Biomarker Discovery and Pathway Activation Profiling in Breast Cancer using Protein Microarrays

Von der Fakultät Energie-, Verfahrens- und Biotechnik
der Universität Stuttgart zur Erlangung der Würde
eines Doktors der Naturwissenschaften (Dr. rer. nat.)
genehmigte Abhandlung

Vorgelegt von
Johanna Sigrid Nelly Sonntag
aus Heilbronn-Neckargartach

Hauptberichter: Prof. Dr. Roland Kontermann
Mitberichter: PD Dr. Stefan Wiemann

Tag der mündlichen Prüfung: 13.06.2013

Institut für Zellbiologie und Immunologie
Universität Stuttgart
2013

„It is far more important to know what person the disease has than what disease the person has."

Hippocrates

Table of Contents

Abbreviations

APS	ammonium persulfate
AREG	amphiregulin
AUC	area under the curve
BTC	betacellulin
BSA	bovine serum albumin
BC	breast cancer
CTC	circulating tumor cell
CV	coefficient of variation
Da	dalton
DNA	deoxyribonucleic acid
DKFZ	Deutsches Krebsforschungszentrum
DTT	dithiothreitol
DCIS	ductal carcinoma in situ
ELISA	enzyme-linked immunosorbent assay
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
EMT	epithelial to mesenchymal transition
ERα	estrogen receptor alpha
EDTA	ethylenediaminetetraacetic acid
FBS	fetal bovine serum
FISH	fluorescence in situ hybridization
FDA	Food and Drug Administration
FFPE	formalin-fixed paraffin-embedded
gal	gene pix array list file
gpr	gene pix result file
GGI	genomic grade index
HBEGF	heparin-binding EGF-like growth factor
HGF	hepatocyte growth factor
HRG	heregulin 1-β
G1	histologic grade 1
G2	histologic grade 2
G3	histologic grade 3
HER2	human epidermal growth factor receptor 2
IgG	immunoglobulin G
IHC	immunohistochemistry
IDC	invasive ductal carcinoma
ILC	invasive lobular carcinoma
LCM	laser capture microdissection
LLOD	lower limit of detection
M-PER	mammalian protein extraction reagent
mRNA	messenger RNA
miRNA	micro RNA

MIA	microspot immunoassay
PBS	phosphate buffered saline
PBST	phosphate buffered saline + 0.1% Tween20®
PAGE	polyacrylamide gel electrophoresis
PVDF	polyvinylidene difluoride
PAM	prediction analysis for microarrays utilizing the nearest shrunken centroid classifier
PR	progesterone receptor
RF-Boruta	random forests with Boruta algorithm for feature selection
RTK	receptor tyrosine kinase
RPPA	reverse phase protein array
RNA	ribonucleic acid
R2LC	RPPA Risk Linear Classification
SDS	sodium dodecyl sulfate
SD	standard deviation
SOP	standard operating procedure
SCAD-SVM	support vector machines using smoothly clipped absolute deviation penalty
TIFF	tagged image file format
TEMED	tetramethylethylenediamine
T-PER	tissue protein extraction reagent
TGFα	transforming growth factor α
TNBC	triple-negative breast cancer
Tris	Tris(hydroxymethyl)aminomethane
TBS	Tris-buffered saline
TBST	Tris-buffered saline + 0.1% Tween20®
TNM	tumor size, lymph node status, distant metastasis
ULOD	upper limit of detection
VEGF	vascular endothelial growth factor

Preface

This dissertation was accomplished at the Division of Molecular Genome Analysis of the German Cancer Research Center (DKFZ) Heidelberg from February 2009 to February 2013.

The work was supported by the Medical Systems Biology initiative (Project: BreastSys) and by the National Genome Research Network (Project: Cellular Systems Genomics in Health and Disease) of the German Federal Ministry of Education and Research. Within the frame of these collaboration projects, tumor samples, blood samples, and corresponding clinical information were provided by the Institute of Pathology at Heidelberg University and the National Center for Tumor Diseases Heidelberg. Dr. Christian Bender and Dr. Annika Jöcker (Division of Molecular Genome Analysis, DKFZ) developed the R-package *bootfs*/R2LC score and the software program QuantProReloaded, respectively. Genome-wide mRNA expression profiling was performed at the DKFZ Proteomics and Genomics Core Facility and data was analyzed by Silvia von der Heyde (University Medical Center Göttingen).

Abstract

Breast cancer, the most frequent cancer entity among women, is nowadays widely recognized as a heterogeneous disease in terms of histopathology as well as on the molecular level. Over the last few years, gene expression profiling studies have improved our understanding of the underlying molecular mechanisms associated with the very heterogeneous outcomes of breast cancer patients. The existence of intrinsic molecular subtypes, which are linked to unique biological and prognostic features, was repeatedly demonstrated and points to the need of tailored therapy options. However on the functional level, breast cancer is not only a genomic but mainly a proteomic disease and gene expression profiling might provide only limited insights.

Following the hypothesis that intrinsic biologic features of breast tumors affect prognosis and also therapy response, the general aim of this thesis was to further explore breast cancer heterogeneity with protein microarrays on the functional proteomics level.

Around 70 – 80% of all breast cancer patients belong to the luminal intrinsic molecular subtype, characterized as a surrogate marker by overexpression of hormone receptors. An improved classification of this subtype is crucial for therapy decision as part of the patients are at higher risk of recurrence requiring chemo-endocrine treatment, whereas the other part is at lower risk and does not benefit from chemotherapy. However, accurate definition of low and high risk hormone receptor-positive breast cancer has remained a challenge so far. Thus, the first objective of this thesis was the identification of a robust and quantitative protein biomarker signature to facilitate risk classification of hormone receptor-positive breast cancer. To approach this aim reverse phase protein arrays were used to screen across over 120 breast cancer relevant proteins and a novel bioinformatics workflow for biomarker hit selection was applied. Using this approach, a biomarker signature consisting of caveolin-1, NDKA, RPS6, and Ki-67, was identified as most promising to distinguish between low and high risk hormone receptor-positive breast cancer.

Since genomic and transcriptomic profiling alone cannot sufficiently predict protein pathway activation, it is important to explore and define the heterogeneity of hormone receptor-positive breast cancer on the proteome level. Especially as protein signaling pathways present the direct targets of new classes of therapeutics. Thus, the second thesis objective addressed the question of whether hormone receptor-positive breast cancer can be further categorized according to similar signaling pathway activation patterns and whether these patterns reflect common molecular mechanisms. Therefore, comprehensive

protein pathway activation profiles of breast cancer specimens were obtained using reverse phase protein arrays. To complement this analysis, a microspot immunoassay was developed, which enabled the simultaneous quantification of eight different growth factors in tumor lysate as well as blood plasma of matching patient samples. Four subgroups were identified, based on differential expression of 90 cancer-relevant signaling proteins. Each subgroup showed unique characteristics which were also related to established clinicopathological features as well as growth factor expression. One subgroup, for example, was characterized by high expression levels of almost all analyzed proteins. In addition, VEGF tumor lysate levels were significantly higher in this subgroup and an enrichment of poorly differentiated tumors was observed underlining the aggressive phenotype. In contrast, another subgroup was characterized by weak signaling activity. Interestingly, this subgroup was mostly associated with invasive lobular carcinoma, the second most common histologic type of breast cancer, reflecting a link between histopathology and underlying molecular mechanisms. In summary, the reverse phase protein array based pathway activation profiling of hormone receptor-positive breast cancer, presented in this thesis, provides a comprehensive snapshot of the heterogeneity of this subtype on the proteomic level. Insights obtained can serve as basis to refine the concept of clinically relevant subtypes towards an improved definition of patient-tailored therapy options.

Zusammenfassung

Brustkrebs ist die häufigste Krebsart bei Frauen und wird mittlerweile als heterogene Erkrankung wahrgenommen. Diese Heterogenität spiegelt sich in der Histopathologie wie auch auf molekularer Ebene wieder. In den letzten Jahren haben vor allem Genexpressionsstudien zum besseren Verständnis der molekularen Mechanismen, die für die unterschiedlichen Krankheitsverläufe der Patientinnen verantwortlich sind, beigetragen. Die Existenz von intrinsischen molekularen Subtypen konnte wiederholt bestätigt werden. Die unterschiedlichen biologischen und prognostische Eigenschaften dieser molekularen Subtypen deuten auf die Notwendigkeit von maßgeschneiderten Therapien hin. Jedoch ist Brustkrebs auf funktionaler Ebene vor allem eine proteomische Erkrankung. Genexpressionsstudien können daher wahrscheinlich nur limitierte Einblicke bieten.

Basierend auf der Hypothese, dass intrinsische biologische Eigenschaften der Tumore die Prognose sowie das Therapieansprechen beeinflussen, war das allgemeine Ziel dieser Dissertation, die Heterogenität von Brustkrebs auf funktionaler Proteome Ebene mittels Protein-Microarrays, besser zu verstehen.

Die Merheit aller Brustkrebs Patientinnen wird mit einem Tumor des luminalen Subtyps diagnostiziert, welcher hauptsächlich durch eine Überexpression von Hormonrezeptoren charakerisiert ist. Eine weitere Klassifizierung dieses Subtyps ist äußerst wichtig, um die Therapieentscheidung zu unterstützen. Bei Patientinnen mit einem hohen Rückfallrisiko ist zusätzlich zu einer antihormonellen Behandlung eine Chemotherapie indiziert, wohingegen Patientinnen mit einem niedrigen Rückfallriskio nicht von einer Chemotherapie profitieren. Jedoch stellt die präzise Unterscheidung zwischen niedrigem und hohem Rückfallrisiko zur Zeit noch eine Herausforderung dar. Der Fokus dieser Dissertation lag daher zunächst auf der Identifizierung einer robusten und quantitativen Biomarker Signatur, die zur verbesserten Risikoklassifizierung bei Hormonrezeptor-positivem Brustkrebs beitragen kann. Reverse Phase Protein-Arrays wurden verwendet, um in Kombination mit einem neu entwickelten Biomarker-Identifikationsprozess die optimale Biomarker Signatur aus über 120 brustkrebsrelevanten Proteinen zu bestimmen. Mit diesem Vorgehen konnte eine Biomarker Signatur, bestehend aus Caveolin-1, RPS6, NDKA und Ki-67, als am vielversprechendsten identifiziert werden, um zwischen einem niedrigen und hohen Rückfallrisiko zu unterscheiden.

Der zweite Teil dieser Dissertation befasste sich mit der Fragestellung, ob Hormonrezeptor-positiver Brustkrebs anhand von ähnlichen Signaltransduktionsaktivitätsmustern weiter

stratifiziert werden kann und ob diese Aktivitätsmuster gemeinsame molekulare Mechanismen wiederspiegeln. Es wurden umfassende Proteinaktivitätsprofile von Tumoren mittels Reverse Phase Protein-Arrays erstellt. Um die Analyse zu erweitern, wurde zusätzlich ein Microspot-Immunoassay entwickelt, welcher die parallele Quantifizierung von acht unterschiedlichen Wachstumsfaktoren in Tumor- und Blutplasmaproben ermöglichte. Anhand der Analyse von 90 verschiedenen krebsrelevanten Signalproteinen konnten vier unterschiedliche Gruppen identifiziert werden. Jede Gruppe zeigte charakteristische Signaltransduktionsaktivitätsmuster. Desweiteren konnten diese vier Gruppen mit etablierten pathologischen Faktoren sowie mit der Expression von Wachstumsfaktoren in Verbindung gebracht werden. Eine Gruppe zeigte zum Beispiel eine stark erhöhte Signalaktivität bei einer Vielzahl der analysierten Proteine. In dieser Gruppe konnte zusätzlich eine signifikant erhöhte VEGF Konzentration sowie eine Anreicherung von schlecht differenzierten Tumoren beobachtet werden, wodurch der aggressive Phänotyp noch weiter bestärkt wurde. Im Gegensatz dazu, zeigte eine andere Gruppe eher schwache Signalaktivität für eine Vielzahl der analysierten Proteine. Interessanterweise wurde in dieser Gruppe ein hoher Anteil an Tumoren vom Typ des invasiven lobulären Karzinoms beobachet. Diese Beobachtung macht die Verbindung von histophatologischen und molekularen Eigenschaften deutlich. Die durchgeführte Analyse der Signaltransduktions-aktivitätsmuster bei Hormonrezeptor-positivem Brustkrebs zeigt die Heterogenität dieses Subtyps auf Proteome Ebene. Die neu gewonnenen Einblicke können als Grundlage für die Verbesserung der Definition von klinisch relevanten Subtypen dienen und zur Entwicklung von personalisierten Therapieansätzen beitragen.

1. Introduction

1.1 Breast anatomy and physiology

The female breast is a mass of glandular and connective tissue positioned over the pectoralis muscles on top of the ribcage. The primary function is to produce and secrete milk in order to feed the infant. While the main components of the connective tissue are adipose and fibrous tissue giving, the breast its size and shape, the components of the glandular tissue are ducts and lobes. These lobes are formed by groups of lobules, each composed of several alveoli (Figure 1). Once milk is produced in the alveoli, it is transported through a network of ducts to the nipple (Benninghoff and Drenckhahn 2004). The mammary gland undergoes highly dynamic changes throughout the different stages of life including embryonic development, pre-puberty, puberty, pregnancy, lactation, and involution. These changes are coordinated by hormones and growth factors. The most prominent hormones involved in these processes are estrogen, progesterone, and prolactin (Brisken and O'Malley 2010).

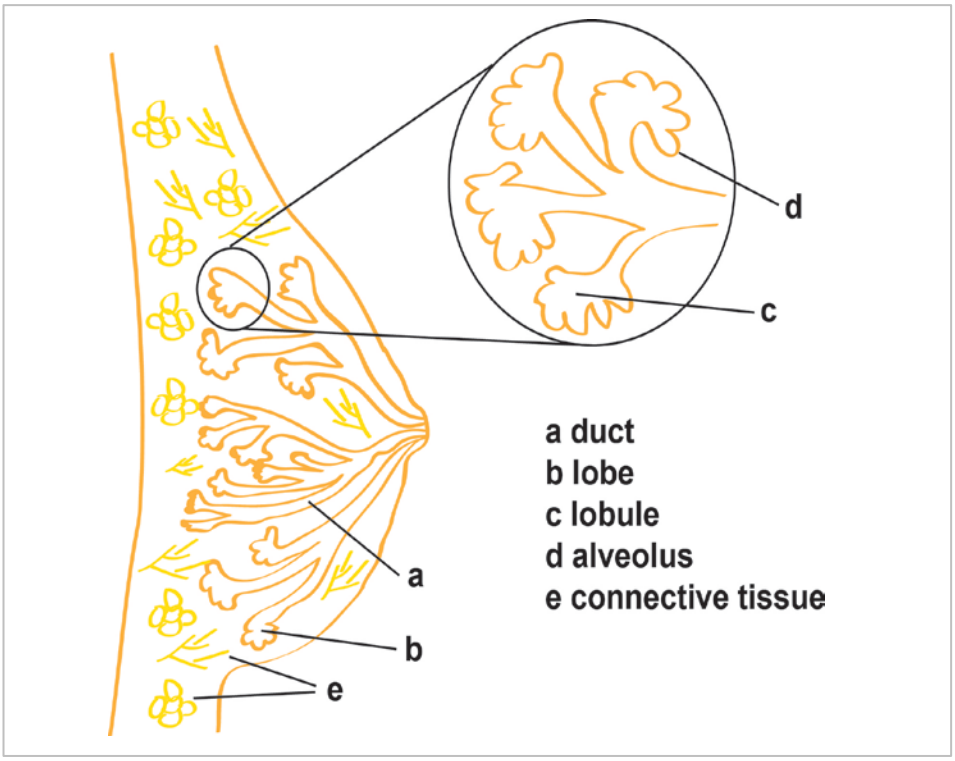


Figure 1: Anatomy of the female breast.

A schematic cross-section of a duct as well as a lobule is shown in Figure 2 and highlights the different cell types the mammary gland is composed of. The differentiation hierarchy of epithelial mammary cells starts with multipotent mammary stem cells. These mammary stem cells develop to specific progenitor cells either of the luminal or the myoepithelial lineage. The luminal progenitors differentiate further to ductal cells or alveolar cells, the latter with the capacity to secrete milk. The myoepithelial progenitors differentiate to myoepithelial cells, which are located between the luminal cells and the basement membrane. These cells have a contractile function and produce components to maintain the basement membrane. The basement membrane separates epithelial cells from the surrounding stroma (Visvader 2009, Van Keymeulen et al. 2011).

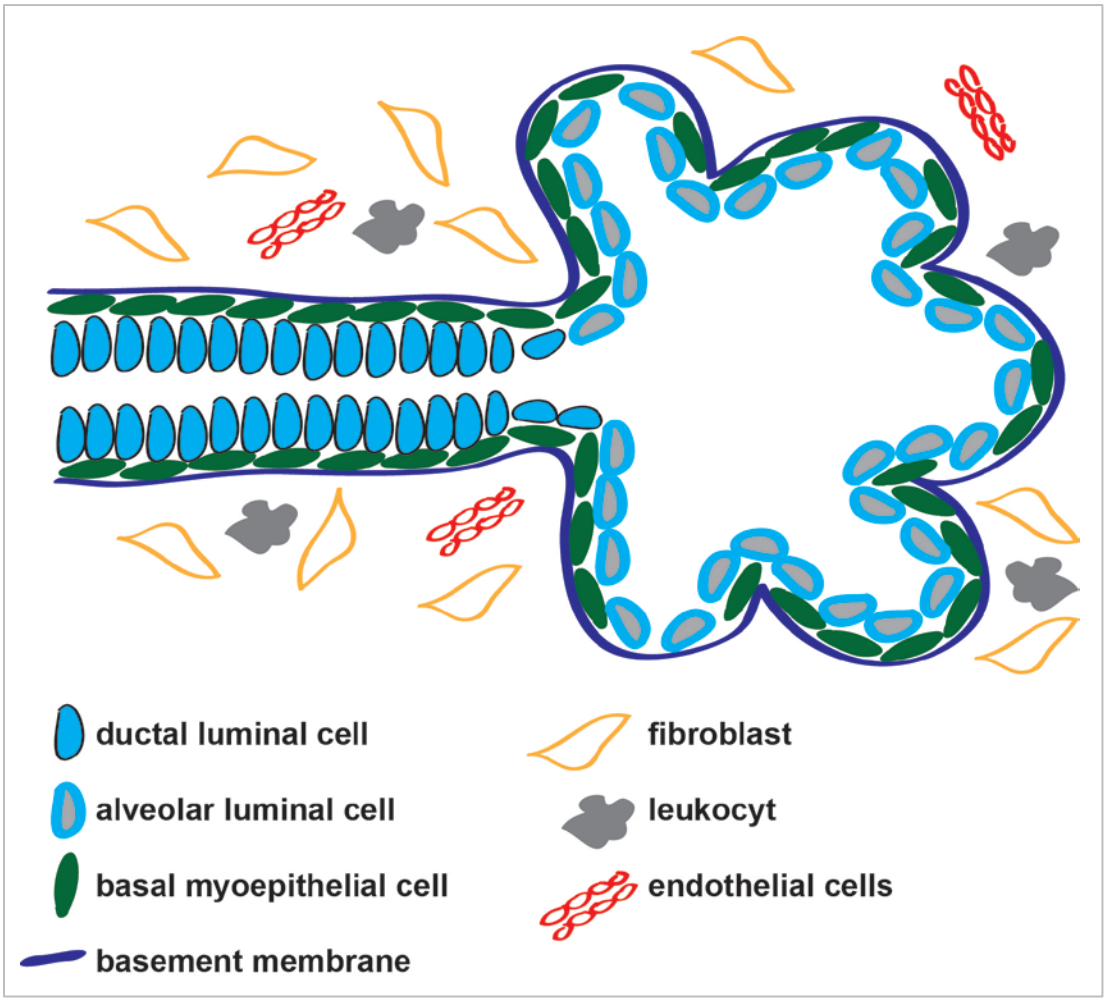


Figure 2: Cross-section of a duct and a lobule (modified after (Polyak and Kalluri 2010)).

1.2 Breast cancer

Breast cancer is the most frequent cancer among women accounting for approximately 30% of all cancer cases. One in eight women will be diagnosed with breast cancer during her life, whereas breast cancer in men is very rare with less than 1% of total breast cancer cases. Breast cancer incidence in Germany has doubled in 2008 compared to 1980 with 72,000 newly diagnosed cases. Nevertheless, breast cancer mortality is constantly decreasing mainly due to improvements in early detection and therapies options (Ferlay et al. 2010, Robert Koch-Institut and Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. 2012).

Besides life style (diet, obesity, physical activity, and alcohol), breast cancer risk factors are mostly related to differences in hormonal exposure during life time. High risk factors include early menarche, late menopause, nulliparity, late pregnancy and hormone replacement therapy (Robert Koch-Institut and Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. 2012). Another risk factor is familial history of breast cancer. In contrast to sporadic cancer, approximately 5 - 10 % of all breast cancer cases are due to inherited genetic predisposition (Fackenthal and Olopade 2007). The first gene which was identified to be associated with hereditary breast cancer was BRCA1 (Hall et al. 1990).

Breast cancer is a complex disorder and develops through accumulation of genetic alterations like chromosomal instability, epigenetic changes, copy number variation, translocation, and mutation, all having consequences on the functional proteomics level. During the multistep process of tumor development, the cancer cells acquire different capabilities known as the “hallmarks of cancer”: sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis (Hanahan and Weinberg 2000). In 2011, Hanahan and Weinberg added two further emerging hallmarks to their concept, which are reprogramming of energy metabolism and evading immune destruction. Furthermore they highlight the contribution of the tumor microenvironment to tumorigenesis (Hanahan and Weinberg 2011).

