: a review* (2005). *Clin Chem*, 51**,** 494-503.

Fehr M, Fink D, Haller U and Maag HC. (2006). *EVEscript (EgoneValidEscript) für Medzinstudierende zur Lehrveranstaltung Gynäkologie in Basel, Bern, Lausanne und Zürich*.

Folkman J. *Tumor angiogenesis: therapeutic implications* (1971). *N Engl J Med*, 285**,** 1182-6.

Frixen UH, Behrens J, Sachs M, Eberle G, Voss B, Warda A, Lochner D and Birchmeier W. *E-cadherin-mediated cell-cell adhesion prevents invasiveness of human carcinoma cells* (1991). *J Cell Biol*, 113**,** 173-85.

Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N and Stratton MR. *A census of human cancer genes* (2004). *Nat Rev Cancer*, 4**,** 177-83.

Fynan TM and Reiss M. *Resistance to inhibition of cell growth by transforming growth factor-beta and its role in oncogenesis* (1993). *Crit Rev Oncog*, 4**,** 493-540.

Garcia MJ, Pole JC, Chin SF, Teschendorff A, Naderi A, Ozdag H, Vias M, Kranjac T, Subkhankulova T, Paish C, Ellis I, Brenton JD, Edwards PA and Caldas C. *A 1 Mb minimal amplicon at 8p11-12 in breast cancer identifies new candidate oncogenes* (2005). *Oncogene*, 24**,** 5235-45.

Gronostajski RM. *Roles of the NFI/CTF gene family in transcription and development* (2000). *Gene*, 249**,** 31-45.

Haber JE and Debatisse M. *Gene amplification: yeast takes a turn* (2006). *Cell*, 125**,** 1237-40.

Hanahan D and Weinberg RA. *The hallmarks of cancer* (2000). *Cell*, 100**,** 57-70.

Heffelfinger SC, Yassin R, Miller MA and Lower EE. *Cyclin D1, retinoblastoma, p53, and Her2/neu protein expression in preinvasive breast pathologies: correlation with vascularity* (2000). *Pathobiology*, 68**,** 129-36.

Herbert BS, Wright WE and Shay JW. *Telomerase and breast cancer* (2001). *Breast Cancer Res*, 3**,** 146-9.

Hewitt M, Breen N and Devesa S. *Cancer prevalence and survivorship issues: analyses of the 1992 National Health Interview Survey* (1999). *J Natl Cancer Inst*, 91**,** 1480-6.

Hoos A, Stojadinovic A, Mastorides S, Urist MJ, Polsky D, Di Como CJ, Brennan MF and Cordon-Cardo C. *High Ki-67 proliferative index predicts disease specific survival in patients with high-risk soft tissue sarcomas* (2001). *Cancer*, 92**,** 869-74.

Hoos A, Stojadinovic A, Singh B, Dudas ME, Leung DH, Shaha AR, Shah JP, Brennan MF, Cordon-Cardo C and Ghossein R. *Clinical significance of molecular expression profiles of Hurthle cell tumors of the thyroid gland analyzed via tissue microarrays* (2002). *Am J Pathol*, 160**,** 175-83.

Huang J, Wei W, Zhang J, Liu G, Bignell GR, Stratton MR, Futreal PA, Wooster R, Jones KW and Shapero MH. *Whole genome DNA copy number changes identified by high density oligonucleotide arrays* (2004). *Hum Genomics*, 1**,** 287-99.

Humphreys RC and Hennighausen L. *Transforming growth factor alpha and mouse models of human breast cancer* (2000). *Oncogene*, 19**,** 1085-91.

Hyman E, Kauraniemi P, Hautaniemi S, Wolf M, Mousses S, Rozenblum E, Ringner M, Sauter G, Monni O, Elkahloun A, Kallioniemi OP and Kallioniemi A. *Impact of DNA amplification on gene expression patterns in breast cancer* (2002). *Cancer Res*, 62**,** 6240-5.

Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F and Pinkel D. *Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors* (1992). *Science*, 258**,** 818-21.

Kang Y and Massague J. *Epithelial-mesenchymal transitions: twist in development and metastasis* (2004). *Cell*, 118**,** 277-9.

Karlseder J, Zeillinger R, Schneeberger C, Czerwenka K, Speiser P, Kubista E, Birnbaum D, Gaudray P and Theillet C. *Patterns of DNA amplification at band q13 of chromosome 11 in human breast cancer* (1994). *Genes Chromosomes Cancer*, 9**,** 42-8.