Breast cancer is increasingly recognized as a very heterogeneous disease in terms of tumor morphology as well as at the molecular level (Weigelt and Reis-Filho 2009, Curtis et al. 2012, Koboldt et al. 2012). Over the last few years, gene expression profiling has improved the understanding of the molecular mechanisms associated with the very heterogeneous

clinical outcome of breast cancer patients (Sotiriou and Pusztai 2009). The seminal work of Sorlie and Perou identified several intrinsic molecular subtypes with unique biological and prognostic features, termed luminal A, luminal B, basal-like, and HER2-enriched. The basal-like and HER-2 enriched subtypes were characterized by low to absent expression of luminal epithelial specific genes (e.g. ESR1 and GATA3). The HER2-enriched subtype showed in addition high expression levels of HER2 and several other genes like GRB7 located at the same amplified genomic region. The basal-like subtype was characterized by high expression of keratins 5 and 17, laminin, and fatty acid binding protein 7 (Perou et al. 2000, Sorlie et al. 2001). Clinical outcome, measured as time to distant metastasis, was best for luminal A, worst for basal-like and HER2-enriched as well as intermediate for the luminal B subtype (Sorlie et al. 2003). Recently, the existence of the breast cancer intrinsic molecular subtypes was confirmed in a large study including also other profiling methods like miRNA sequencing, DNA methylation chips, genomic DNA copy number arrays, and reverse phase protein arrays (Koboldt et al. 2012).

1.2.1 Breast cancer diagnosis

The process of breast cancer diagnosis starts when a suspicious lump is discovered in the breast by self-exam or if an abnormal area is detected by mammography during preventive medical checkup. After a diagnostic mammogram the next step is a diagnostic biopsy of the suspicious tissue. Based on histopathological analysis of this tissue sample the pathologist can decide whether the tumor is benign or malignant. If the tumor is malignant, it is further reported whether it is a non-invasive or an invasive form. In case of invasive breast cancer, tumor cells have already infiltrated the surrounding stroma by passing the basement membrane (Ellis et al. 2003). Further, the histologic type and the histologic grade are reported. The most common type of non-invasive breast cancer is ductal carcinoma in situ (DCIS) (Burstein et al. 2004). The highest prevalence of invasive breast cancer has invasive ductal carcinoma (IDC) of no special type with 50 - 80% followed by invasive lobular carcinoma (ILC) with 5-15% of all cases. Further histologic types include for example mucinous carcinoma, medullary carcinoma, and papillary carcinoma among several others (Ellis et al. 2003, Weigelt and Reis-Filho 2009). The assessment of histologic grade provides

information about tumor proliferation and differentiation status. The Nottingham Grading System is based on a semi-quantitative evaluation of morphologic tumor characteristics (tubule or gland formation, nuclear pleomorphism, and mitotic count). Based on the resulting score, tumor samples are assigned to three different categories, either well-differentiated (grade 1), moderately differentiated (grade 2) or poorly differentiated (grade 3) (Elston and Ellis 1991). In addition to histologic type and histologic grade, the expression level of at least three different biomarkers is routinely assessed using immunohistochemistry. These biomarkers are, in detail, estrogen receptor alpha (ERα), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2). Different scoring systems are used to classify the patient either as positive or negative for the respective biomarker. In case of ERα and PR, either the immunoreactive score (Remmele and Stegner 1987), the Allred score (Harvey et al. 1999) or most recently immunoreactivity of ≥ 1% tumor nuclei (Hammond et al. 2010) is used as cut-off. A positive HER2 status is defined by a positive staining result in > 30% of the tumor cells (stain 3+). HER2-negative samples are defined by staining of less than 10% of the tumor cells (stain 0/1+). In case of a staining result between 10% and 30% (stain 2+), an alternative method like fluorescence in situ hybridization (FISH) is done. If the average HER2 gene copy number is less than 4, patients are grouped into the HER2-negative group (Wolff et al. 2007).

After surgical removal of the tumor mass, assessment of histologic grade as well as ERα, PR, and HER2 status is repeated for confirmation. In addition, extend of lymphatic infiltration and residual cancer burden is reported. Finally, the pathological stage is determined using the TNM classification system. This classification system combines information about tumor size (T), lymph node status (N), and distant metastasis (M). Stage I and II are considered as early stage breast cancer, stage III as locally advanced breast cancer, and stage IV as metastatic breast cancer (Ellis et al. 2003, Kreienberg et al. 2012).

In summary, information about breast cancer diagnosis including tumor type, histologic grade, TNM classification, and residual cancer burden as well as expression status of ERα, PR, and HER2 is provided by the pathologist to the clinicians responsible for therapy decision. The impact of all these prognostic and predictive factors on therapy decision is discussed in the following section.

1.2.2 Breast cancer treatment

The choice of the appropriate standard therapy for patients with invasive breast cancer depends on the stage of the tumor and several other factors. The different stages comprise early, advanced, metastatic, and recurrent breast cancer as well as secondary breast cancer. The first treatment for early breast cancer will most likely be surgery followed by radiotherapy. Subsequently, the kind of systemic adjuvant therapy is decided based on defined clinicopathological features including tumor size, lymph node status, age, histologic grade, as well as few biomarkers like ERα, PR, and HER2 receptor. Breast cancer can be grouped roughly according to the receptor status into three therapeutic subgroups.

The first therapeutic subgroup is characterized by lack of overexpression of ERα, PR, and HER2 receptor and is therefore called triple-negative. About 15% of all breast cancer cases account for this subgroup (Bauer et al. 2007). Patients with triple-negative tumors are rather young and tumors are poorly differentiated. This subtype has in general a worse prognosis and is treated with adjuvant chemotherapy to reduce the risk of recurrence and mortality (Perez et al. 2010, Metzger-Filho et al. 2012). Chemotherapy regimens are for example combinations or sequences of anthracyclines (e.g. epirubicin), taxanes (e.g. paclitaxel), alkylating agents (e.g. cyclophosphamide), or anti-metabolites (e.g. 5-fluorouracil). Chemotherapeutic agents target highly proliferating cells by inhibition of cell proliferation and DNA replication. Besides targeting cancer cells, also normal cells with high proliferation rate are harmed. Mainly cells of the bone marrow, hair follicles, and digestive tracts resulting in immunosuppression, hair loss, and mucositis. In addition side-effects of chemotherapy frequently include fatigue, nausea, and vomiting.

The second therapeutic subgroup is defined by overexpression of the HER2 receptor. HER2 is overexpressed in about 10 - 20% of all breast cancer tumors and associates with worse prognosis of these patients (Slamon et al. 1987, Press et al. 1997, Heil et al. 2012). However, this subgroup benefits by incorporation of trastuzumab, a humanized monoclonal antibody directed against HER2, into adjuvant chemotherapy (Slamon et al. 2001, Baselga et al. 2006). Two years after approval of trastuzumab by the FDA for the treatment of metastatic breast cancer, it was approved in 2000 in the European Union for the same indication. Since 2006, trastuzumab is also approved for the treatment of HER2-positive non-metastatic breast cancer. Trastuzumab belongs to a new class of drugs called targeted therapeutics. In contrast to conventional chemotherapeutic drugs, targeted therapeutics are believed to overcome severe side-effects by exclusively blocking cancer cells and

disturbing functions specific for tumor growth. Although the mode of action of trastuzumab is not fully understood, several possible mechanisms have been described. For example, blocking of extracellular HER2 shedding and thus preventing the formation of constitutively active p95-HER2 (Molina et al. 2001), induction of cell cycle arrest by inhibition of downstream signaling (Yakes et al. 2002), and triggering of antibody-dependent cellular cytotoxicity have been reported (Clynes et al. 2000). Further targeted therapeutics with approval for the treatment of HER2-positive breast cancer are pertuzumab and lapatinib. Pertuzumab, approved in 2012 by the FDA, is like trastuzumab a humanized monoclonal antibody directed against HER2, but binds to a different region of HER2. Pertuzumab inhibits the ligand-dependent dimerization of HER2 with other HER2 family members (Adams et al. 2006) and is used in combination with trastuzumab and docetaxel as first-line therapy for patients with HER2-positive metastatic breast cancer (Baselga et al. 2012b). Lapatinib, a dual specific tyrosine kinase inhibitor directed against HER2 and the epidermal growth factor receptor (EGFR), is approved since 2007 by the FDA in combination with capecitabine to treat patients who have progressed under trastuzumab treatment (Xia et al. 2002, Geyer et al. 2006). Tyrosine kinase inhibitors like lapatinib represent, besides monoclonal antibodies, another important class of targeted therapeutics.

The last therapeutic subgroup is characterized by overexpression of ERα and PR accounting for 70 - 80% of all breast cancer cases. This subgroup has in general a more favorable prognosis compared to the other therapeutic subgroups and in addition is characterized by higher proportion of postmenopausal women (Anders et al. 2008, Heil et al. 2012). Further classification of this subgroup is crucial for therapy decision as some patients are at higher risk of recurrence and require chemo-endocrine treatment whereas other patients are at lower risk of recurrence and poorly respond to chemotherapeutic regimens. Therefore a robust classification of low and high risk hormone receptor-positive patients is essential for therapeutic decisions to avoid over- or under-treatment (Coates et al. 2012). To approximate this risk classification, histologic grading or abundance of a cell proliferation marker (e.g. Ki-67) is recommended, besides well established independent prognostic factors like age, tumor size, and lymph node status (Goldhirsch et al. 2011). However, assignment of histologic grade and measurement of Ki-67 expression using immunohistochemistry are subject to high interobserver variability and thus assign an intermediate risk to a substantial number of patients (Mengel et al. 2002, Longacre et al. 2006). In addition, despite the introduction of gene expression signatures and protein based approaches (Cardoso et al. 2008, Sparano and Paik 2008, Viale et al. 2008, Parker et

al. 2009, Kantelhardt et al. 2011), classification of low and high risk hormone receptor-positive breast cancer has remained a challenge. Besides the crucial decision for or against chemotherapy most patients with hormone receptor overexpression benefit from adjuvant endocrine therapy (Davies et al. 2011). Different types of endocrine therapy exist. These drugs all have the aim to inhibit estrogen receptor signaling, for example by blocking the binding of estrogen to the receptor or by depriving the tumor of estrogen. Tamoxifen, approved since 1977 and the first targeted therapy available for cancer treatment in general, is a selective estrogen receptor modulator and competes with endogenous estrogen for binding at the estrogen receptor (Cole et al. 1971, de Cupis et al. 1999). Tamoxifen is the standard endocrine therapy for pre-menopausal women (Kreienberg et al. 2012) whereas aromatase inhibitors like anastrozole are preferred for post-menopausal women (Burstein et al. 2010, Cuzick et al. 2010). Aromatase inhibitors block the peripheral production of estrogens by inhibiting the activity of the enzyme aromatase, which has the function to convert androgens to estrogens. Aromatase inhibitors lower the estrogen levels in the blood plasma and consequently at the tumor site (Dowsett et al. 1995).

Neoadjuvant systemic therapy can be indicated if patients present with locally advanced or primary inoperable breast cancer. It is used to shrink the tumor to enable surgery or to enhance the chance for breast conserving surgery (Kreienberg et al. 2012). Recent studies have shown, that overall survival does not differ between patients treated with systemic therapy before or after surgery (van der Hage et al. 2001). Therefore, neoadjuvant therapy is also increasingly used to treat early breast cancer with the aim to give immediate and valuable information of tumor response to the therapy of choice (Kaufmann et al. 2007).

Although treatment of breast cancer has made good progress over the last years, still a number of patients eventually relapses or shows progressive disease, due to acquired or intrinsic therapy resistance. Therefore, new therapy options with special focus on targeted therapeutics are continuously evaluated in clinical trials. Examples are drugs targeting mTOR (e.g. everolimus), HER2 (e.g. T-DM1), EGFR (e.g. cetuximab), ERBB family members (e.g. afatinib), and PARP (e.g. iniparib) (Baselga et al. 2012a, Verma et al. 2012, Carey et al. 2012, Schuler et al. 2012, O'Shaughnessy et al. 2011). However, in addition to the development of new therapeutics and combinatorial strategies, the discovery and validation of predictive biomarkers is needed to improve the definition of therapeutic subgroups. This will help to fully enable the concept of personalized medicine of finding the right treatment at the right time for the individual patient.

1.2.3 Deregulated signaling pathways in breast cancer

Development of breast cancer is a multistep process characterized by accumulation of genomic alterations. These genomic alterations have direct implications on deregulation of various signaling pathways crucial for normal cell function.

Deregulation of estrogen receptor signaling plays an important role in the majority of breast cancer cases. Over 100 years ago it was shown that surgical removal of the ovaries can lead to remission of metastatic breast cancer (Beatson 1896). This finding paved the way for the discovery of the interconnection of estrogen receptor signaling and breast cancer. Estrogens are a class of steroid hormones with estradiol (E2 or 17β-estradiol) as most common variant. In pre-menopausal women, estrogens are mainly produced in the ovaries, whereas in post-menopausal women other sites in the body become predominant, like mesenchymal cells of adipose tissue (Simpson and Davis 2001). The biologic effects of estradiol, which are mostly in the regulation of cell growth and differentiation, are mediated through binding to estrogen receptors in target cells. These estrogen receptors (ERα and ERβ) belong to the super family of nuclear receptors and act, once activated through ligand binding and subsequent dimerization, as transcription factors and transcription cofactors (Green et al. 1986, Enmark et al. 1997, Nilsson et al. 2001). The latter function also known as transcriptional cross-talk (Gottlicher et al. 1998). In addition, non-genomic functions of activated estrogen receptor, located near the plasma membrane, have been described. For example, the estradiol-receptor complex can activate the MAP-kinase cascade via SRC (Migliaccio et al. 1996) and insulin-like growth factor receptor 1 can be rapidly phosphorylated in the presence of estradiol (Kahlert et al. 2000).

Besides deregulation of hormone receptor signaling, aberrant receptor tyrosine kinase (RTK) signaling plays a fundamental role in breast cancer. These transmembrane receptors, due to oncogenic mutation or overexpression, can be involved in all steps of cancer development and progression, ranging from cell proliferation over angiogenesis to migration and invasion. Among the 58 receptor tyrosine kinases encoded in the human genome (Robinson et al. 2000), especially members of the MET, INSR, FGFR, VEGFR, and EGFR subfamilies have been extensively described in the context of breast cancer (Raghav et al. 2012, Iqbal et al. 2012, Turner et al. 2010, Dhakal et al. 2012, Nieto et al. 2007). The most prominent is the EGFR subfamily, with the members EGFR, ERBB2, ERBB3, and ERBB4 (Yarden and Sliwkowski 2001). For example, overexpression of ERBB2 (also known as HER2, NEU, NGL, TKR1, CD340, MLN 19, and HER-2/neu) has been noted in about 10 - 20% of all

breast cancer tumors (Heil et al. 2012, Slamon et al. 1987). In contrast to the other EGFR family members, no ligand is known for this receptor (Garrett et al. 2003). However, several ligands of the other family members have been described and deregulation of their expression has also been linked to the development and progression of breast cancer as well as to drug resistance (LeJeune et al. 1993, Revillion et al. 2008, Eckstein et al. 2008, McIntyre et al. 2009). Based on their receptor specificity, they are grouped into four categories. The first category comprises amphiregulin (AREG), epidermal growth factor (EGF), epithelial mitogen (EPGN), and transforming growth factor alpha (TGFα), which all bind exclusively to EGFR. The ligands of the second category are betacellulin (BTC), heparin-binding EGF-like growth factor (HBEGF), and epiregulin (EREG) having dual specificity for EGFR and ERBB4. The last two categories are composed of different neuregulins (or heregulins, HRG1-4) with various isoforms which bind either exclusively to ERBB3 or also to ERBB4 (Hynes and Watson 2010).

Besides deregulation of RAS/RAF and PI3K/PTEN signaling, which are the canonic downstream pathways of RTKs, at least ten other signaling pathways, as summarized in Figure 3, have been described to be prevalently involved in cancer development and progression.

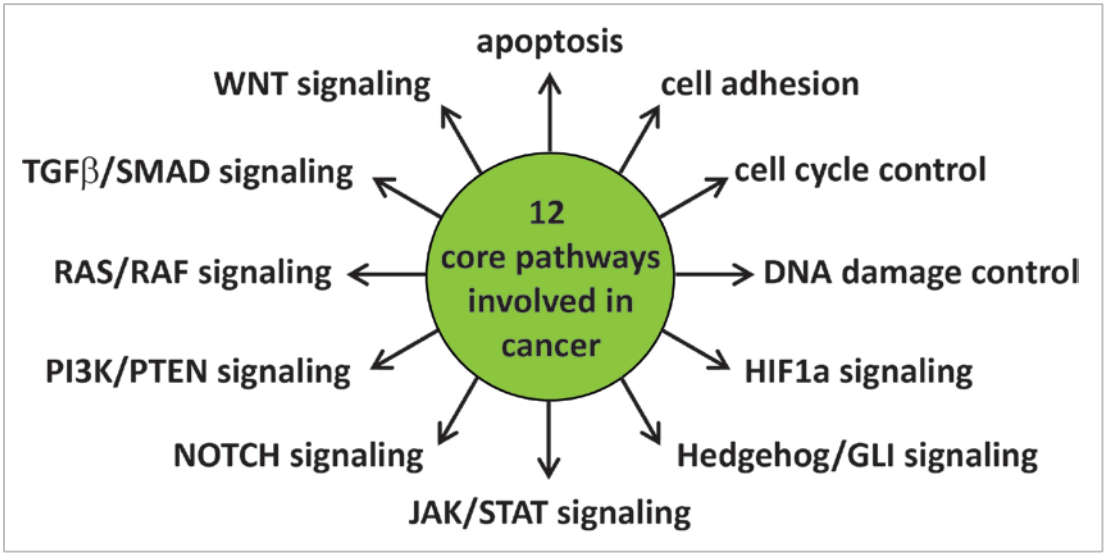


Figure 3: Overview of main signaling pathways with implication in cancer. (Figure inspired by lecture of Bert Vogelstein, *The sequence of all 185,000 coding exons in each of 100 human tumors: What has it taught us?*, AACR 101st Annual Meeting 2010).

For example, Hedgehog/GLI, NOTCH, and WNT signaling pathways, which are required for normal embryonic development, have recently also come into the focus of breast cancer research (Zardawi et al. 2009). In addition, deregulation of TGFβ/SMAD signaling as well as DNA damage control is frequently described in the context of breast cancer (Kretzschmar 2000, Jasin 2002).

Without doubt, the complexity of signaling networks increases with cross-talk between individual signaling pathways and exact pathway delineation is not possible due to redundant components shared by several signaling pathways. In addition, it is likely that not all factors playing a role in breast cancer relevant signaling pathways have been elucidated yet.

1.3 Biomarkers

In the area of personalized medicine, biomarkers are key in supporting the clinicians during their decision making process of finding the right diagnosis/treatment at the right time for the individual patient. A biomarker has been defined by the National Institutes of Health (Bethesda) Biomarkers Definitions Working group as follows:

> "Biological marker (biomarker): A characteristic
> that is objectively measured and evaluated as an
> indicator of normal biological processes,
> pathogenic processes, or pharmacologic responses
> to a therapeutic intervention."
> (Biomarkers Definitions Working Group 2001)

Different application areas of biomarkers have been described so far. These include aiding early diagnosis, determining prognosis, predicting response or resistance to different therapies as well as monitoring of disease progression, regression, and recurrence. In addition, pharmacodynamic biomarkers have been proven to be useful to evaluate drug safety and efficacy in preclinical development and early clinical trials (Frank and Hargreaves 2003). Diagnostic biomarkers have the aim to identify those individuals with disease or

abnormal conditions. Prognostic biomarkers can give valuable hints on the severity and natural course of a certain disease and subsequently facilitate the decision about individual adjuvant treatment. This is a fundamental issue, especially in cancer treatment, since numerous patients are overtreated whereas others are treated ineffectively. A predictive biomarker is intended to identify the subgroup of patients which is most likely to respond to the treatment of choice or to identify the subgroup of patients which is most likely to be resistant to a given treatment. Clinically useful predictive biomarkers in breast cancer, for example, are the ones indicating a positive estrogen receptor status, or a positive HER2 status, to select patients for endocrine therapy, or for therapy with trastuzumab, respectively (Goldhirsch et al. 2009). Moreover, biomarkers can be used to monitor treatment response or recurrence of disease over time. For instance, the prostate-specific antigen is well-accepted as biomarker that monitors the recurrence of prostate cancer (Lilja et al. 2008), whereas in ovarian cancer serum cancer antigen 125 can serve as biomarker to monitor disease recurrence (Rustin et al. 2001). The application area of a certain biomarker is not mutually exclusive. HER2, for example, is a biomarker combining prognostic as well as predictive features (Wolff et al. 2007).

Besides tumor tissue, the main sources of cancer biomarkers are body fluids like blood and urine. Blood and urine are preferred sources for biomarkers as they can be sampled in a less-invasive way compared to tumor tissue and should reflect the various physiological or pathological states of a patient. However caution has to be taken as confounding factors of secondary diseases are much more likely.

A variety of biomarker types exists, as a biomarker, by definition, could be any characteristic, which can be objectively measured. Popular biomarker types in cancer are circulating tumor cells (CTC), auto-antibodies, proteins, mRNAs, miRNAs, gene fusions, and gene mutations. The measurement of CTCs, proteins (e.g. uPA and PAI-1), and gene expression profiles (e.g. 21-gene recurrence score) has been introduced as prognostic biomarkers for breast cancer (Wallwiener et al. 2013, Kantelhardt et al. 2011, Paik et al. 2004). In lung cancer, specific autoantibodies were identified to facilitate diagnosis (Lam et al. 2011) and the EML/ALK gene fusion is used as predictive biomarker for the treatment of lung cancer patients with the tyrosine kinase inhibitor crizotinib (Kwak et al. 2010). A signature of miRNAs has been proposed to assess tumor progression in prostate cancer (Brase et al. 2011), whereas mutation status of KRAS and BRAF have predictive impact for treatment of colorectal cancer with the monoclonal antibody cetuximab and melanoma with the tyrosine kinase inhibitor vemurafenib, respectively (Karapetis et al. 2008,

Chapman et al. 2011). To expedite the development of personalized medicine further, new biomarkers with excellent sensitivity and specificity are required. Especially in clinical trials, companion diagnostics to select appropriate patients beforehand have a high potential to reduce the number of failed clinical trials.

1.4 Protein microarrays

Protein microarrays have emerged as well-established research tools for basic and applied proteomics. Due to the high-throughput and multiplex capacities, they are mainly applied in the context of systems biology approaches or for biomarker discovery and validation studies. Two main formats of protein microarrays exist, the forward phase format and the reverse phase format.