Kawai H, Li H, Chun P, Avraham S and Avraham HK. *Direct interaction between BRCA1 and the estrogen receptor regulates vascular endothelial growth factor (VEGF) transcription and secretion in breast cancer cells* (2002). *Oncogene*, 21**,** 7730-9.

Keyomarsi K, Tucker SL, Buchholz TA, Callister M, Ding Y, Hortobagyi GN, Bedrosian I, Knickerbocker C, Toyofuku W, Lowe M, Herliczek TW and Bacus SS. *Cyclin E and survival in patients with breast cancer* (2002). *N Engl J Med*, 347**,** 1566-75.

Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G and Kallioniemi OP. *Tissue microarrays for high-throughput molecular profiling of tumor specimens* (1998). *Nat Med*, 4**,** 844-7.

Korsching E, Packeisen J, Agelopoulos K, Eisenacher M, Voss R, Isola J, van Diest PJ, Brandt B, Boecker W and Buerger H. *Cytogenetic alterations and cytokeratin expression patterns in breast cancer: integrating a new model of breast differentiation into cytogenetic pathways of breast carcinogenesis* (2002). *Lab Invest*, 82**,** 1525-33.

Kossakowska AE, Huchcroft SA, Urbanski SJ and Edwards DR. *Comparative analysis of the expression patterns of metalloproteinases and their inhibitors in breast neoplasia, sporadic colorectal neoplasia, pulmonary carcinomas and malignant non-Hodgkin's lymphomas in humans* (1996). *Br J Cancer*, 73**,** 1401-8.

Kurose K, Zhou XP, Araki T, Cannistra SA, Maher ER and Eng C. *Frequent loss of PTEN expression is linked to elevated phosphorylated Akt levels, but not associated with p27 and cyclin D1 expression, in primary epithelial ovarian carcinomas* (2001). *Am J Pathol*, 158**,** 2097-106.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, et al. *Initial sequencing and analysis of the human genome* (2001). *Nature*, 409**,** 860-921.

Larue L, Ohsugi M, Hirchenhain J and Kemler R. *E-cadherin null mutant embryos fail to form a trophectoderm epithelium* (1994). *Proc Natl Acad Sci U S A*, 91**,** 8263-7.

Li C and Hung Wong W. *Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application* (2001). *Genome Biol*, 2**,** RESEARCH0032.

Li C and Wong WH. *Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection* (2001). *Proc Natl Acad Sci U S A*, 98**,** 31-6.

Liotta LA, Espina V, Mehta AI, Calvert V, Rosenblatt K, Geho D, Munson PJ, Young L, Wulfkuhle J and Petricoin EF, 3rd. *Protein microarrays: meeting analytical challenges for clinical applications* (2003). *Cancer Cell*, 3**,** 317-25.

Livingstone LR, White A, Sprouse J, Livanos E, Jacks T and Tlsty TD. *Altered cell cycle arrest and gene amplification potential accompany loss of wild-type p53* (1992). *Cell*, 70**,** 923-35.

McClintock B. *The Fusion of Broken Ends of Chromosomes Following Nuclear Fusion* (1942). *Proc Natl Acad Sci U S A*, 28**,** 458-63.

Moriya T, Kasajima A, Ishida K, Kariya Y, Akahira J, Endoh M, Watanabe M and Sasano H. *New trends of immunohistochemistry for making differential diagnosis of breast lesions* (2006). *Med Mol Morphol*, 39**,** 8-13.

Mosse YP, Greshock J, Weber BL and Maris JM. *Measurement and relevance of neuroblastoma DNA copy number changes in the post-genome era* (2005). *Cancer Lett*, 228**,** 83-90.

Mueller C, Riese U, Kosmehl H, Dahse R, Claussen U and Ernst G. *Telomerase activity in microdissected human breast cancer tissues: association with p53, p21 and outcome* (2002). *Int J Oncol*, 20**,** 385-90.

Myllykangas S, Himberg J, Bohling T, Nagy B, Hollmen J and Knuutila S. *DNA copy number amplification profiling of human neoplasms* (2006). *Oncogene*.

Myllykangas S and Knuutila S. *Manifestation, mechanisms and mysteries of gene amplifications* (2006). *Cancer Lett*, 232**,** 79-89.