In case of forward phase protein microarrays, defined capture molecules are immobilized on planar surfaces such as chemically modified glass or plastic (Kusnezow and Hoheisel 2003, Wingren et al. 2007). Commonly used capture molecules are antibodies directed against the analytes of interest (Nielsen and Geierstanger 2004). For protein-protein interaction studies or autoantibody profiling, recombinant proteins are used as capture molecules (MacBeath and Schreiber 2000, Robinson et al. 2002). The arrayed capture molecules are then incubated with the analyte containing sample and the binding event is visualized and quantified by different strategies. For example by labelling of the sample with fluorescent dyes (Schroder et al. 2010) or biotinylation of the samples and subsequent use of a biotin-streptavidin detection system (Ingvarsson et al. 2007). Another option employing an indirect approach is the use of a detection antibody mix. This kind of detection can be done as a one-step approach by coupling of the detection antibodies to specialized tags which enable a fluorescent, chemiluminescent, colorimetric, electrochemical, or radiometric based analyte quantification. In case of a two-step approach, a tagged secondary antibody is used and therefore overcomes the need for direct labelling of each individual analyte specific detection antibody (Korf et al. 2008b). A special type of forward phase protein microarrays are bead based systems also called suspension arrays. In this case, the capture molecules are immobilized on specialized color coded beads instead of a planar surface (Fulton et al. 1997, Schwenk et al. 2008).

The reverse phase protein array format, first described in 2001, is characterized by direct immobilization of the samples (e.g. tumor or cell line lysates) instead of the capture molecules (Paweletz et al. 2001). This format allows the relative quantification of target protein expression across hundreds of samples in parallel while consuming only a minute amount of sample. Several variations of the reverse phase protein array format have been published so far, mainly with differences in detection system and subsequent data analysis strategies (Pawlak et al. 2002, Wulfkuhle et al. 2003, Loebke et al. 2007, Troncale et al. 2012). The two different protein microarrays formats (microspot immunoassays and reverse phase protein arrays) applied in this thesis are described in detail in the following sections.

1.4.1 Microspot immunoassays

Microspot immunoassays (MIA) belong to the forward phase category of protein microarrays and are in principle miniaturized multiplex sandwich immunoassays first described in 1989 (Ekins 1989). MIA enable the absolute quantification of multiple analytes in parallel and overcome technical limitation of traditional enzyme-linked immunosorbent assays by saving sample amount, time, and reagent consumption. The quantification of analytes in cell culture samples using the MIA approach was already established in the Division of Molecular Genome Analysis (DKFZ) (Korf et al. 2008a) and was further adjusted to the measurement of analytes in blood plasma and tumor tissue samples within the scope of this thesis. As illustrated in Figure 4, different capture antibodies are printed as technical replicates and in a predefined number of identical subarrays on nitrocellulose coated glass slides. After blocking of free binding sites on the nitrocellulose coating, the slides are mounted in a specialized incubation chamber to create distinct wells for each subarray. These wells are then used to either incubate with a serially diluted standard protein mix or with plasma/tumor lysate samples. The detection of the analytes is done in a two-step procedure by incubation with biotinylated detection antibody mix followed by incubation with near-infrared-dye labelled streptavidin. In between the different working steps, a wash procedure is employed to remove even minute amount of material that potentially

increases experimental noise. Next, slides are scanned on an near-infrared imaging system to determine the signal intensities of the different spots. The resulting data is then used to calculate a calibration curve for each of the different analytes present in the multiplexed standard mixture and to assess the different analyte concentrations in all samples. This data processing step is facilitated with the program QuantProReloaded, specifically tailored for this purpose.

A crucial step for the development of a new MIA is the identification of suitable antibody pairs. Besides common requirements of an immunoassay like accuracy, linearity, spike-in recovery, specificity, and sensitivity (Sweep et al. 2006) suitable antibody pairs must be compatible in the multiplexed set-up (Gonzalez et al. 2008).

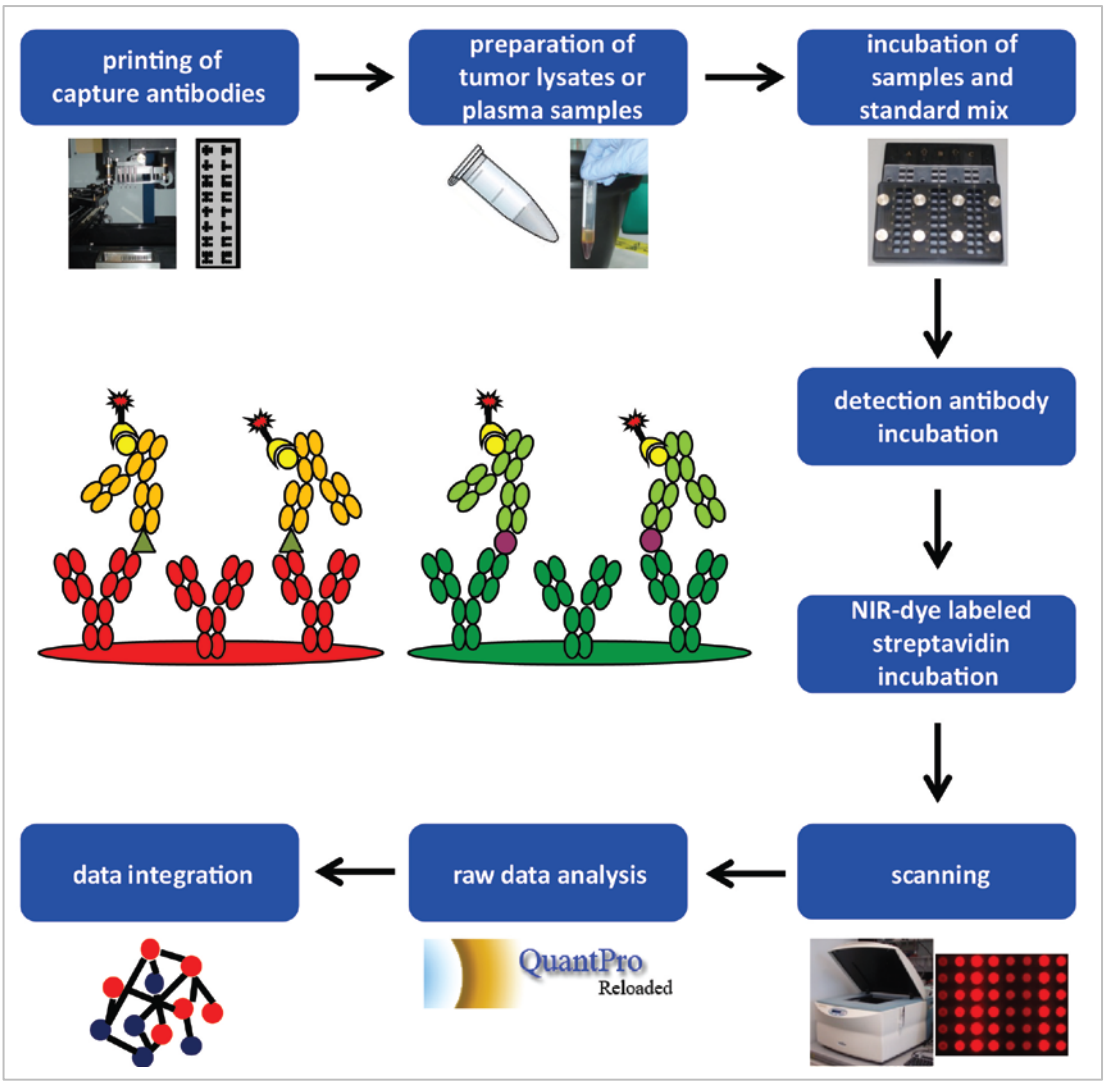


Figure 4: Microspot immunoassay workflow for quantification of analyte concentrations in tumor lysate or plasma samples.

1.4.2 Reverse phase protein arrays

Reverse phase protein arrays (RPPA) are in principle miniaturized dot blot immunoassays and enable the quantitative analysis of target protein abundance including posttranslational modifications in large sample sets. Over the past decade, the RPPA approach has been successfully used for different applications in the field of systems biology (Iadevaia et al. 2010, Uhlmann et al. 2012, Henjes et al. 2012) as well as for biomarker discovery projects in different tumor entities including prostate cancer, acute myeloid leukemia, and breast cancer (Grubb et al. 2009, Kornblau et al. 2010, Wulfkuhle et al. 2012).

The RPPA platform at the Division of Molecular Genome Analysis (DKFZ) was established by Christian Löbke in 2007 and since then further developed in terms of automation and data processing (Loebke et al. 2007, Mannsperger et al. 2010, Mannsperger 2011, Henjes et al. 2012). In contrast to colorimetric detection approaches used by the majority of other RPPA research groups (Wulfkuhle et al. 2003, Spurrier et al. 2008, Hennessy et al. 2010), the DKFZ RPPA platform is based on a near-infrared fluorescent detection system. Advantages of this approach are for example low and uniform background signals as well as a high dynamic range of signal detection.

A general overview of the RPPA workflow as presented in this thesis for tumor sample profiling is depicted in Figure 5. First, snap-frozen tumor samples are macro-dissected and total protein is isolated. The resulting tumor lysates are printed with equal total protein concentration on a series of nitrocellulose coated glass slides alongside with dilution series of appropriate controls. In principle, up to thousands of spots, containing each the whole proteome of different cell line or tumor samples can be arrayed on one slide. Best results are obtained using an accurate contact printer equipped with solid pins. The deposited sample amount should be in the range of 2 - 6 ng total protein per spot and result in a spot diameter of around 250 µm. Depending on the number of different samples several identical subarrays can be printed per slide to optimize the throughput. Each of the arrays is subsequently incubated with a target specific primary antibody. These primary antibodies have to be thoroughly validated beforehand to be highly target specific, as in contrast to Western blot, proteins are not resolved by molecular weight. The incubation of the arrays with primary antibodies is then followed by the final detection step using a near-infrared-dye labeled secondary antibody. Appropriate blocking and washing steps are applied throughout the whole procedure. After air-drying, the slides are scanned with an infrared

imaging system and individual spot intensities are quantified with an appropriate image analysis software. The resulting raw data is processed using the R-Package RPPanalyzer (Mannsperger et al. 2010) followed by quality control checks and a detailed analysis integrating clinical or further experimental data.

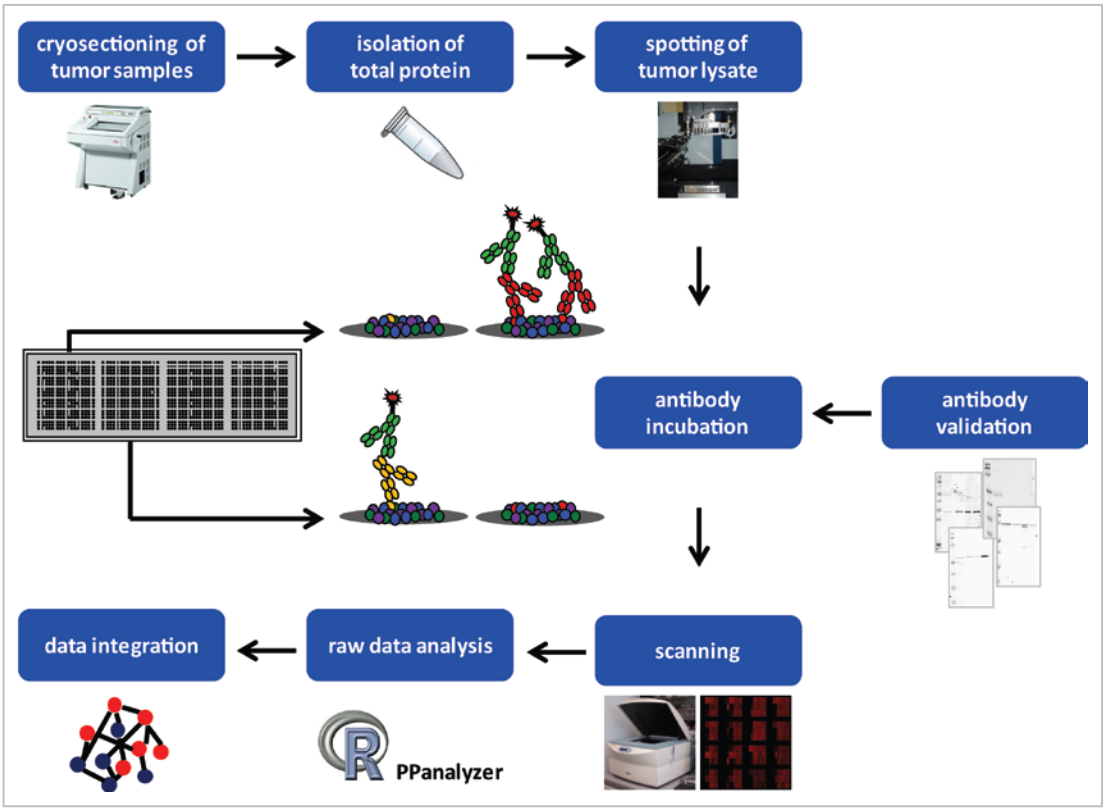


Figure 5: Reverse phase protein array workflow for tumor profiling.

1.5 Project aim

Breast cancer is nowadays widely recognized as a heterogeneous disease. With this knowledge in mind and following the hypothesis that intrinsic biologic features of breast tumors affect prognosis and therapy response, the general aim of this thesis was to further explore breast cancer heterogeneity with protein microarrays.

The first objective was the identification of a robust and quantitative protein biomarker signature to facilitate the risk classification of hormone receptor-positive breast cancer. To achieve this aim, RPPA based tumor profiling was used in combination with a novel biomarker hit selection workflow to screen across over 120 candidate biomarker proteins. As technical prerequisite it was first tested if well-known characteristics of breast cancer subtypes can be reproduced by this method.

The second objective was to generate pathway activation profiles of hormone receptor-positive breast cancer tumor samples covering components of major signaling pathways known to be involved in cancer. With this approach the question was addressed if tumors can be grouped according to similar signaling pathway activation patterns and whether these patterns reflect common molecular mechanisms. Moreover, the signaling activation patterns obtained using RPPA were correlated with established clinicopathological features such as age, tumor size, lymph node status, histologic type, and histologic grade. To complement this analysis, a MIA was developed to enable the measurement of eight different growth factors in tumor lysates as well as blood plasma of matching patient samples. The growth factor expression data was then compared to clinicopathological features and analyzed in context of the matched tumor pathway activation profiles. New insights obtained on the heterogeneity of hormone receptor-positive breast cancer should have the potential to serve as basis to define new pathway-specific and patient-tailored therapy options.

2. Materials and Methods

2.1 Materials

2.1.1 Instruments

aspiration device	Vacuboy (INTEGRA Biosciences, Fernwald, Germany)
balances	BP121S and BP2100S (Sartorius, Göttingen, Germany)
bead mill	TissueLyser (Qiagen, Hilden, Germany)
centrifuges	Biofuge fresco (Heraeus, Hanau, Germany); Varifuge 3.0R (Heraeus, Hanau, Germany); Rotina 35R (Andreas Hettich, Tuttlingen, Germany); Galaxy mini (Merck Eurolab,Darmstadt, Germany)
contact printer	2470 Arrayer (Aushon BioSystems, Billerica, USA)
cryomicrotome	cryostat CM-1950 (Leica Mikrosysteme, Wetzlar, Germany)
dessicator	DURAN® dessicator (Schott, Mainz, Germany)
electroblotting system	Trans-Blot®SD (BioRad, München, Germany)
electrophoresis system	Mini-Protean®II electrophoresis cell system (BioRad, München, Germany)
heating block	Dri-block®DB-2D (Bibby Scientific Limited, Stone, UK)
incbation chambers	incubation chamber 3/4 and 3/16 (Metecon, Mannheim, Germany)

magnetic stirrer	MR3001 (Heidolph, Schwabach, Germany)
manual hand-held dispenser	Multipette®plus (Eppendorf, Hamburg, Germany)
microplate reader	Infinite M200 (Tecan, Männedorf, Switzerland)
multichannel pipette	Biohit Proline (Sartorius, Göttingen, Germany)
Odyssey® Infrared Imaging System	Odyssey® Infrared Imager (LI-COR Biosciences, Lincoln, USA)
pH meter	inoLab (WTW, Weilheim, Germany)
pipettes	Pipetman® (Gilson, Limburg, Germany); ErgoOne® (Starlab International, Hamburg, Germany); Pipetboy acu (INTEGRA Biosciences, Fernwald, Germany)
rocking platforms	RM5-30 and ST5 (CAT Ingenieurbüro M. Zipperer, Staufen, Germany); Titramax 100 (Heidolph, Schwabach, Germany); Tube Rotator (VWR, Darmstadt, Germanx)
spectrophotometer	NanoDrop ND-1000 (NanoDrop Products, Wilmington, USA)
Thermomixer	Thermomixer comfort (Eppendorf, Hamburg, Germany)
vortexter	VortexMixer 7-2020 (neoLab, Heidelberg, Germany)
wash box	Western incubation box (LI-COR Biosciences, Lincoln, USA)
water purification system	Milli-Q Biocel System (Merck Millipore, Darmstadt, Germany)

2.1.2 Chemicals and consumables

348-well plates	AB-1056 (Abgene, Epsom, UK)
4x sample loading buffer	Roti®-Load 1 (Carl Roth, Karlsruhe, Germany)
96-well plates	V-shape and flat bottom (Greiner Bio-One, Frickenhausen, Germany)
acrylamide (30%)/bisacrylamide (0.8%)	Rotiphorese® Gel 30 (Carl Roth, Karlsruhe, Germany)
blocking buffer	Fluorescent Western Blotting (Rockland, Gilbertsville, USA)
blunt needle	(B. Braun, Melsungen, Germany)
combitips	(Eppendorf, Hamburg, Germany)
Complete Mini Protease Inhibitor Cocktail	(Roche Diagnostics, Mannheim, Germany)
dessicant bags	(Conrad Electronics, Hirschau, Germany)
Fast Green FCF	(Carl Roth, Karlsruhe, Germany)
homogenizer tubes	QIAshredder (Qiagen, Hilden, Germany)
mammalian protein extraction reagent	M-PER (Thermo Fischer Scientific, Rockford, USA)
nitrocellulose-coated glass slides	Oncyte® Avid (Grace Bio-Labs, Bend, USA)
PhosSTOP Phosphatase Inhibitor Cocktail	(Roche Diagnostics, Mannheim, Germany)
pipette tips	(STARLAB, Hamburg, Germany); (Becton Dickinson, Heidelberg, Germany)

polyvinylidene difluoride membrane	Immobilon-FL (Merck Millipore, Darmstadt, Germany)
prestained protein ladder	peqGOLD Protein Marker IV and V (PEQLAB Biotechnologie, Erlangen, Germany)
scalpel	Feather No21 (pfm medical, Köln, Germany)
silica gel beads (orange)	(AppliChem, Darmstadt, Germany)
stainless steel beads (5 mm)	(Qiagen, Hilden, Germany)
staurosporine	(Merck Millipore, Darmstadt, Germany)
tuberculin syringe	(Becton Dickinson, Heidelberg, Germany)
tissue protein extraction reagent	T-PER (Thermo Fischer Scientific, Rockford, USA)
tubes	(Eppendorf, Hamburg, Germany)
Whatman paper	(GE Healthcare, München, Germany)

All other chemicals not listed above were either purchased from Sigma-Aldrich (St. Louis, USA) or Carl Roth (Karlsruhe, Germany),

2.1.3 Antibodies and recombinant proteins

Table 1: Primary antibodies for Western blot and RPPA

target protein	phosphorylation site	order ID	company[1]	project[2]
acetyl-CoA carboxylase		3662	CST	A
AKT1		610860	BD	A, B
AKT1/2	S473	9271	CST	A, B
AKT1/2	T308	9275	CST	A, B
AKT2		3063	CST	A, B
APC		2504	CST	A, B
ATM		2873	CST	A, B
ATR		2790	CST	A, B
BAX		2772	CST	A, B
βCatenin		9562	CST	A, B
βCatenin	S33/S37/T41	9561	CST	A, B
BCL-2		2876	CST	A, B
BCL-XL		2764	CST	A, B
BRCA2		9012	CST	A, B
caveolin-1		ab32577	Abcam	A
CBL		2747	CST	A
CDK1		9112	CST	A, B
CDK2		C5223	Sigma	A, B
CDK6		sc-177	SC	A, B
cJUN		9165	CST	A
Claudin 1		DP153-05	Acris	A, B
Claudin 3		DP155-05	Acris	A, B
COL4A3BP		HPA035645	KTH	A
cRAF	S259	9421	CST	A, B
CREB		9197	CST	A
cyclin B1		1495-1	Epitomcs	A, B
cyclin D1		sc-718	SC	A, B
cytokeratin 18		1924-1	Epitomcs	A
cytokeratin 8	S23	2147-1	Epitomcs	A

target protein	phosphorylation site	order ID	company[1]	project[2]
E-cadherin		4065	CST	A, B
EGFR		2646	CST	A
EpCAM		2929	CST	A, B
HER2		AB-17	Thermo	A
HER2	Y1112	04-294	Merck	A
HER2	Y1248	ab47755	Abcam	A
ERBB3		AB-2	Thermo	A
ERBB4		sc-283	CST	A
ERBB4	Y1162	2295-1	Epitomcs	A
ERK1		AF1575	R&D	A, B
ERK1/2	T202/Y204/T185/Y187	4370	CST	A, B
ERα		SP1	DCS	A
FAK	S843	F7928	Sigma	A, B
FIH		4426	CST	A, B
FoxO3a	S318/S321	9465	CST	A
GATA3		5852	CST	A
GRB2		3972	CST	A
GSK3A		9338	CST	A, B
GSK3A	S21	9316	CST	A, B
GSK3A/B	Y279/Y216	2309-1	Epitomcs	A, B
GSK3B		9315	CST	A, B
GSK3B	S9	9323	CST	A, B
Integrin B1		4706	CST	A, B
Integrin B3		4702	CST	A, B
Ki-67		M7240	Dako	A
LAMB1		3575-1	Epitomcs	A
MCL-1		5453	CST	A, B
MEK		610122	BD	A, B
MEK	S217/S221	M7683	Sigma	A, B
MET		3148	CST	A
metadherin		9596	CST	A

target protein	phosphorylation site	order ID	company[1]	project[2]
MNK1		2195	CST	A, B
mTOR		2983	CST	A, B
mTOR	S2448	2971	CST	A, B
N-cadherin		4061	CST	A, B
NDKA		5353	CST	A
NFκB		sc-372	SC	A, B
NFκB	S536	3033	CST	A, B
NOTCH2		4530	CST	A, B
NOTCH3		5276	CST	A, B
p27		610241	BD	A, B
p38		9212	CST	A, B
p38	T180/Y182	9211	CST	A, B
p53		sc-126	SC	A, B
p53	S15	9286	CST	A, B
p70S6K		2708	CST	A, B
p70S6K	T389	9234	CST	A, B
p70S6K	T421/S424	1135-1	Epitomcs	A, B
p90RSK	S380	9341	CST	A, B
PAK1		2602	CST	A, B
PAK2		2247-1	Epitomcs	A, B
PARP		AM30	Merck	A, B
P-cadherin		2130	CST	A, B
PCNA		sc-7907	SC	A
PDI		3501	CST	A
PDK1		3062	CST	A, B
PDK1	S241	3061	CST	A, B
PI3K_p110		4249	CST	A, B
PI3K_p85		ab40755	Abcam	A, B
PKA		sc-903	SC	A
PKCα		ab32376	Abcam	A, B
PKCα	S657/Y658	ab23513	Abcam	A, B

target protein	phosphorylation site	order ID	company[1]	project[2]
PLCγ		ab41433	Abcam	A
PR		1483-1	Epitomcs	A
PRAS40		2691	CST	A, B
PRAS40	T246	2997	CST	A, B
PTEN		9552	CST	A, B
PTEN	T366/S370	2195-1	Epitomcs	A, B
pyruvate dehydrogenase		3205	CST	A
RB		9309	CST	A, B
RB	S807/S811	9308	CST	A, B
RKIP		07-137	Merck	A, B
ROCK1		4035	CST	A, B
ROCK2		HPA007459	Sigma	A, B
RPS6		2217	CST	A, B
RPS6	S235/S236	4858	CST	A, B
RPS6	S240/S244	2215	CST	A, B
RSK		9355	CST	A, B
SDHA		sc-59687	SC	A
SHP1		HPA001466	Sigma	A
SHP2		1609-1	Abcam	A
SMAD2		3103	CST	A, B
SMAD7		ab76498	Abcam	A, B
SMURF2		ab53316	Abcam	A, B
SRC		2123	CST	A, B
SRC	Y416	2101	CST	A, B
STARD10		HPA026661	KTH	A
STAT1		9175	CST	A, B
STAT1	Y701	612232	BD	A, B
STAT3		610189	BD	A, B
STAT3	Y705	9131	CST	A, B
STAT5	Y694/Y699	05-495	Merck	A, B
TIE2		4224	CST	A

target protein	phosphorylation site	order ID	company[1]	project[2]
TOP2A		HPA006458	KTH	A
TSC1		4906	CST	A, B
TSC2		3990	CST	A, B
TSC2	T1462	3617	CST	A, B
VEGFR2		2479	CST	A
vimentin		3932	CST	A