Nannya Y, Sanada M, Nakazaki K, Hosoya N, Wang L, Hangaishi A, Kurokawa M, Chiba S, Bailey DK, Kennedy GC and Ogawa S. *A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays* (2005). *Cancer Res*, 65**,** 6071-9.

Narayanan V, Mieczkowski PA, Kim HM, Petes TD and Lobachev KS. *The pattern of gene amplification is determined by the chromosomal location of hairpin-capped breaks* (2006). *Cell*, 125**,** 1283-96.

Nishizuka S, Charboneau L, Young L, Major S, Reinhold WC, Waltham M, Kouros-Mehr H, Bussey KJ, Lee JK, Espina V, Munson PJ, Petricoin E, 3rd, Liotta LA and Weinstein JN. *Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays* (2003). *Proc Natl Acad Sci U S A*, 100**,** 14229-34.

Nocito A, Bubendorf L, Tinner EM, Suess K, Wagner U, Forster T, Kononen J, Fijan A, Bruderer J, Schmid U, Ackermann D, Maurer R, Alund G, Knonagel H, Rist M, Anabitarte M, Hering F, Hardmeier T, Schoenenberger AJ, Flury R, et al. *Microarrays of bladder cancer tissue are highly representative of proliferation index and histological grade* (2001). *J Pathol*, 194**,** 349-57.

Osborne RJ, Merlo GR, Mitsudomi T, Venesio T, Liscia DS, Cappa AP, Chiba I, Takahashi T, Nau MM, Callahan R and et al. *Mutations in the p53 gene in primary human breast cancers* (1991). *Cancer Res*, 51**,** 6194-8.

Parkin DM. *Global cancer statistics in the year 2000* (2001). *Lancet Oncol*, 2**,** 533-43.

Parkin DM, Bray F, Ferlay J and Pisani P. *Global cancer statistics, 2002* (2005). *CA Cancer J Clin*, 55**,** 74-108.

Pavlickova P, Schneider EM and Hug H. *Advances in recombinant antibody microarrays* (2004). *Clin Chim Acta*, 343**,** 17-35.

Paweletz CP, Charboneau L, Bichsel VE, Simone NL, Chen T, Gillespie JW, Emmert-Buck MR, Roth MJ, Petricoin IE and Liotta LA. *Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front* (2001). *Oncogene*, 20**,** 1981-9.

Pawitan Y, Bjohle J, Amler L, Borg AL, Egyhazi S, Hall P, Han X, Holmberg L, Huang F, Klaar S, Liu ET, Miller L, Nordgren H, Ploner A, Sandelin K, Shaw PM, Smeds J, Skoog L, Wedren S and Bergh J. *Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts* (2005). *Breast Cancer Res*, 7**,** R953-64.

Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO and Botstein D. *Molecular portraits of human breast tumours* (2000). *Nature*, 406**,** 747-52.

Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW and Albertson DG. *High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays* (1998). *Nat Genet*, 20**,** 207-11.

Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D and Brown PO. *Genome-wide analysis of DNA copy-number changes using cDNA microarrays* (1999). *Nat Genet*, 23**,** 41-6.

Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, Tibshirani R, Botstein D, Borresen-Dale AL and Brown PO. *Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors* (2002). *Proc Natl Acad Sci U S A*, 99**,** 12963-8.

Rubens JR, Lewandrowski KB, Kopans DB, Koerner FC, Hall DA and McCarthy KA. *Medullary carcinoma of the breast. Overdiagnosis of a prognostically favorable neoplasm* (1990). *Arch Surg*, 125**,** 601-4.

Sabbir MG, Roy A, Mandal S, Dam A, Roychoudhury S and Panda CK. *Deletion mapping of chromosome 13q in head and neck squamous cell carcinoma in Indian patients: correlation with prognosis of the tumour* (2006). *Int J Exp Pathol*, 87**,** 151-61.

Sasco AJ, Kaaks R and Little RE. *Breast cancer: occurrence, risk factors and hormone metabolism* (2003). *Expert Rev Anticancer Ther*, 3**,** 546-62.

Sauter G, Simon R and Hillan K. *Tissue microarrays in drug discovery* (2003). *Nat Rev Drug Discov*, 2**,** 962-72.