[1]Abcam (Abcam, Cambridge, UK); Acirs (Acris Antibodies, Herford, Germany); BD (Becton Dickinson, Heidelberg, Germany); CST (Cell Signaling Technology, Danvers, USA); Dako (Dako Deutschland, Hamburg, Germany); DCS (DCS Innovative Diagnostik-Systeme, Hamburg, Germany); Epitomics (Epitomics, Burlingame, USA); KTH (courtesy Prof. Uhlen, KTH Sweden); Merck (Merck Millipore, Darmstadt, Germany); R&D (R&D Systems, Wiesbaden, Germany); SC (Santa Cruz Biotechnology, Heidelberg, Germany); Sigma (Sigma-Aldrich, St. Louis, USA); Thermo (Thermo Fischer Scientific, Rockford, USA)

[2]Used for biomarker discovery project (A); Used for pathway activation profiling project (B)

Table 2: Capture antibodies for MIA

target protein	order number	company[1]
AREG	MAB262	R&D
BTC	MAB261	R&D
EGF	MAB636	R&D
HBEGF	AF-259-NA	R&D
HGF	MAB694	R&D
HRG	N195	Leinco
TGFα	AF-239-NA	R&D
VEGF	MAB293	R&D

[1]R&D (R&D Systems, Wiesbaden, Germany); Leinco (Leinco Technologies, St. Louis, USA)

Table 3: Biotinylated detection antibodies for MIA

target protein	order number	company[1]
AREG	BAF262	R&D
BTC	BAF261	R&D
EGF	BAF236	R&D
HBEGF	BAF259	R&D
HGF	BAF294	R&D
HRG	BAF377	R&D
TGFα	BAF239	R&D
VEGF	BAF293	R&D

[1]R&D (R&D Systems, Wiesbaden, Germany)

Table 4: Recombinant human proteins for MIA

target protein	order number	company[1]
AREG	262-AR-100	R&D
BTC	262-CE-010	R&D
EGF	236-EG	R&D
HBEGF	259-HE-050	R&D
HGF	294-HG-005	R&D
HRG	377-HB-050	R&D
TGFα	239-A-100	R&D
VEGF	293-VE-010	R&D

[1]R&D (R&D Systems, Wiesbaden, Germany)

Table 5: Secondary detection reagents (Life Technologies, Darmstadt, Germany)

format	reactivity	host	conjugate
full-length IgG	rabbit IgG (H+L)	goat	Alexa Flour® 680
full-length IgG	mouse IgG (H+L)	goat	Alexa Flour® 680
F(ab')2	rabbit IgG (H+L)	goat	Alexa Flour® 680
F(ab')2	mouse IgG (H+L)	goat	Alexa Flour® 680
streptavidin	biotin	---	Alexa Flour® 680

2.1.4 Buffers and solutions

10x PBS:

1.37 M NaCl, 27 mM KCl, 18 mM KH_2PO_4, 100 mM Na_2PO_4, pH 7.4

10x TBS:

1.37 M NaCl, 200 mM Tris, pH 7.6

PBST:

0.1% Tween20® in PBS

TBST:

0.1% Tween20® in TBS

cell line lysis buffer:

mammalian protein extraction reagent (M-PER) with PhosSTOP Phosphatase Inhibitor Cocktail and Complete Mini Protease Inhibitor Cocktail

tissue lysis buffer:

tissue protein extraction reagent (T-PER) with 1 mM EDTA, 5 mM NaF, 2 µM staurosporine, PhosSTOP Phosphatase Inhibitor Cocktail, and Complete Mini Protease Inhibitor Cocktail

4x separation buffer:

1.5 M Tris, pH 8.8

4x stacking buffer:

0.5 M Tris, pH 6.8

SDS-PAGE running buffer:

192 mM glycine, 25 mM Tris, 0.1% SDS

anode buffer I:

300 mM Tris, 20% methanol

anode buffer II:

25 mM Tris, 20% methanol

cathode buffer:

40 mM 6-aminocaproic acid, 20% methanol

modified blocking buffer:

50% blocking buffer, 5 mM NaF, and 1 mM Na_3VO_4 in TBS

4x printing buffer:

10% Glycerol, 4% SDS, 10 mM DTT, 125 mM Tris, pH 6.8

FCF staining solution:

0.005% Fast Green FCF, 10% acetic acid, 30% ethanol

FCF destaining solution:

10% acetic acid, 30% ethanol

recombinant protein standard mix:

3 ng/ml AREG (262-AR-100), 3 ng/ml BTC (262-CE-010), 0.5 ng/ml EGF (236-EG), 0.5 ng/ml HBEGF (259-HE-050), 3 ng/ml HGF (294-HG-005), 3 ng/ml HRG (377-HB-050), 0.5 ng/ml TGFα (239-A-100), and 3 ng/ml VEGF (293-VE-010) in PBS with 20% fetal bovine serum (Life Technologies, Darmstadt, Germany)

detection antibody mix:

0.42 µg/ml anti-AREG (BAF262), 0.42 µg/ml anti-BTC (BAF261), 0.28 µg/ml anti-EGF (BAF236), 0.36 µg/ml anti-HBEGF (BAF259), 0.42 µg/ml anti-HGF (BAF294), 0.42 µg/ml anti-HRG (BAF377), 0.28 µg/ml anti-TGFα (BAF239), 0.36 µg/ml anti-VEGF (BAF293) in TBST with 0.1% BSA (PAA Laboratories, Pasching, Austria)

2.1.5 Kits

bicinchoninic acid assay	Thermo Fischer Scientific, Rockford, USA
miRNeasy Mini Kit	Qiagen, Hilden, Germany

2.1.6 Software

Odyssey Application Software 3.0	LI-COR Biosciences, Lincoln, USA
GenePix Pro 5.0	Molecular Devices, Sunnyvale, USA
R version 2.13.1	R Development Core Team (2011)
SigmaPlot 12.0	Systat Software, Erkrath, Germany
KNIME 2.5.0	www.knime.org

2.2 Methods

2.2.1 Biobanking of clinical samples

Tumor specimens from patients diagnosed with invasive breast carcinoma were obtained from Heidelberg University Women's Hospital/National Center of Tumor Diseases Heidelberg at the time of surgery. The samples were collected between 2008 and 2010 with Institutional Review Board approval and informed consent was obtained from all patients. The patients had not received neoadjuvant therapy. Tumor samples were reviewed by histopathology to contain > 70% tumor cells and were stored snap-frozen at -80°C. To obtain aliquots of each sample, they were further processed in two different ways depending on sample size. Tumors were either cut with a scalpel or with a cryomicrotome. In case of cryosectioning, the tumor was cut into 60 µm slices and the slices were distributed equally between different aliquots. These procedures allow optimal usage of the finite tumor sample amount, as aliquots can be used for different downstream processing protocols e.g. for DNA, RNA, or protein isolation.

Blood samples were collected from patients with suspicion of breast cancer at the day of diagnostic biopsy at Heidelberg University Women's Hospital/National Center of Tumor Diseases Heidelberg. Institutional Review Board approval and informed consent was obtained from all patients. Blood samples (9 ml) were collected in EDTA coated tubes and kept at 4°C until processing at the same day. Blood samples, collected from December 2008 till May 2010, were centrifuged at 4000 x g for 30 min at 4°C. The blood plasma was subsequently aliquoted (10 x 100 µl) and stored at -80°C until further use. Blood samples collected from May 2010 till August 2011, were centrifuged at 1300 x g for 20 min at 10°C. The blood plasma was additionally centrifuged at 15,500 x g for 10 min at 10°C and the supernatant was stored aliquoted (10 x 100 µl) at -80°C until further use.

2.2.2 Preparation of protein extracts from cell lines

The breast cancer cell lines MDA-MB-231 (HTB-26), HCC1954 (CRL-2338), and MCF7 (HTB-22) were purchased from the American Type Culture Collection and cell line authentication was performed at the DKFZ Genomics and Proteomics Core Facility (Castro et al. 2013). Pre-chilled cell line lysis buffer (500 µl) was added to frozen cell pellets (1×10^7 cells) and incubated on a tube rotator for 20 min at 4°C. The cell line lysates were subsequently centrifuged for 10 min at 16,000 x g and the supernatant was stored at -80°C until further use. The total protein concentration of cell line lysates was determined by bicinchoninic acid assay with a modified protocol adapted to sample volumes of 5 µl (Korf et al. 2008a).

2.2.3 Preparation of protein extracts from tumor samples

Pre-chilled tissue lysis buffer (10 µl / 1 mg tumor) was added to frozen tumor samples and samples were thawed on ice for 5 min. One stainless steel bead was added per tube and the samples were homogenized for 4 min at 30 Hz with a bead mill. Afterwards, the tumor lysates were placed on ice for 5 min and subsequently frozen on dry ice. These tumor lysates were thawed on ice and centrifuged at 16,000 x g for 10 min at 4°C. The supernatants were transferred to homogenizer tubes (QIAshredder) and centrifuged at 16,000 x g for 1 min at 4°C. The homogenized tumor lysates were aliquoted and stored at -80°C until further use. The total protein concentration of tumor lysates was determined by bicinchoninic acid assay with a modified protocol adapted to sample volumes of 5 µl (Korf et al. 2008a).

2.2.4 SDS-PAGE and Western blot

Polyacrylamide gels for SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) were prepared according to the protocol listed in Table 6. Separation gels with 7.5% or 12.5% acrylamide were used for separation of target proteins with 60 – 350 kDa and 15 – 100 kDa, respectively. Tumor or cell line lysates (20 µg total protein) were mixed with 4x sample loading buffer and heated for 5 min at 95°C prior to loading on the gel. A prestained protein ladder was used as molecular mass marker. Run time of gels was 1.5 – 2 h at 120 – 150 V, depending on target protein size.

Table 6: Protocol for polyacrylamide gels composed of separation and stacking gel

	separation gel 7.5%	**separation gel 12.5%**	**stacking gel**
4x separation buffer	5 ml	5 ml	
4x stacking buffer			2.5 ml
H_2O	9.7 ml	6.4 ml	6 ml
acrylamide (30%)/ bisacrylamide (0.8%)	5 ml	8.3 ml	1.3 ml
SDS (10%)	200 µl	200 µl	100 µl
APS (10%)	100 µl	100 µl	100 µl
TEMED	6.7 µl	6.7 µl	5 µl

After separation of the proteins via SDS-PAGE, the proteins were transferred from the polyacrylamide gel to a polyvinylidene difluoride (PVDF) membrane by electrophoresis. The "semi dry" blotting set-up was as follows: Four Whatman paper soaked in anode buffer I, two Whatman paper soaked in anode buffer II, PVDF membrane activated in methanol and equilibrated in anode buffer II, polyacrylamide gel, six Whatman paper soaked in cathode buffer. The blotting was done at 25 V for 1 h. Afterwards the membrane was blocked for 1 h with modified blocking buffer and subsequently incubated with target specific primary antibody over night at 4°C on a rocking platform. The membrane was washed 4 x 5 min in TBST followed by incubation with Alexa Flour® 680 conjugated secondary antibody (1:10,000 in TBST) for 1 h at room temperature. The membrane was protected from light. After washing for 4 x 5 min, the membrane was scanned at an excitation wavelength of 685 nm and a resolution of 84 µm using the Odyssey® Infrared Imaging System.

2.2.5 Selection and validation of antibodies for reverse phase protein arrays

Target proteins for pathway activation profiling and biomarker discovery were identified based on literature research. First, antibodies directed against proteins and posttranslational modifications were selected to cover twelve different signaling pathways known to be involved in cancer (see Figure 3 chapter 1.2.3). Next, additional antibodies directed against proteins and posttranslational modifications with known or hypothesized link to breast cancer were selected to complete the set of target proteins used for biomarker discovery.

Prior to use in RPPA each antibody had to be tested for specificity to assure the detected signal is representative for the target of interest. The gold standard for antibody validation is Western blot. A pool of different tumor samples and three different breast cancer cell lines (MDA-MB-231, MCF7, HCC1954) were used as test samples. All antibodies resulting in a target specific single band or characteristic band pattern for the tumor pool and/or the cell line (Figure 6 A – D) were used for RPPA. In addition, antibodies showing no signal on Western blot (Figure 6 E) were also used for RPPA and revalidated on Western blot if a signal could be detected on RPPA using the sample that generated the signal. Antibodies resulting in several unspecific bands (Figure 6 F) were not used for RPPA.

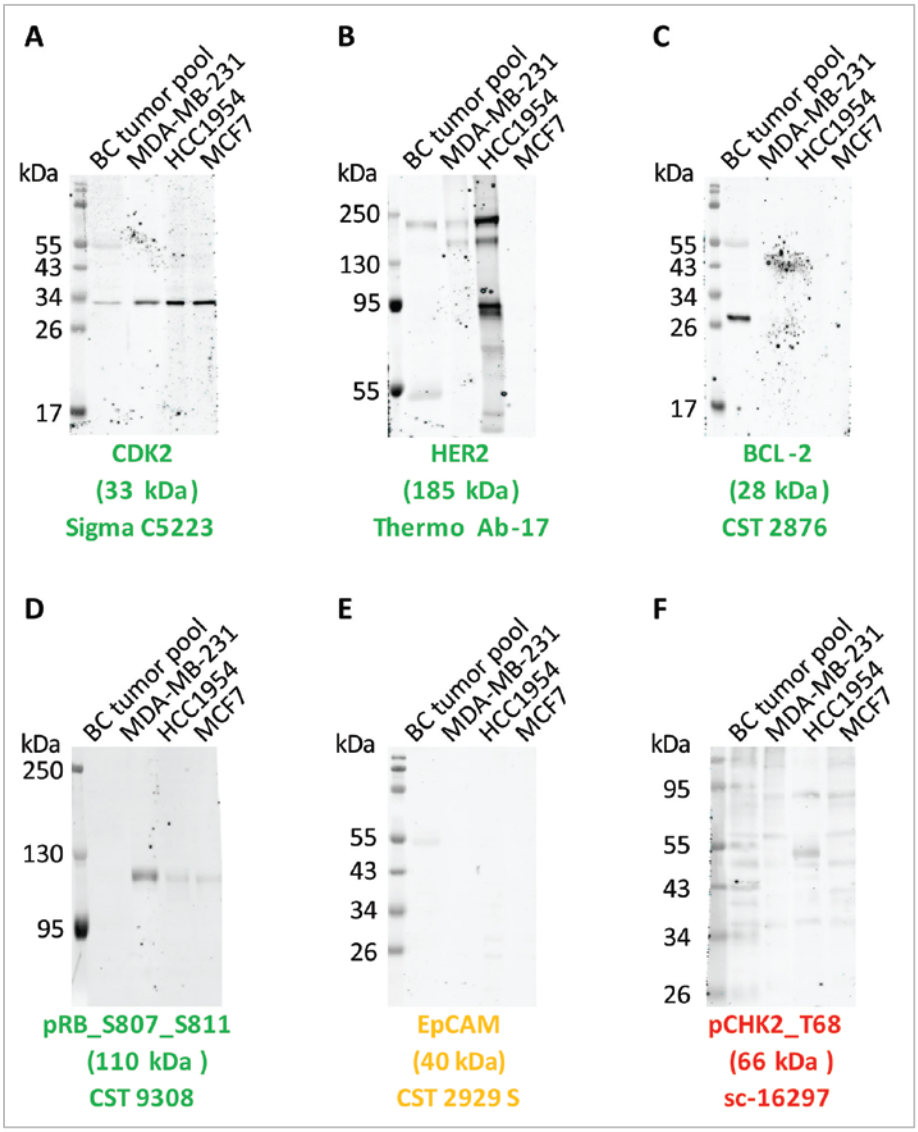


Figure 6: Representative Western blots of the antibody validation process. A pool of breast cancer (BC) tumor lysates as well as three different cell lines (MDA-MB-231, HCC1954, MCF7) was used as test sample set. Antibodies were grouped into six different categories according to the result. (A) Single band at predicted protein size. (B) Target specific band pattern. (C) Signal at predicted protein size only for tumor pool. (D) Signal at predicted protein size only for cell lines. (E) No signal for tumor pool as well as for cell lines. (F) Antibodies showing several unspecific bands.

2.2.6 Reverse phase protein arrays (printing and antibody incubation)

Prior to printing of slides, all samples were mixed with 4x printing buffer containing SDS/DTT and boiled for 5 min at 95°C. Subsequently, 24 µl of each sample was transferred to 348-well plates, which were filled with water at the edges to avoid evaporation effects, and centrifuged 2 min at 200 x g. Samples comprised tumor lysates with uniform total protein concentration of 2 µg/µl (for ≥ 5% of the samples) as well as several control samples as two-fold six step dilution series. In detail, three different tumor pools consisting of either ERα/PR-positive, HER2-positive, or triple-negative tumor samples as well as three different breast cancer cell lines (MDA-MB-231, MCF7, HCC1954). The concentration of dilution step 1 for the tumor pool samples was 3 µg/µl and for the cell line samples 2 µg/µl. Tissue lysis buffer served as negative control. All samples were printed as technical triplicates and four identical subarrays on nitrocellulose-coated glass slide using a contact printer equipped with 185 µm solid pins (1.6 nl sample per spot, average spot diameter 250 µm). The print head configuration was 4 x 3 pins with 4.5 mm pin spacing. The humidity during the printing run was kept constant at 80%. In total, 72 slides were printed at two different days. The RPPA printing layout, as depicted in Figure 7, was identical for both runs. Slides were stored after the print run with desiccant at -20°C.

Slides were mounted in 4-pad incubation chambers and blocked with modified blocking buffer for 2 h at room temperature. Each subarray was subsequently incubated with a target specific primary antibody at 4°C over night. Representative subarrays were incubated without primary antibody serving as “blank” control. The next day, slides were washed 4 x 5 min with TBST. Detection of the primary antibodies was done with Alexa Fluor® 680 F(ab')2 fragments of goat anti-mouse IgG or anti-rabbit IgG in 1:8000 dilution for 1 h at room temperature. Slides were washed 4 x 5 min with TBST followed by a final washing step with ultra pure water for 5 min before air drying of the slides. All incubation and washing steps were performed on a rocking platform and slides were protected from light. Three slides of each run were stained with Fast Green FCF for total protein quantification to serve as normalization reference as described before (Loebke et al. 2007). Images of all slides were obtained at an excitation wavelength of 685 nm and a resolution of 21 µm using the Odyssey® Infrared Imaging System and saved as 16-bit TIFF files.

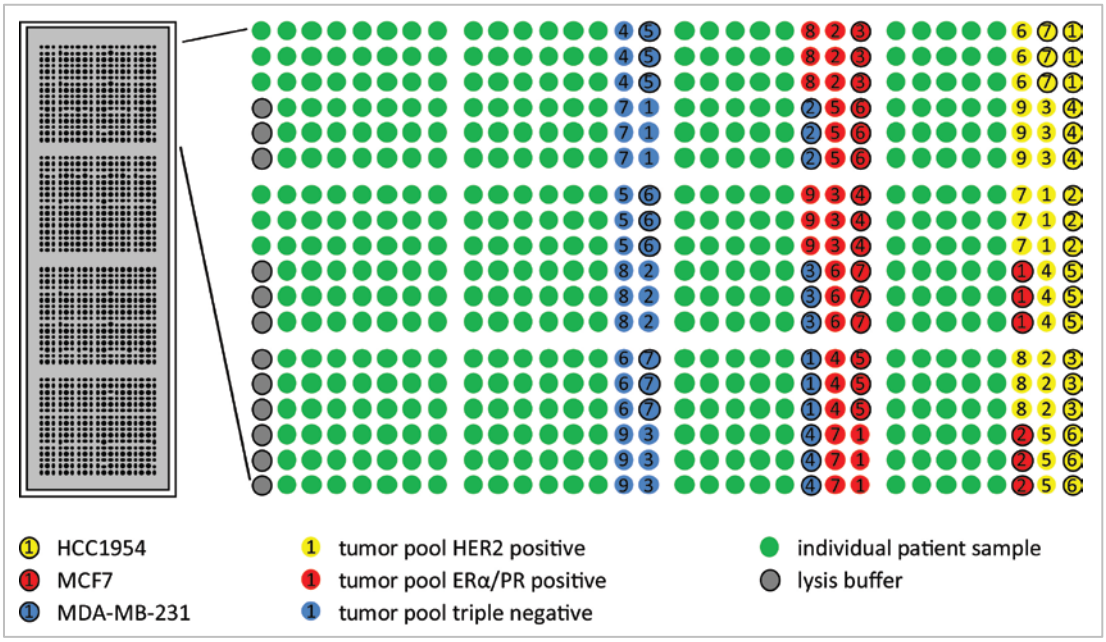


Figure 7: RPPA printing layout. Four identical sub-arrays were printed on each slide. Cell line samples and tumor pool samples serving as positive controls were printed as dilution series, whereas individual patient samples were printed with a uniform total protein concentration of 2 µg/µl. Tissue lysis buffer was used as negative control. All samples were printed as technical triplicate.

2.2.7 Reverse phase protein arrays (image analysis and raw data processing)

The software GenePixPro 5.0 was used to quantify the signal intensity of each individual spot. Therefore, the RPPA slide images generated with the Odyssey® Infrared Imaging System, as well as the *gene pix array list* file (.gal), generated by the printer to map the sample location on the slide were required. The resulting data was saved as *gene pix result* file (.gpr). At this step a visual inspection of the slides was performed. Slides without uniform background signal were excluded from further analysis.

RPPA raw data processing and quality control were done with the R-package RPPanalyzer (Mannsperger et al. 2010). Therefore, the .gpr files as well as two .txt files with sample annotation and slide-specific information were required. First, the raw signal intensities of the control samples were plotted against the respective total protein concentration. Only data of antibodies resulting in a linear correlation between target signal intensity and protein concentration were used for further analysis. The results of the anti-ERα and anti-

HER2 antibody are shown as example in Figure 8 A and Figure 8 B. Next, target signals were normalized to the total protein amount per spot as assessed by staining of representative slides with the dye Fast Green FCF. This was done to account for small variance in total protein concentration due to potential pipetting and printing errors. After calculation of the median of technical replicates, normalized target signal intensities were plotted against the signal intensities obtained by incubation of subarrays without primary antibody (blank signal). Only data of antibodies with target signals which differed markedly from corresponding blank signals were used for further analysis. The results of the anti-ERα and anti-HER2 antibody are shown as example in Figure 8 C. Antibodies and respective target proteins, which passed all the aforementioned validation criteria, are listed in Table 1.

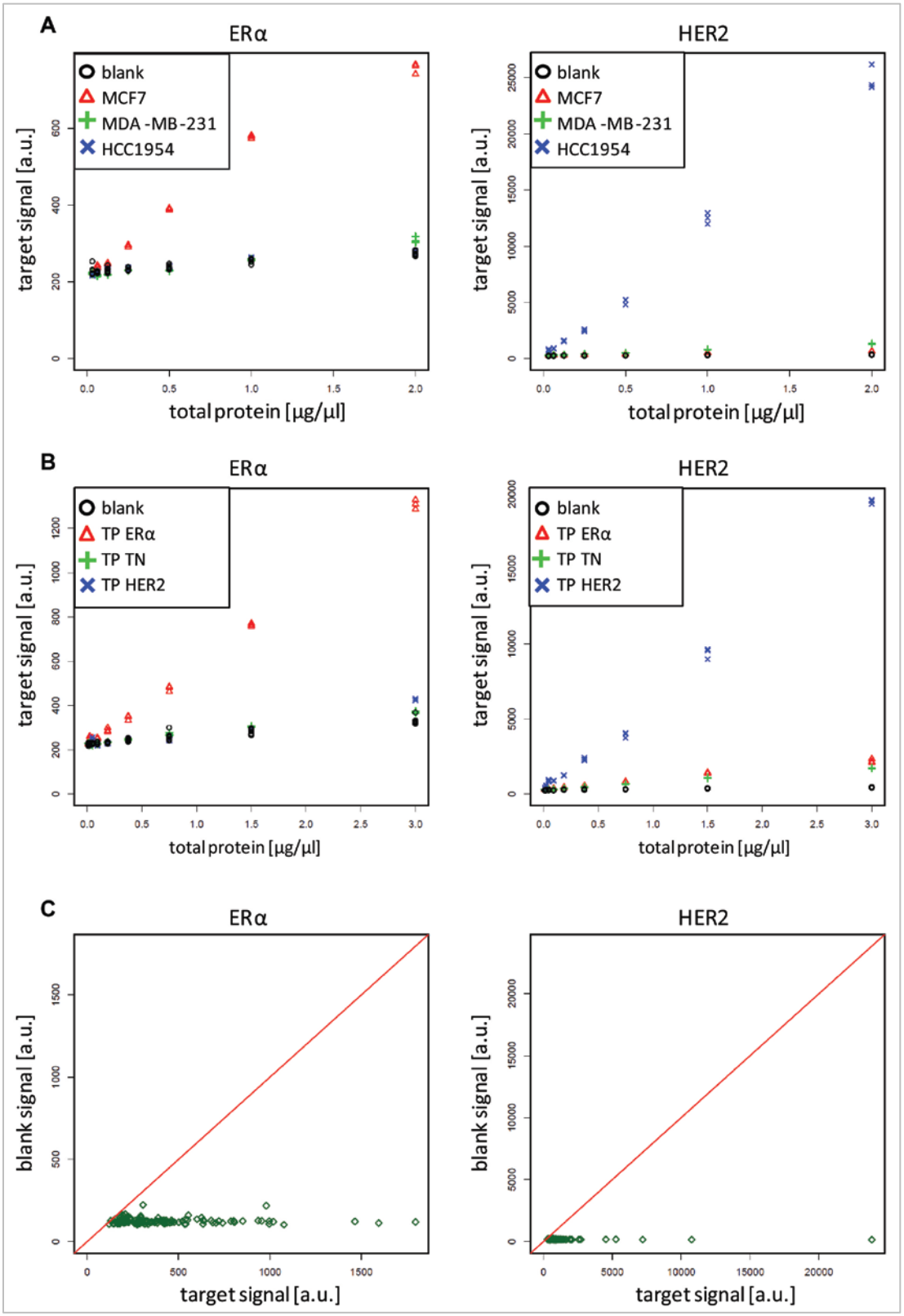


Figure 8: RPPA data quality assessment of individual antibodies. Results of the anti-ERα and anti-HER2 antibody are shown exemplarily. A linear correlation between target signal intensity and protein concentration for cell lines (A) and tumor pools (B) was observed. In addition, target signals differed markedly from corresponding blank signal (C).

2.2.8 Microspot immunoassays (printing of capture antibody slides)

Capture antibodies were printed on nitrocellulose coated glass slides using a contact printer equipped with 185 µm solid pins. The printing layout is depicted in Figure 9.

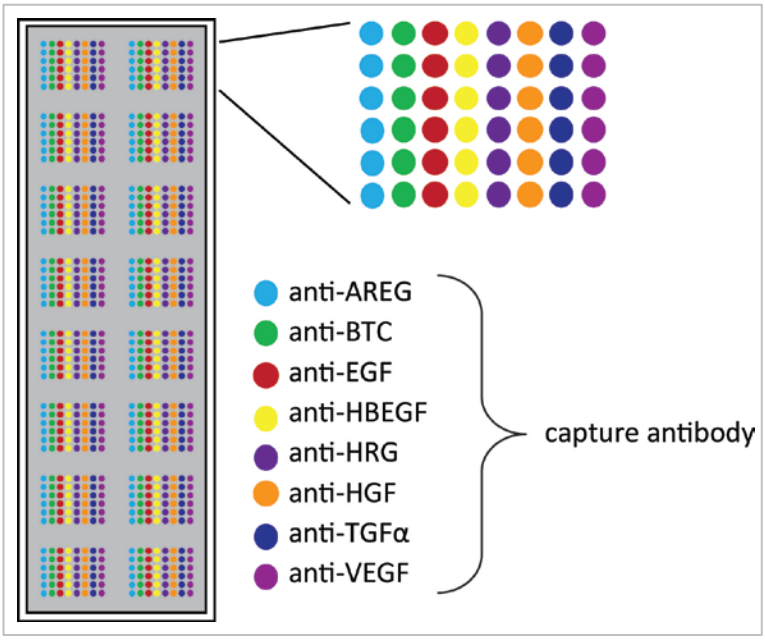
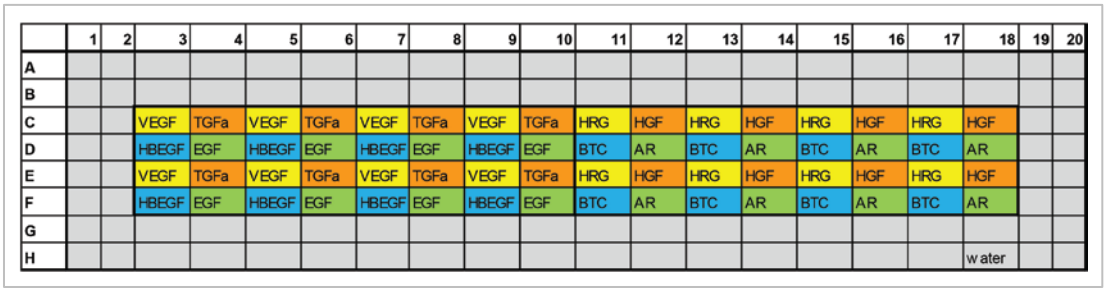


Figure 9: MIA printing layout. Identical subarrays (n = 16) were printed on nitrocellulose-coated glass slides. Each subarray consisted of eight different capture antibodies printed as six fold technical replicate.

The printhead configuration was 2 × 4 pins with 9 mm pin spacing. The 8-pins printhead configuration required that each capture antibody (7.5 µl/well) had to be present in eight different wells of the source well plate according to the scheme depicted in Figure 10. The stock solutions of anti-AREG, anti-BTC, anti-HGF, anti-HRG, and anti-VEGF capture antibodies were diluted to a concentration of 250 µg/ml for printing. In case of anti-EGF and anti-HBEGF capture antibody, the concentration was 125 µg/ml and in case of anti-HBEGF capture antibody 31 µg/ml. All capture antibodies were diluted in PBS and Tween®20 was added to a final concentration of 0.05%. All capture antibodies were printed with two depositions per spot and the humidity during the printing run was kept constant at 80%. After the print run, slides were kept at 4°C over night and were afterwards stored at room temperature in a desiccator with silica gel beads.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A																				
B																				
C			VEGF	TGFa	VEGF	TGFa	VEGF	TGFa	VEGF	TGFa	HRG	HGF	HRG	HGF	HRG	HGF	HRG	HGF		
D			HBEGF	EGF	HBEGF	EGF	HBEGF	EGF	HBEGF	EGF	BTC	AR	BTC	AR	BTC	AR	BTC	AR		
E			VEGF	TGFa	VEGF	TGFa	VEGF	TGFa	VEGF	TGFa	HRG	HGF	HRG	HGF	HRG	HGF	HRG	HGF		
F			HBEGF	EGF	HBEGF	EGF	HBEGF	EGF	HBEGF	EGF	BTC	AR	BTC	AR	BTC	AR	BTC	AR		
G																				
H																		water		

Figure 10: Capture antibody pipetting scheme. Each capture antibody solution was present in eight different wells of the source well plate according to the scheme. The surrounding wells were filled with water to avoid evaporation effects during the print run.

2.2.9 Microspot immunoassays (analyte quantification)

The following protocol is optimized for the parallel quantification of analytes in plasma of ten individual patients as technical triplicates. Capture antibody slides (n = 3) were blocked in blocking buffer over night at 4°C on a rocking platform. The next day, a recombinant protein standard mix was prepared and two-fold serially diluted to result in 8 dilution steps. In addition, a spike-in sample used as internal control was prepared with 1.5 ng/ml of recombinant AREG, BTC, HGF, HRG, and VEGF as well as 0.25 ng/ml of recombinant EGF, HBEGF, and TGFα. The plasma samples were thawed at room temperature for 10 min and diluted 1:6 with PBST. The diluted plasma samples were subsequently centrifuged at 16,000 x g for 10 min. A sample master plate was prepared by transferring 160 µl of standard mix, internal control, and plasma samples to a 96-well plate as indicated in Figure 11. The capture antibody slides were mounted in specialized incubation chambers to create distinct wells for each of the 16 subarrays/slide. Using a multichannel pipette, 150 µl of samples were transferred from the master plate to the capture antibody slides. After incubation for 2 h at 22°C and 300 rpm on a thermomixer, samples were aspirated and each subarray was washed 2 x 5 min with PBST. Afterwards, slides were removed from the incubation chamber, placed in a wash box, and washed 2 x 5 min with PBST. Slides were then incubated with detection antibody mix for 1 h followed by washing for 4 x 5 min. Slides were incubated with Alexa Fluor® 680 streptavidin (1:5000 in PBST) for 30 min and subsequently washed 4 x 5 min with PBST and 2 x 5 min with ultra pure water. All incubation and washing steps were performed on a rocking platform at room temperature.

Slides were protected from light during the whole procedure. After air drying of slides, images were obtained at an excitation wavelength of 685 nm and a resolution of 21 µm using the Odyssey® Infrared Imaging System and saved as 16-bit TIFF files.

In case of tumor lysate samples, the same microspot immunoassay protocol was applied. However, the total protein concentration of all tumor lysates was adjusted to a total protein concentration of 250 ng/µl using PBST as diluent prior to quantification of analytes.

	1	2	3 4	5	6	7 8	9	10
A	standard mix 1	standard mix 2		standard mix 1	standard mix 3		standard mix 1	standard mix 4
B	standard mix 5	standard mix 8		standard mix 6	standard mix 8		standard mix 7	standard mix 8
C	FBS/PBS (1:5)	internal control		FBS/PBS (1:5)	internal control		FBS/PBS (1:5)	internal control
D	sample 1	sample 2		sample 1	sample 2		sample 1	sample 2
E	sample 3	sample 4		sample 3	sample 4		sample 3	sample 4
F	sample 5	sample 6		sample 5	sample 6		sample 5	sample 6
G	sample 7	sample 8		sample 7	sample 8		sample 7	sample 8
H	sample 9	sample 10		sample 9	sample 10		sample 9	sample 10

Figure 11: Layout of sample master plate. The standard mix and the samples were directly transferred from the master plate to the three capture antibody slides. Standard mix 1 (highest concentration) and standard mix 8 (lowest concentration) as well as the buffer and the internal control were present on all three slides as technical replicates. Standard mix 2 to 7 were only present on one slide and distributed as indicated. Each individual sample was present as technical replicate on all three slides.

Absolute quantification of analyte concentrations involved two steps. First, slide images were analyzed with the software GenePixPro 5.0. Therefore, the slide images as well as the *gene pix array list* file (.gal) generated by the printer to map the sample location on the slide were required. The resulting signal intensity data was saved as *gene pix result* file (.gpr). Second, the software QuantProReloaded was used for automated calibration curve fitting and subsequent calculation of analyte concentrations (Jocker et al. 2010). Therefore, the .gpr files as well as a .txt file combining sample annotation and slide-specific information were required. Analyte concentrations were reported as mean of triplicate measurements.

2.2.10 RNA isolation from tumor samples

Total RNA was isolated from tumor samples using the miRNeasy Mini Kit. As first step, 700 µl QIAzol® lysis reagent and one stainless steel bead were added to frozen tumor samples (≈20 mg) followed by homogenization for 4 min at 30 Hz with a bead mill. All further protocol steps were performed according to manufacturer's instructions. Total RNA concentration as well as purity was determined using the NanoDrop ND-1000 spectrophotometer.

2.2.11 Genome-wide mRNA expression profiling

Total RNA (500 ng) of tumor samples was submitted to the DKFZ Proteomics and Genomics Core Facility for genome-wide mRNA expression profiling. There labeling of RNA and hybridization to Illumina BeadChip Sentrix arrays (Human HT-12 v4) was performed. Transcriptional profiling data was log-transformed and quantile-normalized (Bolstad et al. 2003) for further analysis.

2.2.12 Statistics

Statistical analysis including Wilcoxon rank sum test, Kruskal-Wallis test, Spearman's rank correlation, and Chi-square test was performed using R version 2.13.1 (www.R-project.org) or SigmaPlot version 12.0. P-values < 0.05 were considered as statistically significant. Cluster analysis of protein expression data and illustration as heat map was performed using R version 2.13.1 (www.R-project.org).

3. Results

3.1 Technical prerequisites for RPPA based tumor profiling

Several technical prerequisites had to be considered before printing of the breast tumor lysates. For example, optimization of the tumor lysis protocol, the evaluation of appropriate slide storage conditions, as well as selection and preparation of control samples. The resulting optimized RPPA protocol is summarized in the Materials and Methods chapter (page 33 and page 37). In addition, RPPA slides had to be printed at two different days. Reasons were on the one hand to limit time of printing in order to maintain sample integrity and printing consistency and on the other hand to produce a sufficient amount of identical RPPA slides required for the profiling of breast cancer tumor samples. To confirm the reproducibility of obtained results, representative slides of each run were incubated with primary antibodies directed against ERα, PR, HER2, Ki-67, and EGFR. The target protein levels measured for each tumor sample with slides of printing run 1 were positively correlated with the respective target protein levels of printing run 2 (Figure 12).

In addition, cluster analysis of the combined results of printing run 1 and 2 revealed no systematic difference between target proteins measured on slides of printing run 1 compared to target proteins measured on slides of printing run 2. The five proteins, which were measured using slides of both printing runs, clustered together across all analyzed tumor samples and target proteins (Figure 13). Furthermore, Figure 13 illustrates that tumor samples do not cluster based on differences in sample preparation demonstrating that cryosectioning of a subset of the samples had no influence on protein expression profiles.

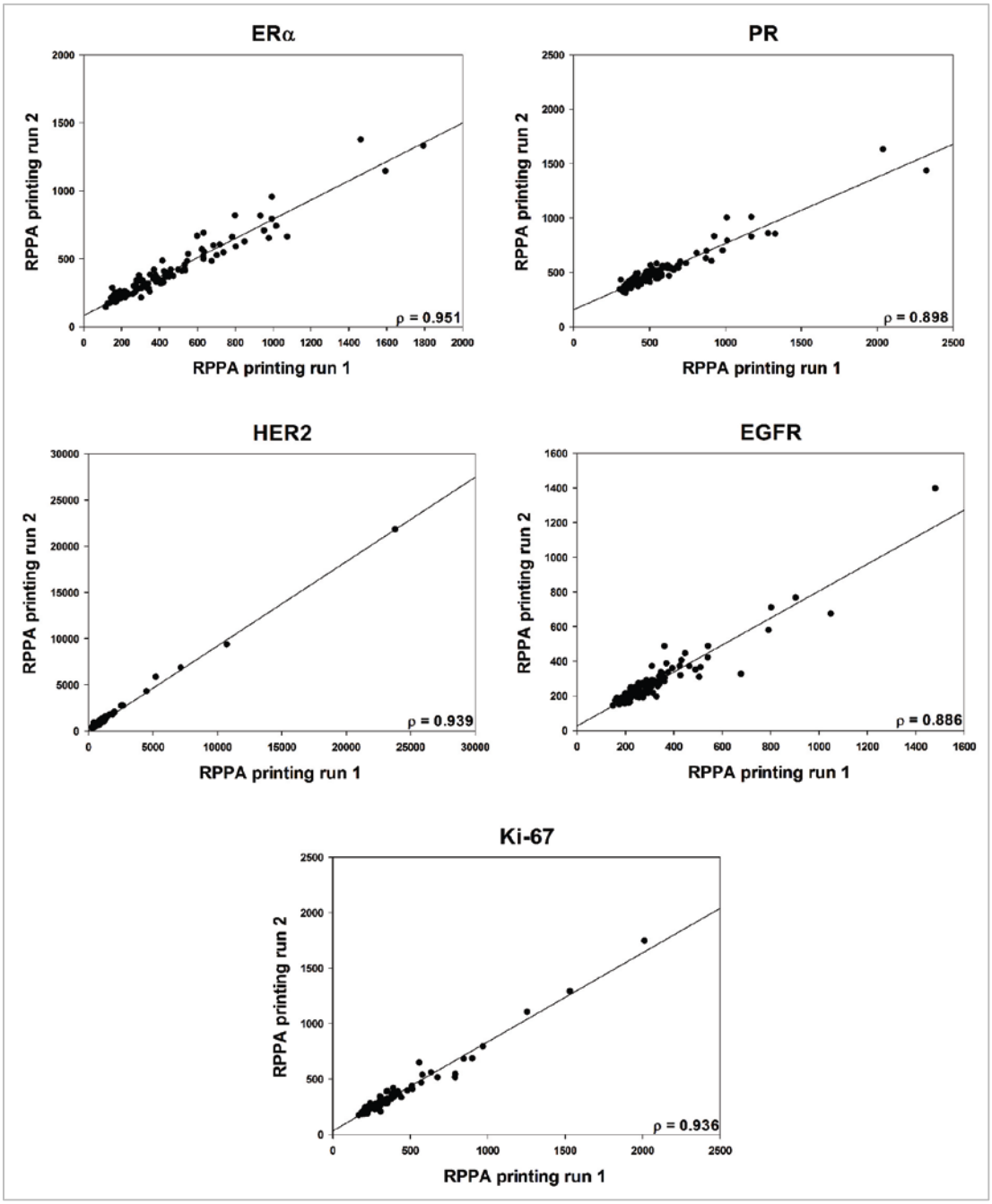


Figure 12: Correlation analysis of RPPA printing run 1 and run 2. Protein expression levels of ERα, PR, HER2, EGFR, and Ki-67 of individual patient samples measured in printing run 1 were highly correlated with corresponding protein expression levels measured in printing run 2. Spearman's rank correlation coefficients ranged from 0.886 to 0.951.