Schoenberg Fejzo M and Slamon DJ. *Frozen tumor tissue microarray technology for analysis of tumor RNA, DNA, and proteins* (2001). *Am J Pathol*, 159**,** 1645-50.

Schroeder JA and Lee DC. *Transgenic mice reveal roles for TGFalpha and EGF receptor in mammary gland development and neoplasia* (1997). *J Mammary Gland Biol Neoplasia*, 2**,** 119-29.

Schwab M. *Oncogene amplification in solid tumors* (1999). *Semin Cancer Biol*, 9**,** 319-25.

Shpitz B, Zimlichman S, Zemer R, Bomstein Y, Zehavi T, Liverant S, Bernehim J, Kaufman Z, Klein E, Shapira Y and Klein A. *Telomerase activity in ductal carcinoma in situ of the breast* (1999). *Breast Cancer Res Treat*, 58**,** 65-9.

Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, Fleming T, Eiermann W, Wolter J, Pegram M, Baselga J and Norton L. *Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2* (2001). *N Engl J Med*, 344**,** 783-92.

Sledge GW, Jr. and Miller KD. *Exploiting the hallmarks of cancer: the future conquest of breast cancer* (2003). *Eur J Cancer*, 39**,** 1668-75.

Sorlie T. *Molecular portraits of breast cancer: tumour subtypes as distinct disease entities* (2004). *Eur J Cancer*, 40**,** 2667-75.

Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lonning P and Borresen-Dale AL. *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications* (2001). *Proc Natl Acad Sci U S A*, 98**,** 10869-74.

Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lonning PE, Brown PO, Borresen-Dale AL and Botstein D. *Repeated observation of breast tumor subtypes in independent gene expression data sets* (2003). *Proc Natl Acad Sci U S A*, 100**,** 8418-23.

Spradling AC and Mahowald AP. *Amplification of genes for chorion proteins during oogenesis in Drosophila melanogaster* (1980). *Proc Natl Acad Sci U S A*, 77**,** 1096-100.

Takeichi M. *Cadherins: a molecular family important in selective cell-cell adhesion* (1990). *Annu Rev Biochem*, 59**,** 237-52.

Tan P, Cady B, Wanner M, Worland P, Cukor B, Magi-Galluzzi C, Lavin P, Draetta G, Pagano M and Loda M. *The cell cycle inhibitor p27 is an independent prognostic marker in small (T1a,b) invasive breast carcinomas* (1997). *Cancer Res*, 57**,** 1259-63.

Thor AD, Liu S, Moore DH, 2nd, Shi Q and Edgerton SM. *p(21WAF1/CIP1) expression in breast cancers: associations with p53 and outcome* (2000). *Breast Cancer Res Treat*, 61**,** 33-43.

Torhorst J, Bucher C, Kononen J, Haas P, Zuber M, Kochli OR, Mross F, Dieterich H, Moch H, Mihatsch M, Kallioniemi OP and Sauter G. *Tissue microarrays for rapid linking of molecular changes to clinical endpoints* (2001). *Am J Pathol*, 159**,** 2249-56.

Tsihlias J, Kapusta L and Slingerland J. *The prognostic significance of altered cyclin-dependent kinase inhibitors in human cancer* (1999). *Annu Rev Med*, 50**,** 401-23.

Vainio H and Bianchini F. (2002). *Breast cancer screening - IARC handbooks of cancer prevention, Vol. 7*. Iarcpress (ed.). IARCPress.

van de Rijn M, Perou CM, Tibshirani R, Haas P, Kallioniemi O, Kononen J, Torhorst J, Sauter G, Zuber M, Kochli OR, Mross F, Dieterich H, Seitz R, Ross D, Botstein D and Brown P. *Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome* (2002). *Am J Pathol*, 161**,** 1991-6.

Venkatraman ES and Olshen A. (2004).

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, et al. *The sequence of the human genome* (2001). *Science*, 291**,** 1304-51.

Vogelstein B and Kinzler KW. *Cancer genes and the pathways they control* (2004). *Nat Med*, 10**,** 789-99.

Wolf M, Mousses S, Hautaniemi S, Karhu R, Huusko P, Allinen M, Elkahloun A, Monni O, Chen Y, Kallioniemi A and Kallioniemi OP. *High-resolution analysis of gene copy number alterations in human prostate cancer using CGH on cDNA microarrays: impact of copy number on gene expression* (2004). *Neoplasia*, 6**,** 240-7.

Yang ZQ, Imoto I, Pimkhaokham A, Shimada Y, Sasaki K, Oka M and Inazawa J. *A novel amplicon at 9p23 - 24 in squamous cell carcinoma of the esophagus that lies proximal to GASC1 and harbors NFIB* (2001). *Jpn J Cancer Res*, 92**,** 423-8.

Yin Y, Tainsky MA, Bischoff FZ, Strong LC and Wahl GM. *Wild-type p53 restores cell cycle control and inhibits gene amplification in cells with mutant p53 alleles* (1992). *Cell*, 70**,** 937-48.

Zucker S, Lysik RM, Zarrabi MH and Moll U. *M(r) 92,000 type IV collagenase is increased in plasma of patients with colon cancer and breast cancer* (1993). *Cancer Res*, 53**,** 140-6.

5. Appendix

5.1. Evaluation of reverse phase protein microarrays (lysate microarrays)

5.1.1. Introduction

Reverse phase protein microarrays (also called lysate microarrays) were first introduced in 2001 by Paweletz et al (Paweletz et al., 2001). Other authors have further developed this relatively new technique (Chan et al., 2004; Nishizuka et al., 2003). The principal aim was to establish a technique for the screening of molecular markers on the proteomic level. The principle of these microarrays involves a process of spotting cell extracts (protein lysates) from a large number of biological samples on a coated glass slide and subsequently probing the array with a large number of antibodies. The very low volume used for such an array spot (circa 0.3nL) allows the spotting of samples in several replicates in serial dilutions. This, in turn, does not only allow for the detection of the linear and dynamic range of binding of the used antibody, but also for the statistical evaluation of the retrieved data. Theoretically, up to 500 different tumor samples can be spotted on one glass slide. In order to circumvent the limitations of the tissue microarrays, such as the limited availability of antibodies suited for paraffin-embedded tissues, we wanted to establish the reverse phase protein microarray technology. We decided to use a dual-colour labelling and detection system in order to assess both, the protein of interest (e.g. HER-2) and the reference protein (e.g. GAPDH), simultaneously on the same array spot.

5.1.2. Establishment of the best settings

We selected a Cy3/Cy5 fluorescent labelling and detection system. In order to determine the best settings for the reverse phase protein microarray technology for our use, we spotted several tissue and cell line lysates onto different commercially available slides. Unfortunately, the protein binding capacity of almost all of the purchased slides was not sufficient for further use in a quantitative fashion (Figure S1 A left). Only the nitrocellulose coated slides from Schleicher&Schüll could be considered

further (Figure S1 A right), since the three-dimensional polymer coating of the nitrocellulose led to a higher surface area and thus to a higher binding capacity. Similar observations have been reported by other researchers ((Chan et al., 2004; Paweletz et al., 2001), Kallioniemi et al, personal communication). Optimizations of the environmental settings (e.g. humidity, spotting buffer used, pin diameter) enabled the production of nearly perfectly round spots, which are prerequisite for automated signal determination (Figure S1 B). The optimal spot morphology was confirmed by the achieved signal linearity of the dilution series (Figure S1 C). The linear region spanned at least five to six two-fold dilutions (dilutions A1-A5, Figure S1 C), i.e. a 32-64 fold dilution range. This was accomplished either by using a GAPDH antibody in conjunction with a Cy5-coupled secondary antibody, or by using a total protein detection system that fluoresces near the Cy3 wavelength. Unfortunately, the usage of Cy3-coupled secondary antibodies was impracticable due to the nitrocellulose of the slides used which revealed an increasing background on more shorten wavelengths. The Cy3 system with a wavelength of around 560 nm (in contrast to Cy5 with 660nm), was strongly affected by this background. These findings made the use of a system in which both, the protein of interest and the reference protein were detected with Cy3-, respectively Cy5-coupled antibodies, impracticable. Hence, we decided to use the Cy3 channel only as reference for the determination of the total protein amount spotted. To this aim, the Deep Purple total protein detection (Amersham Biosciences) was selected. This procedure was feasible since the total protein amount was several times higher than the amount of a specific reference protein, and therefore above background and thus detectable. Obviously, the simultaneous detection of the protein of interest on the same slide was not possible any more.