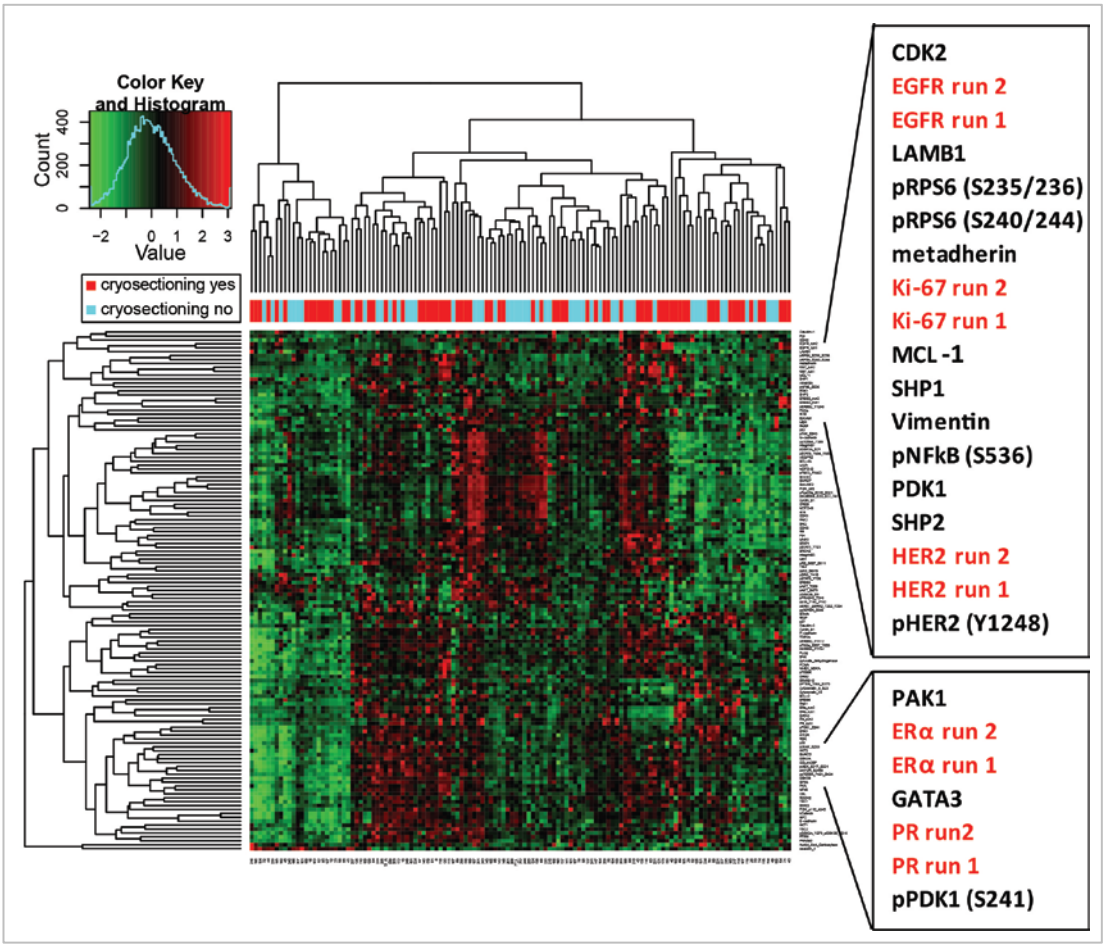


Figure 13: Heat map illustrating the reproducibility of RPPA based tumor profiling. No systematic difference between two individual printing runs with identical tumor sample set was observed. ERα, PR, HER2, EGFR, and Ki-67 protein expression levels were measured on slides of both printing runs. The respective proteins clustered together across all analyzed tumor samples and target proteins of both printing runs. In addition, tumor samples do not cluster based on difference in sample preparation demonstrating that cryosectioning of a subset of the samples had no influence on protein expression profiles.

3.2 Tumor profiling using RPPA in comparison with IHC biomarkers

Out of the initial 140 tumor samples used for RPPA based tumor profiling, data obtained for 129 tumor samples was considered for further analysis. The characteristics of the sample set are summarized in Table 7 and the exclusion criteria of the 11 samples omitted for further analysis in Table 8.

Table 7: Patient characteristics of the entire study cohort (n = 129). Median age was 63 years (range 31-86 years).

characteristic	number of patients	%
pT category (UICC 2009)		
pT1	54	42
pT2	63	49
pT3	7	5
pT4	5	4
lymph node status[1]		
positive	43	33
negative	85	66
histologic grade		
1	14	11
2	78	60
3	37	29
ERα status[2]		
positive	109	84
negative	20	16
PR status[2]		
positive	104	81
negative	25	19
HER2 status[3]		
positive	7	5
negative	122	95
triple negative[4]		
yes	15	12
no	114	88

Table 8: Samples removed from further analysis (n = 11).

patient ID	cause of removal from sample set
ID_64	ductal carcinoma in situ
ID_75	ductal carcinoma in situ
ID_135	ductal carcinoma in situ
ID_180	histology not further specified
ID_205	histology not further specified
ID_212	Angiosarcoma
ID_134	neoadjuvant treatment
ID_260	local recurrence
ID_259	RPPA profile
ID_18	not enough sample
ID_153	not enough sample

[1]Number does not add up to 129 due to a case with missing data; [2]positive: IRS 3-12, negative: IRS 0-2 (Remmele and Stegner 1987); [3]positive: IHC 3+ or average HER2 gene copy number > 6, negative: IHC 0/1+ or average HER2 gene copy number < 4 (Wolff et al. 2007); [4] negative for ERα, PR, and HER2.

Few biomarkers are used in the routine clinical setting for therapy decision of early breast cancer. These biomarkers are ERα, PR, and HER2 and their expression level is assessed with immunohistochemistry (IHC). RPPA derived expression levels of ERα, PR, and HER2 for 129 tumor samples were compared with the routine classification based on IHC for the respective target proteins. RPPA derived data significantly separated the IHC ERα-positive from the ERα-negative group of patients ($p < 0.001$, Wilcoxon rank sum test). A significant positive correlation of RPPA and IHC results could also be shown for PR and HER2 as summarized in Figure 14.

In addition, several proteins were identified by RPPA based tumor profiling to be highly associated with the subgroup of triple-negative breast cancer patients. For example, protein abundance of EGFR, MET, Vimentin, P-cadherin, Ki-67, and CDK6 was significantly higher in triple-negative breast cancer tumor samples ($p < 0.001$, Wilcoxon rank sum test). In contrast, protein abundance was lower in case of GATA3, STARD10, pCytokeratin 8 (S23), Cytokeratin 18 as well as ERα and PR ($p < 0.001$, Wilcoxon rank sum test). The differential expression of the aforementioned proteins is illustrated as boxplots in Figure 15.

In summary, known characteristics of breast cancer subtypes were reproduced using RPPA based tumor profiling. The rather small number of triple-negative samples ($n = 15$) and HER2-positive samples ($n = 7$) does not allow to draw any further conclusions about these two subgroups. For this reason, all further studies were limited to the sample group with positive ERα status as defined by routine IHC ($n = 109$).

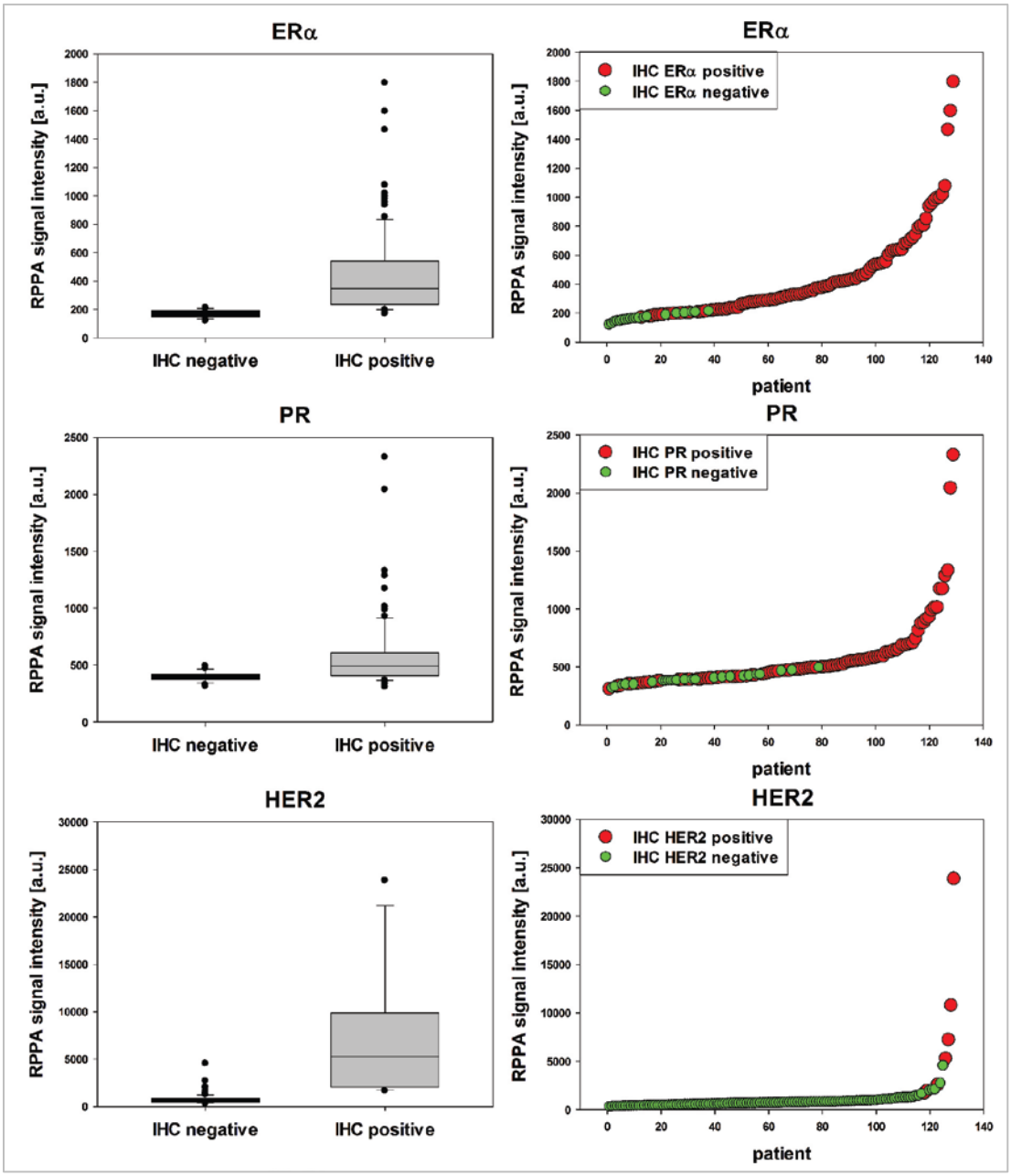


Figure 14: Comparison of RPPA data with classification of patients based on routine IHC. RPPA derived ERα, PR, and HER2 expression levels significantly separated the respective IHC based groups ($p < 0.001$, Wilcoxon rank sum test).

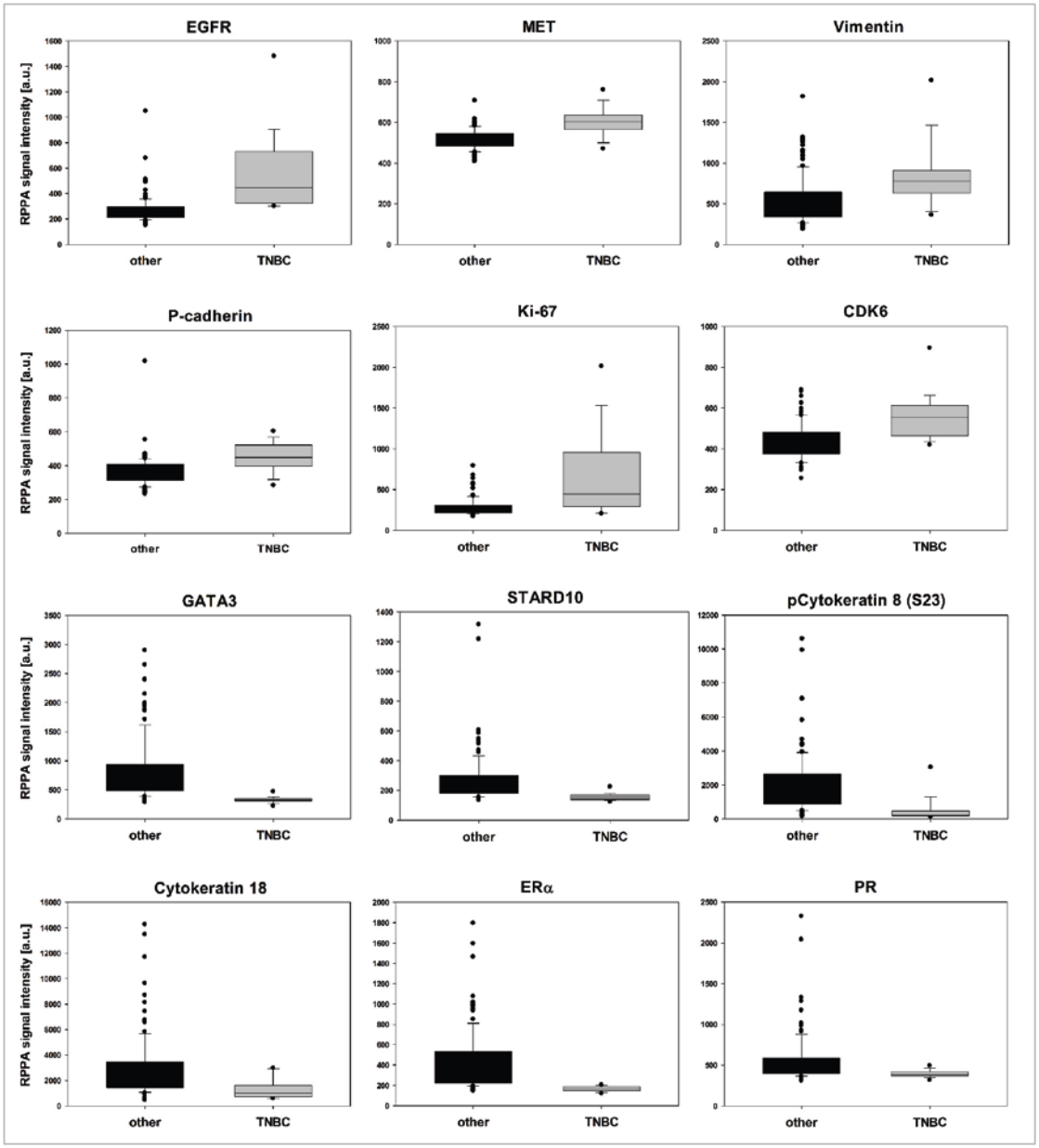


Figure 15: Protein abundance associated with triple-negative breast cancer (TNBC). High protein abundance in TNBC samples was observed for EGFR, MET, Vimentin, P-cadherin, Ki-67, and CDK6 compared to other cases. Low protein abundance was observed for GATA3, STARD10, pCytokeratin 8 (S23), Cytokeratin 18, ERα, and PR ($p < 0.001$, Wilcoxon rank sum test).

3.3 Identification of a protein biomarker signature for risk classification of hormone receptor-positive breast cancer

Tumor specimens from patients diagnosed with hormone receptor-positive primary invasive breast cancer were analyzed with RPPA aiming at the identification of a robust set of protein biomarkers to classify patients into either a low or a high risk group of cancer recurrence.

3.3.1 Biomarker identification process

Quantitative protein expression data was obtained for 128 different proteins or phosphoproteins (Table 11, appendix). Biomarker identification was based on the idea of using quantitative protein expression data of tumor samples with histologic grade 1 (n = 14) and histologic grade 3 (n = 22) as surrogates for the low and high risk group, respectively. Characteristics of this patient cohort are summarized in Table 12 (appendix).

Numerous algorithms have been described to perform two-group classification based on high-throughput gene or protein expression data in order to identify an optimal set of biomarkers. To take advantage of different characteristics of these algorithms a combination of three known classification algorithms, including a bootstrap approach on the samples, was used to classify histologic grade 1 (G1) versus grade 3 (G3). This method was developed by Christian Bender and implemented in the R programming language as R-package *bootfs* (available at https://r-forge.r-project.org/projects/bootfs/). As illustrated in Figure 16, *bootfs* employs a bootstrap approach to generate 100 bootstrap data sets out of the initial RPPA data set as first step. Next, feature sets suitable for classification were individually computed for each of the 100 bootstrap data sets using the three different classification algorithms. In detail, SCAD-SVM (Support Vector Machines using Smoothly Clipped Absolute Deviation penalty), RF-Boruta (Random Forests with Boruta algorithm for feature selection), and PAM (Prediction analysis for microarrays utilizing the nearest shrunken centroid classifier). Subsequently, all features/proteins which had been chosen by each of the three classification algorithms were included in the resulting intersected feature set. As last step, the frequency of each feature/protein was calculated across 100 intersected feature sets.

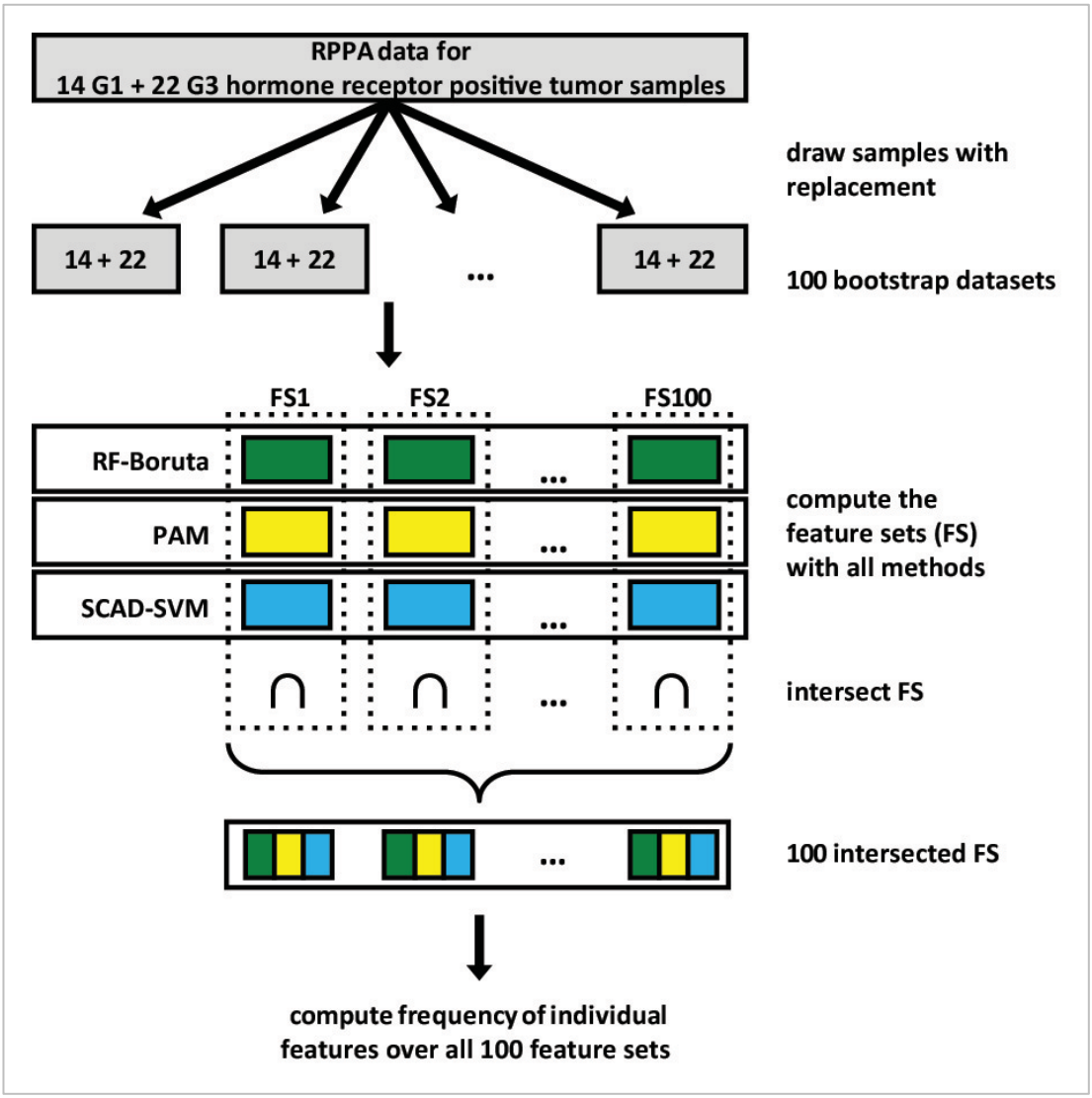


Figure 16: Biomarker (feature) selection workflow using the R-package *bootfs*. For 100 bootstrapped RPPA data sets the intersection of the three individual feature sets, computed with RF-Boruta, PAM, and SCAD-SVM, is used to generate 100 intersected feature sets. Counting the occurrences of each feature in the intersected feature sets yields the final ranking of the proteins according to their relevance for the classification. RF, Random Forests with Boruta algorithm for feature selection (Kursa and Rudnicki 2010); PAM, Prediction analysis for microarrays utilizing the nearest shrunken centroid classifier (Kursa and Rudnicki 2010); SCAD-SVM, Support Vector Machines using Smoothly Clipped Absolute Deviation penalty (Zhang et al. 2006).

The result of the *bootfs* biomarker selection process is visualized as importance graph in Figure 17A representing abundance and co-occurrence of selected features. To evaluate the robustness of biomarker selection, *bootfs* was repeated 20 times. Selected proteins were ranked according to their relative frequency and rank variation was calculated (Figure 17B).