5.1.3. The Her-2 experiment

In order to test our established reverse phase protein microarray system, we lysed 20 breast cancer tissue samples and four tumor cell lines. The samples were spotted in four two-fold dilutions on replicates (Figure S1 D). Deep Purple total protein detection system was used for the determination of total protein amount spotted. HER-2 was chosen as protein of interest and its expression was determined by usage of an anti-HER-2 antibody and a Cy5-coupled secondary antibody. Detection of

the spots and its signal quantification were done automatically by the scanner and its corresponding software. The two used channels (Cy3 and Cy5) were acquired separately, and the two images were overlaid, as shown in Figure S1 D. The Her-2 expression statuses of these samples were first determined independently from this study by using the commercially available Her-2 diagnostic kit (HercepTest, DAKO). Its scoring system ranges between 0 (no Her-2 expression) and 3+ (very high Her-2 protein expression). Two samples were classified as 3+ (sample 19 and the cell line SKBR3), one sample as 2+ (sample 11) and four samples (sample 15, 18 and the cell lines MCF7 and T47D) as 1+. Using the total protein as reference, we determined the expression status of each of the spotted samples. The summary is shown as bar graph in Figure S1 E. The 3+ and 2+ positive samples could undoubtedly be detected in all of the cases.

5.1.4. Discussion

With this project, we showed the semi-successful establishment of the relatively new technology of the reverse phase protein microarrays. Several arising issues, like spot morphology, optimal environmental settings, determination of the linear range, could be solved. The experiment shown in Figure S1 D-E revealed that our system was capable to recognize and detect high expressing proteins, like Her-2 2+ and 3+ overexpressing samples, even when the data acquisition and evaluation was done in an automated manner. Nevertheless, we did not have the ability to detect the Her-2 1+ expressing samples, i.e. our system was not as sensitive as expected. We suggest that is due to the lack of a simultaneous incubation and detection system, in which both channels (Cy3 and Cy5) can be used. Since the FAST slides seem the most suited slides for the spotting of proteins, other investigators, which work with reverse phase protein microarrays, were confronted with similar problems. Chan et al (Chan et al., 2004) showed an interesting approach to handle this issue: they used a tyramide signal amplification system for the amplification of the Cy3 signal. This led to an usable signal above background, but also to a smaller linear range. Kallioniemi et al (personal communication) decided to replace the Cy3 detection with an infrared detection system. Since the wavelength of infrared signal is longer (> 700nm), the background diminishes and it can even be used as a double labelling and

detection system with the Cy5 channel (~650nm). Anyway, the acquisition of a scanner feasible for Cy5 and infrared signal is associated with tremendous costs. Finally, it can be summarized, that the idea of the reverse phase protein microarrays for screening of molecular markers on proteomic level, was very promising. But several issues, especially concerning the detection system have to be solved, first. The use of an infrared detection system is the first step in the right direction, but is accompanied be extremely high costs.