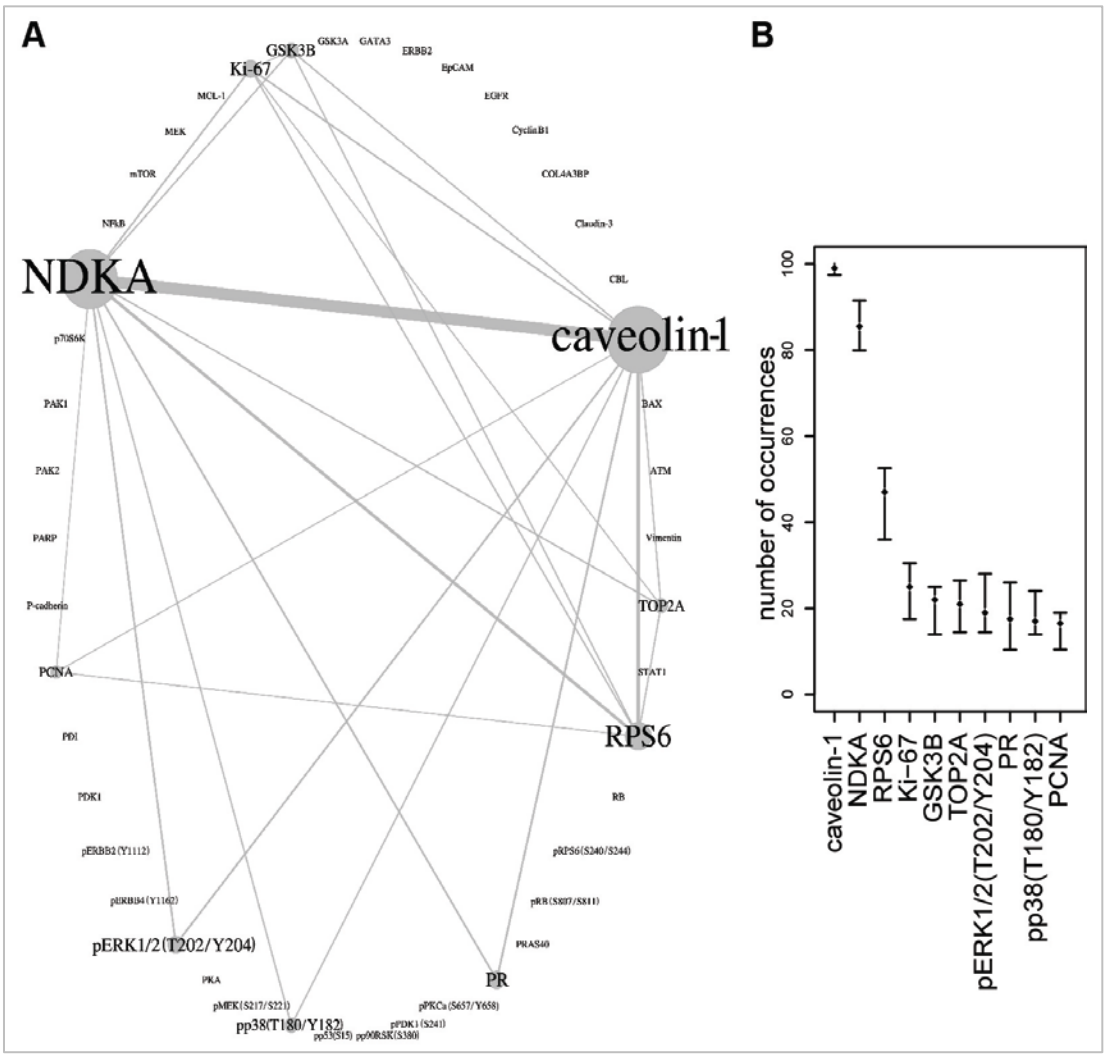


Figure 17: Summary of *bootfs* biomarker selection result. A, Importance graph: Node size represents the frequency of one protein to be present in the 100 intersected feature sets. Edge width is proportional to co-occurrence of the nodes adjacent to the edge. B, The bootstrapped feature selection was repeated 20 times with differing random seeds and the ranks of each feature across these 20 repeats are shown, sorted in decreasing order. The error bars indicate the 95% confidence interval of the observed ranks for each feature.

Ki-67, TOP2A, and PCNA were listed consistently among the top 10 hits of proteins able to discriminate between histologic G1 and G3 tumor samples. All three target proteins had a significantly higher expression level in histologic G3 samples (Figure 18A). These proteins present well known proliferation markers and confirm that a major discriminator between G1 and G3 tumors is the cell proliferation rate. However, the three top hits to classify the tumor samples were identified as caveolin-1, NDKA, and RPS6. Caveolin-1 revealed a higher expression in histologic G1 tumors, whereas NDKA and RPS6 were more strongly expressed in histologic G3 tumors (Figure 18B). The differential expression of caveolin-1, NDKA, and RPS6 identified by RPPA was subsequently confirmed by Western blot (Figure 19).

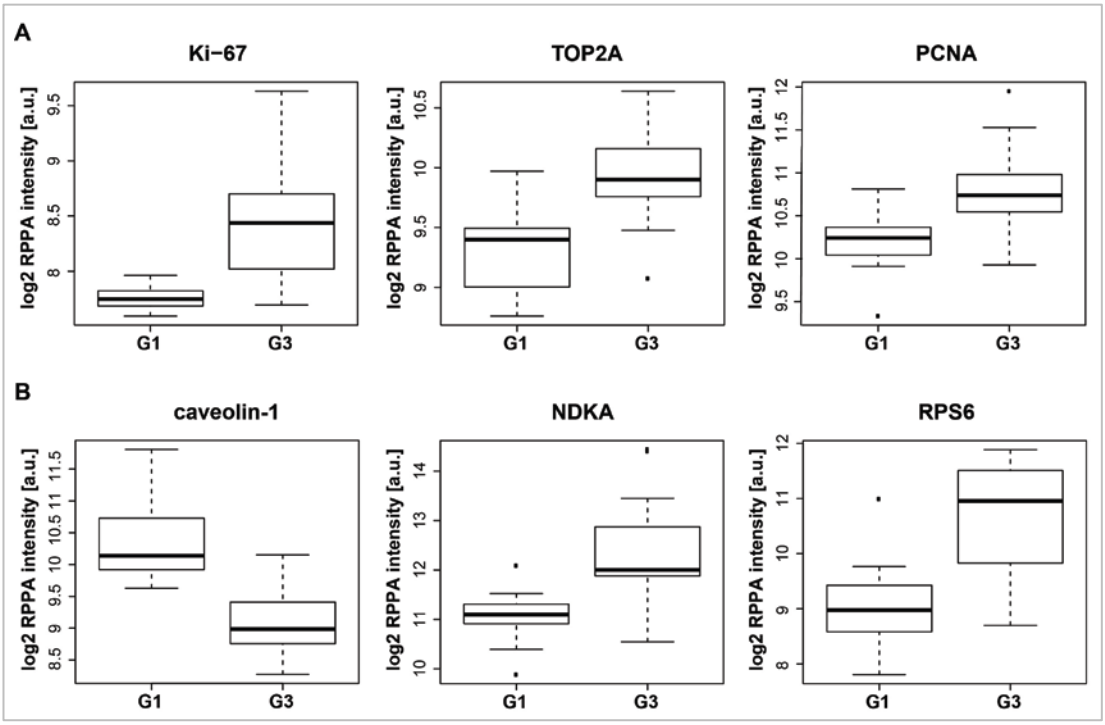


Figure 18: Differential expression of selected proteins. A, Cell proliferation marker Ki-67, TOP2A, and PCNA were among the top 10 selected proteins showing significantly higher expression in histologic G3 tumor samples ($p < 0.001$, Wilcoxon rank sum test). B, The top 3 frequently selected biomarkers, were caveolin-1, NDKA and RPS6. Caveolin-1 was significantly lower expressed in histologic G3 tumor samples whereas RPS6 and NDKA were significantly higher expressed compared to G1 tumor samples ($p < 0.001$, Wilcoxon rank sum test).

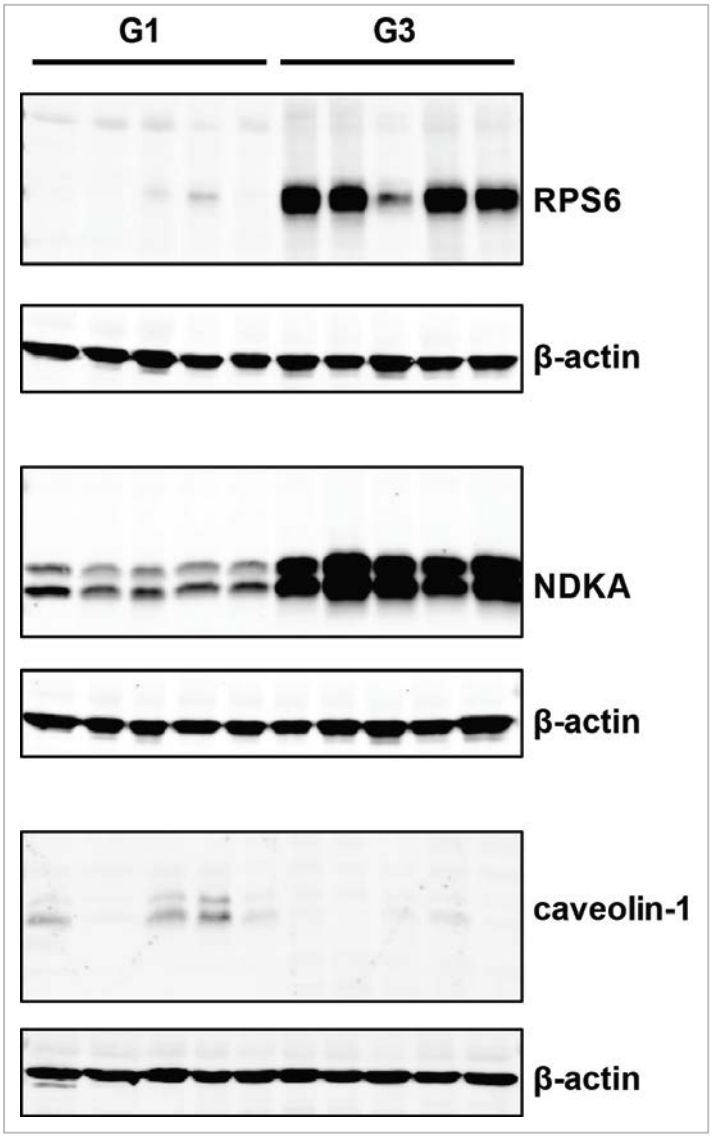


Figure 19: Western blot analysis of the three top hits RPS6, NDKA, and caveolin-1. Five tumor samples with histologic grade 1 (G1) and five tumor samples with histologic grade 3 (G3) were probed with antibodies directed against RPS6, NDKA, and caveolin-1. RPS6 and NDKA showed a higher expression in G3 tumor samples, whereas a trend of lower expression in G3 tumor samples was observed for caveolin-1. β-actin staining was used as loading control.

3.3.2 Comparison of biomarker protein and mRNA expression levels

To evaluate the selected biomarker set consisting of caveolin-1, NDKA, RPS6, and Ki-67 further, a comparison of mRNA and protein expression was carried out for a subset of 71 samples. Correlation analysis revealed that caveolin-1 mRNA and protein level were positively correlated ($p < 0.001$) with a Spearman correlation coefficient of $\rho = 0.665$. NDKA and Ki-67 also had a significant positive correlation with $\rho = 0.702$ and $\rho = 0.404$, respectively. In case of RPS6, no correlation between mRNA and protein expression was observed (Figure 20A).

A recently published gene expression data set (Curtis et al. 2012) comprising 406 estrogen receptor-positive breast cancer samples was used to compare expression levels of caveolin-1, NDKA, and Ki-67 with the annotated histologic grading status. In line with RPPA derived results, mRNA levels of caveolin-1 were significantly higher in histologic G1 samples compared to G3 samples. In addition, NDKA and Ki-67 revealed a higher expression in histologic G3 samples Figure 20B.

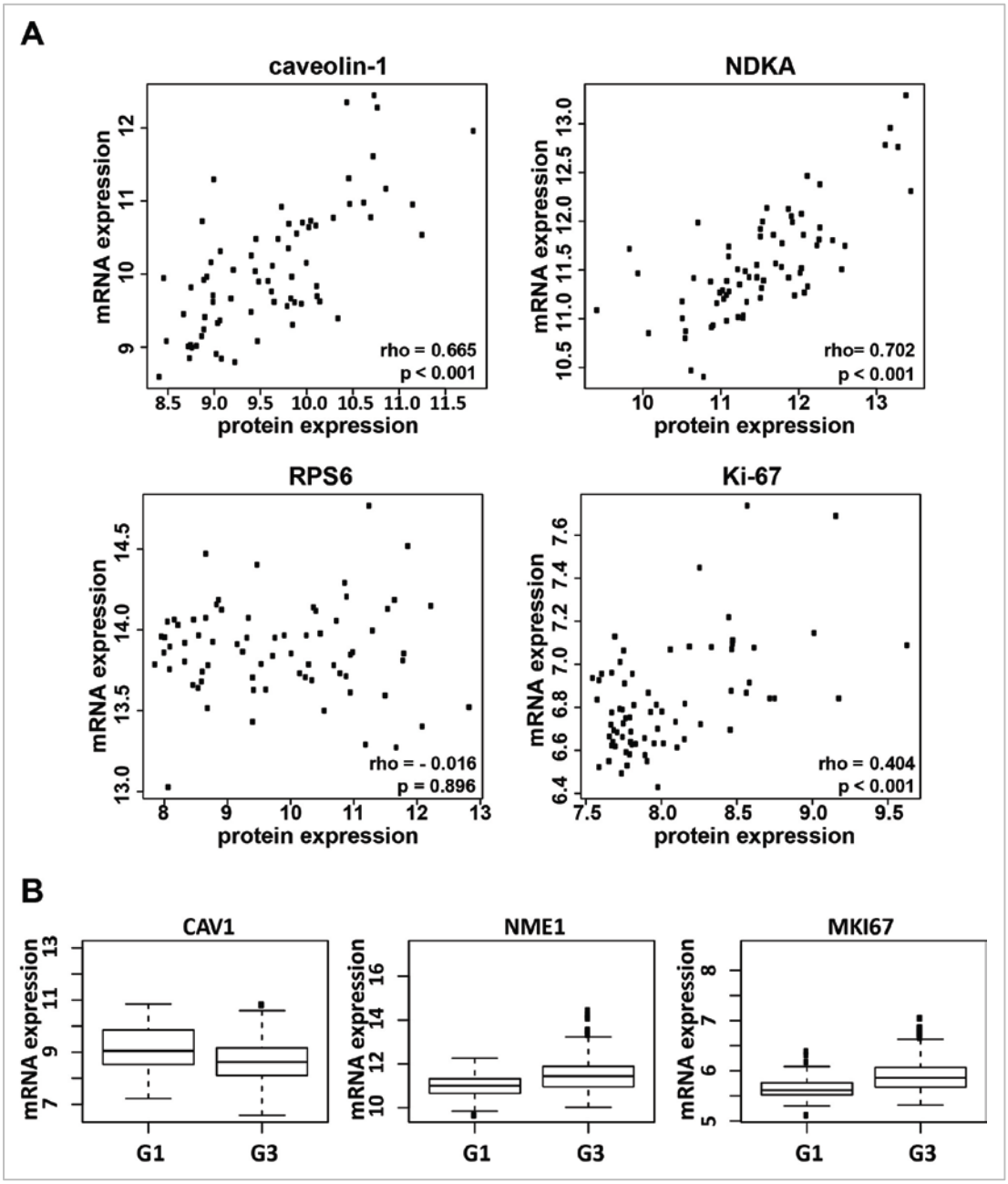


Figure 20: Comparison of biomarker protein and mRNA expression levels. A, Correlation of mRNA and protein expression derived by Illumina whole genome gene expression profiling and RPPA, respectively. Significant correlation was observed for caveolin-1, NDKA, and Ki-67 ($p < 0.001$, Spearman's rank correlation), but not for RPS6. B, Association of high NDKA (NME1) and Ki-67 (MKI67) mRNA expression with histologic G3 tumor samples as well as high caveolin-1 (CAV1) mRNA expression of histologic G1 tumor samples was confirmed using an independent sample set composed of 406 estrogen receptor-positive tumor samples (Curtis et al. 2012).

3.3.3 Two-way hierarchical cluster analysis using the selected biomarkers

Caveolin-1, NDKA, and RPS6 followed by Ki-67 were the most important proteins, as indicated by the selection workflow ranking, to discriminate between histologic G1 and G3 patients. This finding was visualized by two-way hierarchical cluster analysis (Figure 21) which separated the 36 samples in two main groups comprising either histologic G1 or histologic G3 samples.

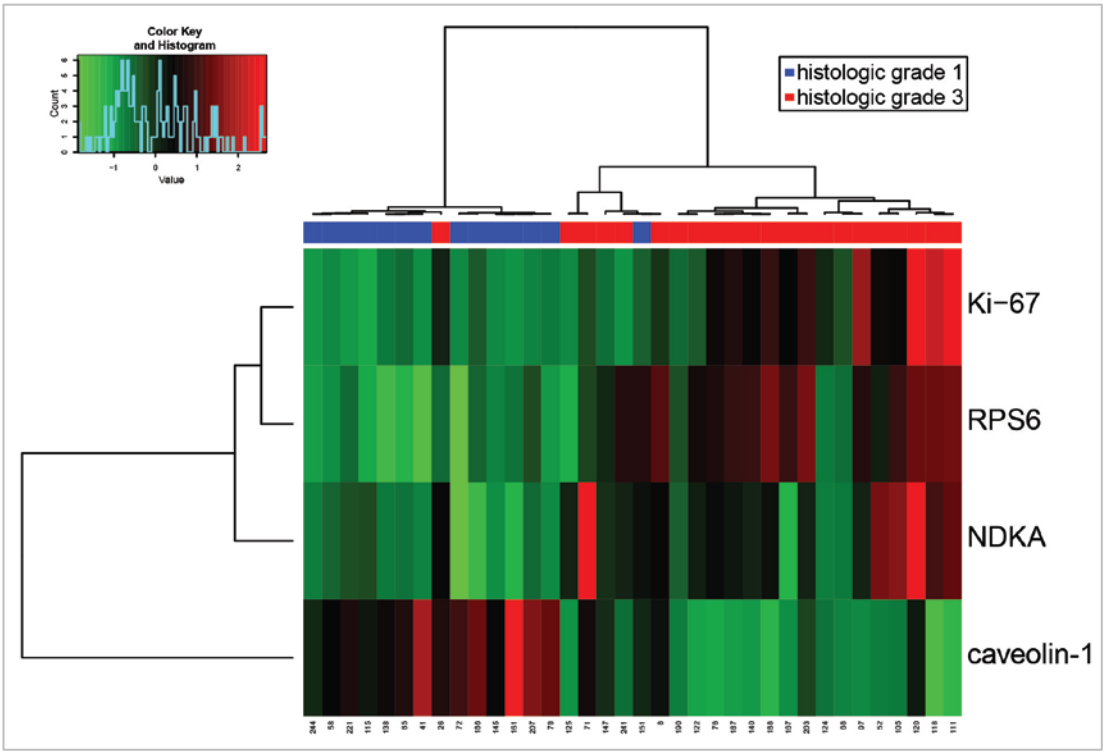


Figure 21: Hierarchical cluster analysis of histologic grade 1 and grade 3 tumor samples. Heat map illustrating a supervised two-way hierarchical cluster analysis based on correlation combined with Ward's linkage rule. Separation of 14 histologic grade 1 (blue) and 22 histologic grade 3 (red) tumor samples was achieved using RPPA derived protein expression data of caveolin-1, NDKA, RPS6, and Ki-67. Only two out of 36 samples were misclassified. Color code of the heatmap: green = low expression, red = high expression.

Protein expression levels of caveolin-1, NDKA, RPS6, and Ki-67 were next assessed by including RPPA data obtained for histologic G2 samples (patient characteristics of this cohort are summarized in Table 13, appendix). The resulting heatmap visualization revealed that histologic G2 samples cover the full expression level range and do not form a distinct group with respect to the expression of the four biomarker proteins (Figure 22). This suggests that histologic G2 patients with high level expression of NDKA, RPS6 or Ki-67 as well as low level expression of caveolin-1 are at high risk for relapse as their protein biomarker profile is highly similar to that of histologic G3 patients.

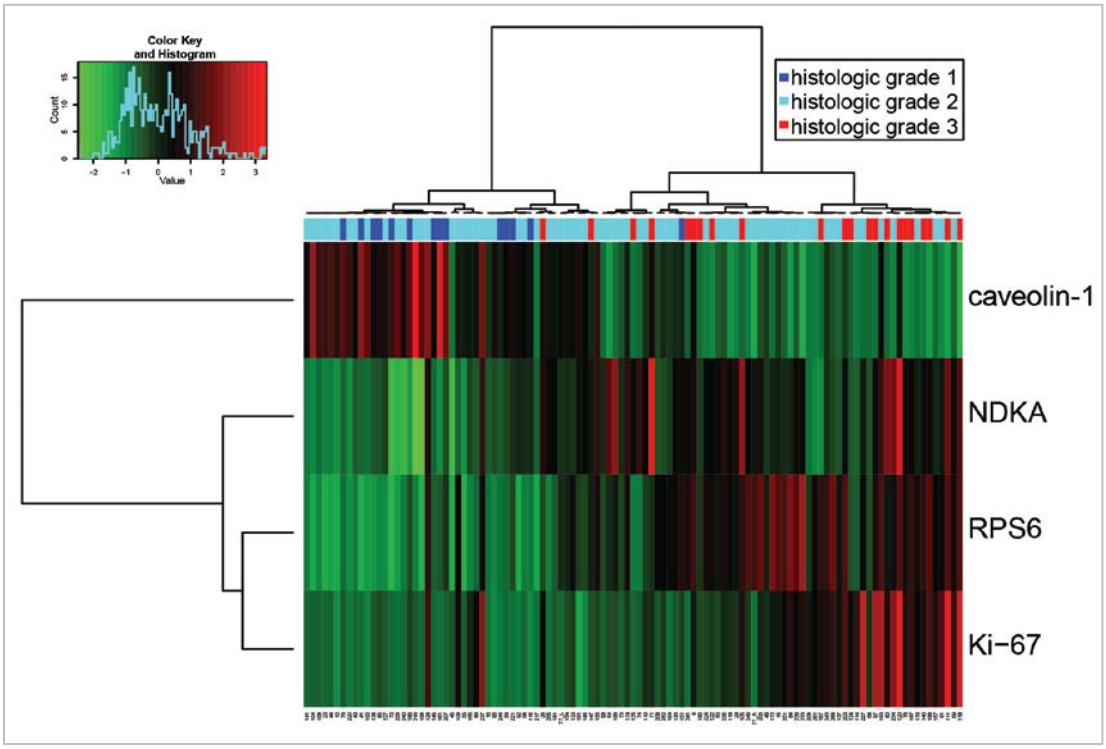


Figure 22: Hierarchical cluster analysis of histologic grade 1, grade 2, and grade 3 tumor samples. Heat map illustrating a two-way hierarchical cluster analysis based on correlation combined with Ward's linkage rule. Histologic grade 2 tumor samples (turquoise) do not represent an independent cluster, but cover the full expression range of histologic grade 1 (blue) and histologic grade 3 (red) samples. Color code of the heatmap: green = low expression, red = high expression.