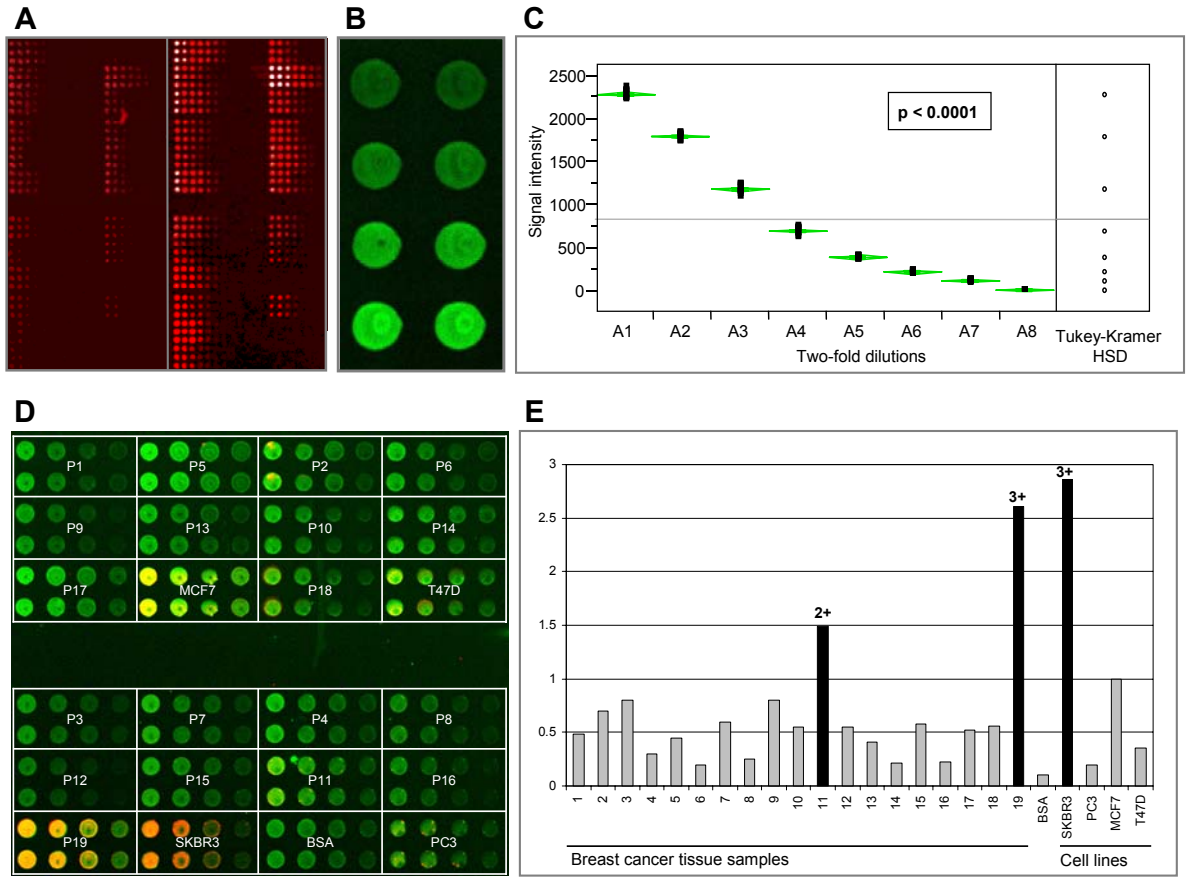


Figure S1. Reverse phase protein microarray experiments. A) shows the tremendous effect of the substratum on protein binding capacity. Left: epoxy-coated surface; right: nitrocellulose coated surface. B) Optimal parameter settings (humidity, buffer, substratum) leads to nearby perfect spot morphology. C) shows the linear range of a spotted dilution serie. The first five two-fold dilutions (A1-A5) can be used for data analysis. D) shows the Her-2 experiment. Image overlay of Deep Purple total protein detection (Cy3, green) and HER-2 antibody determination using a Cy5-coupled secondary antibody (red). E) shows the data analysis of the Her-2 experiment (D): Only the cell lines and samples that have an DAKO IHC score of 2+ or 3+ were detected as positive samples.

Curriculum Vitae

Christian Ruiz
Klybeckstrasse 251
CH - 4057 Basel

phone: +41 61 631 03 19
mobile: +41 76 435 54 53
email: christian.ruiz@unibas.ch

nationalities: Spanish
date of birth: 16.04.1977

EDUCATION

2004-2007	**Graduate (PhD) Program of the Biozentrum (University of Basel)** Research group of Prof. Dr. Guido Sauter Performed in the Department of Molecular Pathology, Institute of Pathology, Basel. Identification and validation of novel amplification target genes in human breast cancer.
2003-2004	**Postgraduate studies in informatics (Executive Master of Information Technology)** University of Applied Sciences (Fachhochschule beider Basel FHBB)
1998-2002	**Diploma in Cell and Molecular Biology (Master of Science)** Biozentrum, University of Basel
1996-1998	**First Propaedeutikum, human medicine** Faculty of Medicine, University of Basel
1996	**University entry diploma** Gymnasium Bäumlihof

UNIVERSITY LECTURES ATTENDED DURING PHD

BioinformaticsI (Prof. Schwede, Prof. Primig, Prof. van Nimwegen, Prof. Zavolan)

Excercises in Bioinformatics I (Prof. Schwede)

Bioinformatics II (Prof. Schwede, Prof. Primig, Prof. van Nimwegen, Prof. Zavolan)

Excercises in Bioinformatics II (Prof. Schwede)

Data Acquisition and Knowledge Representation (Prof. Schwede, Prof. Primig)

New literature in cell-cell and cell-extracellular matrix adhesion (Prof. Chiquet, Prof. Christofori)

Extracellular Matrix and Cell Adhesion (Prof. Chiquet, Prof. Rüegg)