3.3.4 Development of the risk classification score R2LC

To assign histologic G2 samples either as being at low or high risk of cancer relapse according to the surrogate marker profile, a risk classification score named R2LC (**R**PPA **R**isk **L**inear **C**lassification) was developed in cooperation with Christian Bender. This score is a weighted linear combination of individual biomarker expression levels which was derived by a bootstrapped linear model fit regressing histologic grade onto protein expression data of selected markers. In detail, a decision rule was defined for the risk classification by setting up a linear model for regression of the histologic grade onto the protein expression levels of the selected biomarkers. The linear model is expressed as: $\mathbf{y=X\beta+\varepsilon}$

Here, **y** is a numeric vector representing the histologic grade of each sample in the training set (14 histologic G1 and 22 histologic G3), **X** is a 36 x 4-matrix of RPPA derived protein expression values for the 36 samples and 4 selected markers, **β** is the vector of 4 coefficients to be estimated and **ε** the random error component in the model. The coefficients were estimated from 1,000 bootstrap data matrices **X'** (samples were selected randomly with replacement, while the group proportions were preserved). The final coefficients (**β**) were determined as the median of the 1,000 bootstrap coefficients for each protein (Figure 23A). These coefficients are subsequently used for calculation of the RPPA Risk Linear Classification (R2LC) score:

[R2LC]= β_caveolin-1*[caveolin-1]+ β_NDKA*[NDKA]+ β_RPS6*[RPS6]+ β_Ki-67*[Ki-67]

The decision for low risk (similar to G1) or high risk (similar to G3) based on the R2LC score is done by selecting the grade which has the smallest difference to the predicted value.

The bootstrapped estimate of the score was derived as:

[R2LC] = **- 0.464** * [caveolin-1] **+ 0.266** * [NDKA] **+ 0.194** * [RPS6] **+ 0.208** * [Ki-67]

The performance of R2LC to classify test sets was assessed additionally in a 5-fold cross validation with 10 repeats, showing good performance with AUC = 0.987 (Figure 23B).

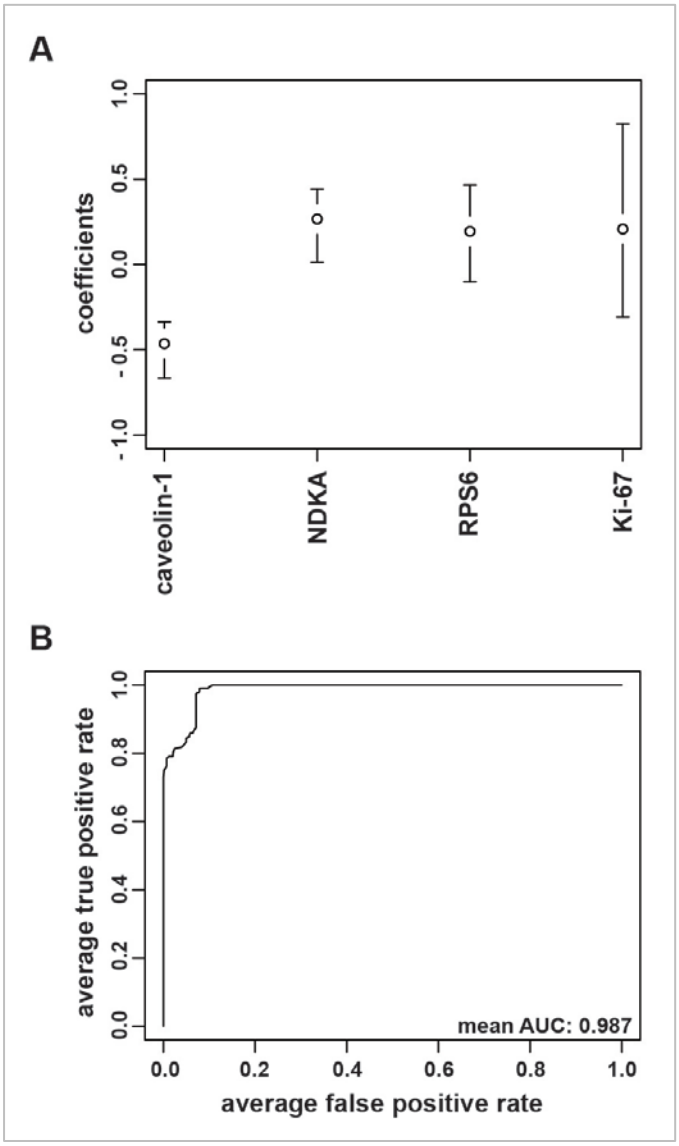


Figure 23: Result of R2LC coefficient determination and performance test of R2LC. A, R2LC coefficient estimates and respective 95% bootstrap confidence interval. B, Performance of the risk classification score R2LC to distinguish between histologic G1 and G3 samples based on the protein expression levels of caveolin-1, NDKA, RPS6, and Ki-67 was tested by 5-fold cross validation with a total of 10 repeats resulting in a mean AUC of 0.987.

Using the risk classification score R2LC, 25 out of 73 histologic G2 patients were classified as low risk whereas the other 48 patients were classified as being at high risk of recurrence. The classification result was visualized by two-way hierarchical cluster analysis of all 109 patient samples (Figure 24). The two resulting main clusters match the classification as derived by application of the risk classification score R2LC.

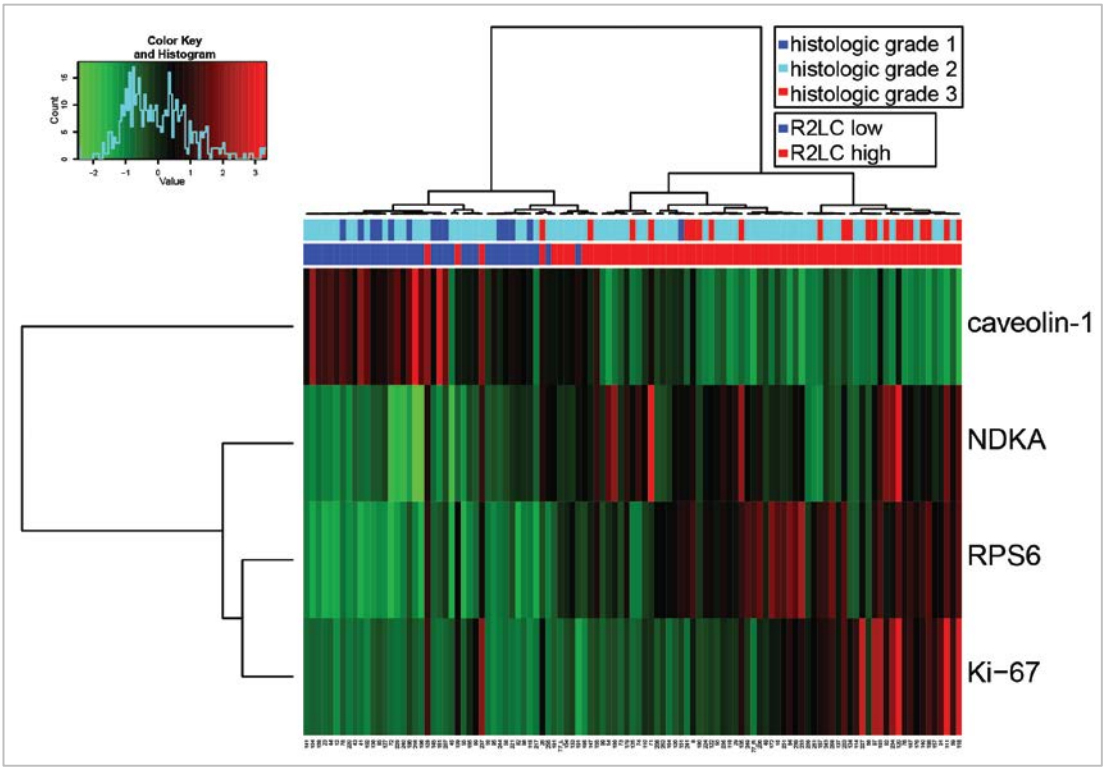


Figure 24: Hierarchical cluster analysis of histologic grade 1, grade 2, and grade 3 tumor samples in comparison with R2LC classification. Heat map illustrating a two-way hierarchical cluster analysis based on correlation combined with Ward's linkage rule. Samples classified as low risk (blue) using to the risk prediction score R2LC cluster together with the group of histologic grade 1 like tumor samples. Color code of the heatmap: green = low expression, red = high expression.

3.3.5 Comparison of R2LC based risk classification with clinical information and other experimental classification methods

The new R2LC based risk classification was compared to standard clinical parameters including age, tumor size, and lymph node status as well as recurrence-free survival. No correlation of the R2LC based risk groups was observed with age and tumor size. In case of lymph node status, a significant higher proportion of lymph node negative patients was classified as R2LC low risk. This trend, although not statistically significant, was conserved when the subgroup of histologic G2 samples was analyzed (Figure 25).

Due to the limited follow-up time, which was between 26 months and 48 months after primary diagnosis, a detailed analysis of the recurrence-free survival in comparison with the R2LC based risk groups is not yet possible. So far, eight out of 109 patients had a recurrence of breast cancer (1x local recurrence [histologic G3], 1x axilla recurrence [histologic G3], 6x distant metastasis [3x histologic G3, 3x histologic G2]). Of these eight patients six were grouped as R2LC high risk and two were grouped as R2LC low risk. However, one of the R2LC low risk patients had bilateral breast cancer, with only one tumor analyzed by RPPA.

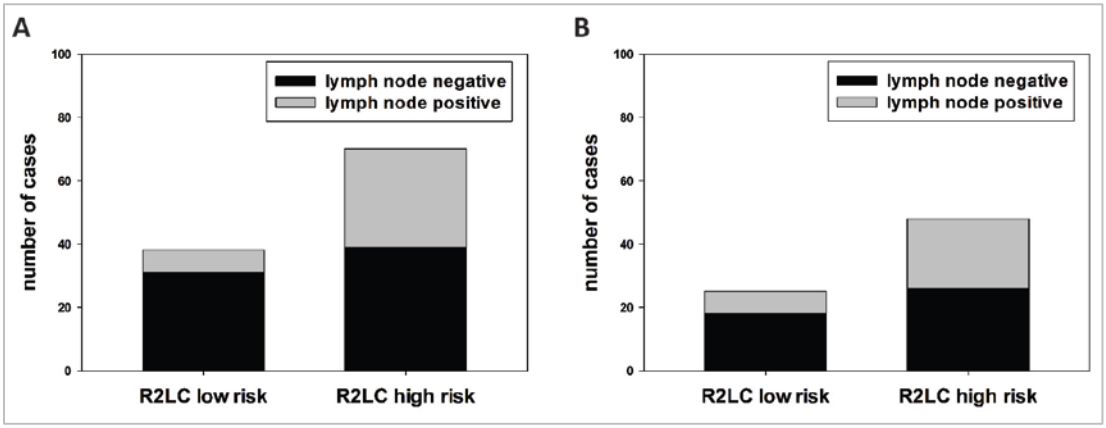


Figure 25: Comparison of R2LC based risk groups with lymph node status. A, The proportion of lymph node negative patients was significantly enriched in the R2LC low risk group (Chi-square test, p = 0.013) when analyzing the whole sample set. B, No significant difference (Chi-square test, p = 0.22) could be observed for the subgroup of patients with histologic G2 tumor. However, the trend to higher proportion of lymph node negative patients classified as R2LC low risk was preserved.

In addition, the new R2LC based risk classification was compared to other experimental classification methods like IHC Ki-67 and the gene expression signatures GGI (Sotiriou et al. 2006) and PAM50 (Parker et al. 2009). The assessment of Ki-67 expression using IHC is recommended as convenient approximation for low and high risk classification of hormone receptor-positive early stage breast cancer (Goldhirsch et al. 2011) with 14% nuclear Ki-67 staining as cut-off (Cheang et al. 2009). Ki-67 data was available for 104 of 109 samples since this marker was assessed routinely in the clinics. Of these 104 patients, 61 had a Ki-67 staining < 14% whereas the other 43 had a Ki-67 staining > 14%. For a subset of 71 samples with whole genome gene expression profiling data, the respective classification according to the gene expression signatures GGI and PAM50 was calculated with the R-package genefu (Haibe-Kains et al. 2011) by Silvia von der Heyde. Based on the GGI signature, 40 patients were classified as low risk and 31 patients as high risk. In case of the PAM50 signature, 54 patients were classified as luminal A, 16 patients as luminal B, and one patient as normal. The result of the comparison is illustrated in Figure 26. Although a good concordance was observed between R2LC based risk classification and the other three experimental classification methods, a trend for R2LC classifying more patients as high risk compared to the others was obvious.

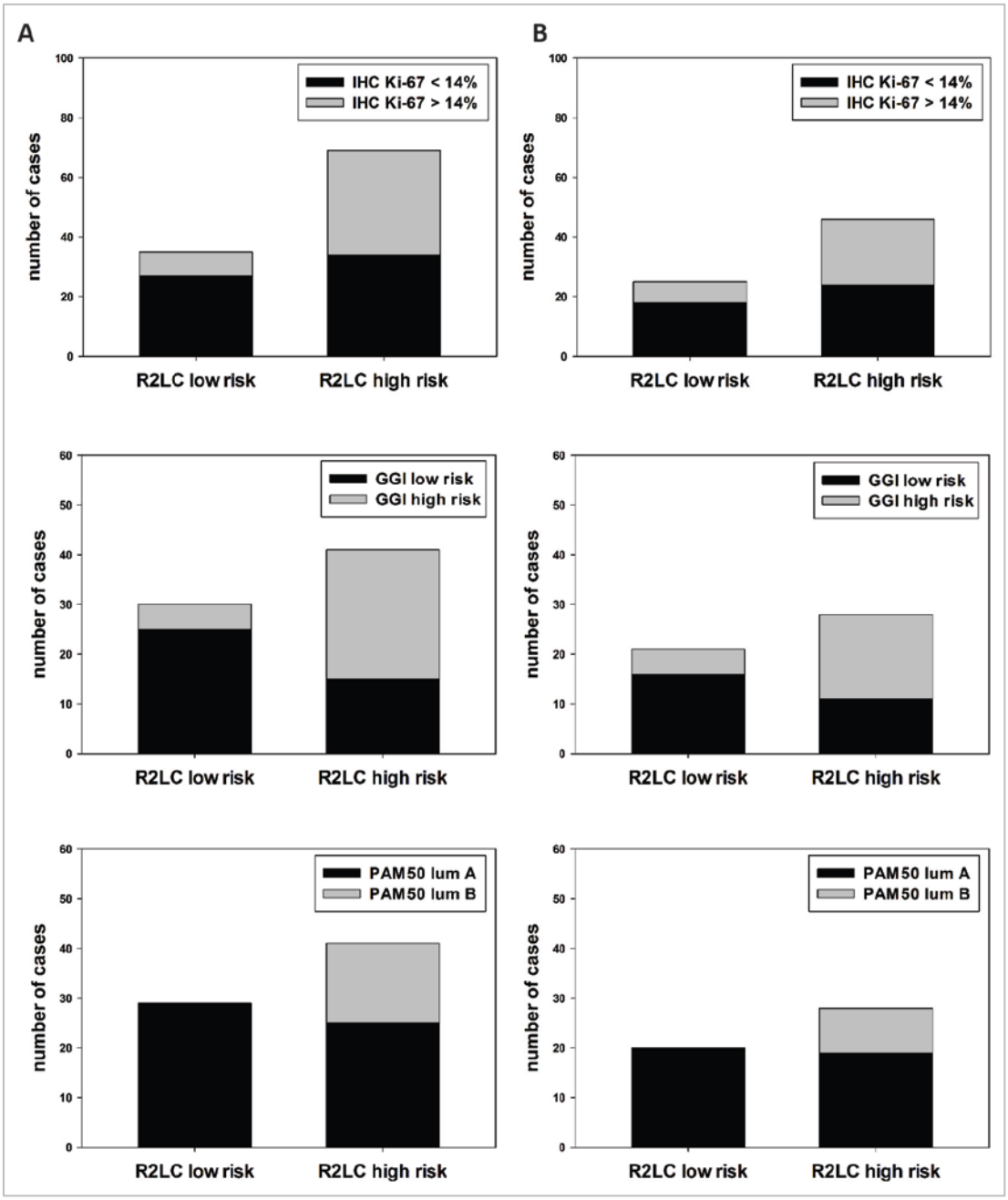


Figure 26: Comparison of R2LC based risk groups with IHC Ki-67, GGI and PAM50 classification. A, The proportion of low risk patients as defined by IHC Ki-67, GGI, or PAM50 was significantly enriched in the R2LC low risk group (Chi-square test, p = 0.012 for IHC Ki-67, p <0.001 for GGI, p < 0.001 for PAM50) when analyzing the whole sample set. B, Analysis of the subgroup of histologic G2 patients confirmed the good concordance of R2LC based risk classification with the classification based on GGI and PAM50 gene signatures (Chi-square test, p = 0.023 and p = 0.006, respectively), but not for IHC Ki-67 (Chi-square test, p = 0.171).

3.4 Development of a MIA for the quantification of eight different growth factors

Deregulation of growth factor expression plays an important role in the development and the progression of breast cancer as well in drug resistance. To evaluate the potential of a selected set of growth factors to serve as biomarkers and to elucidate the expression patterns of these growth factors in the context of breast cancer, a microspot immunoassay (MIA) was developed. Eight different growth factors can be quantified simultaneously with this 8-plex MIA. In detail, amphiregulin (AREG), betacellulin (BTC), epidermal growth factor (EGF), heparin-binding EGF-like growth factor (HBEGF), hepatocyte growth factor (HGF), heregulin 1-β (HRG or NRG1), transforming growth factor α (TGFα), and vascular endothelial growth factor ($VEGFA_{165}$ or VEGF). All steps involved in the development of this MIA are described in the following sections.

3.4.1 Selection of antibody pairs, recombinant standard proteins, standard mix diluents, and blocking buffer

A crucial step for the development of a new microspot immunoassay is the identification of suitable antibody pairs. In principle, matched antibodies pairs for conventional ELISA as proposed by the manufacturer were used, except for the HRG antibody pair. In this case a capture antibody from another manufacturer showed a better performance in terms of signal to noise ratio. The concentrations of capture and detection antibodies were adjusted for each single antibody pair in the context of the multiplexed assay. The final antibody concentrations are summarized in the Materials and Methods chapter (page 30 and page 41). In parallel, the optimal concentration range of the recombinant protein standard mix was determined according to pilot experiments and literature research. For AREG, BTC, HGF, HRG, and VEGF the analyte starting concentration was selected to be 3 ng/ml and for EGF, HBEGF, and TGFα 0.5 ng/ml.

To mimic the complex matrix of the plasma samples also for the standard mix, PBS with 20% FBS was used as diluent for the recombinant standard proteins. Compared to PBS with 1% BSA, which is commonly recommended as diluent for the standard proteins by ELISA

manufactures, no differences in signal intensities for the blank (incubation without analyte) were observed. However, signal intensities of the calibration curve resulted overall in weaker signal intensities compared to PBS with 1% BSA as diluent (Figure 27). This fact has to be kept in mind when comparing results obtained with the MIA approach presented here with data from other studies using different diluents.

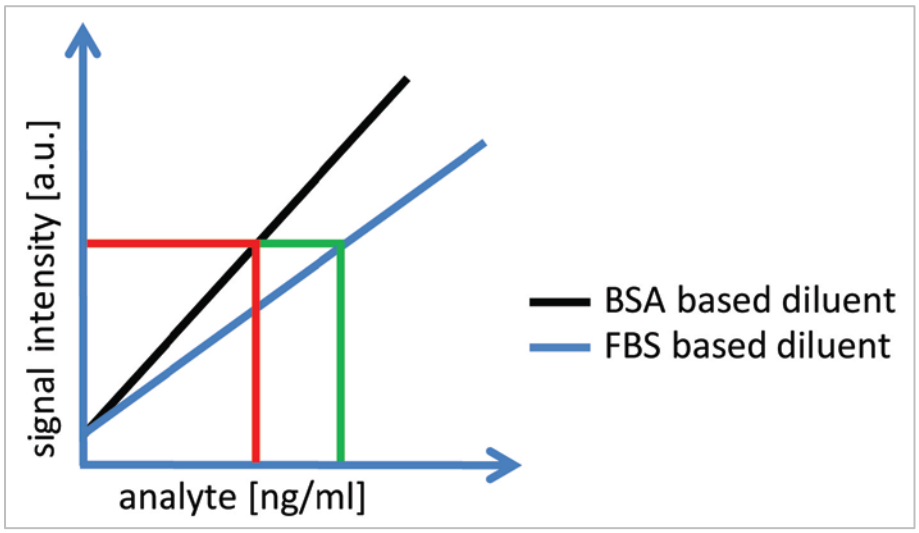


Figure 27: Comparison of 1% BSA and 20% FBS based standard mix diluents. The calibration curve with BSA based diluent (black) results in higher signal intensities compared to FBS based diluent (blue). The calculated target protein concentration of a sample is higher using the standard curve of the FBS based diluent (green) compared to using the BSA based diluent (red).

Another important step for the development of the 8-plex MIA was the choice of an appropriate blocking buffer. Inefficient blocking of free binding sides of the nitrocellulose coating results in high background signals especially with plasma samples. Therefore, three different blocking solutions were tested. In detail, “Odyssey Blocking Buffer” (LI-COR Biosciences, Lincoln, USA), “Blocking Buffer for Fluorescent Western Blotting” (Rockland, Gilbertsville, USA), and a milk based blocking buffer (5% milk, 0.5% NP40, 50 mM Tris, 150 mM NaCl, and 1 mM EDTA). All three blocking buffers showed a good performance in combination with the recombinant protein standard mix. However, assaying plasma samples after blocking with the “Odyssey Blocking Buffer” resulted in high background

signals. An optimal result was obtained with the “Blocking Buffer for Fluorescent Western Blotting”. Background signals were low with this blocking buffer as well as with the milk-based blocking buffer. However the “Blocking Buffer for Fluorescent Western Blotting” resulted in a more favourable target to background signal ratio compared to the milk-based blocking buffer (Figure 28).

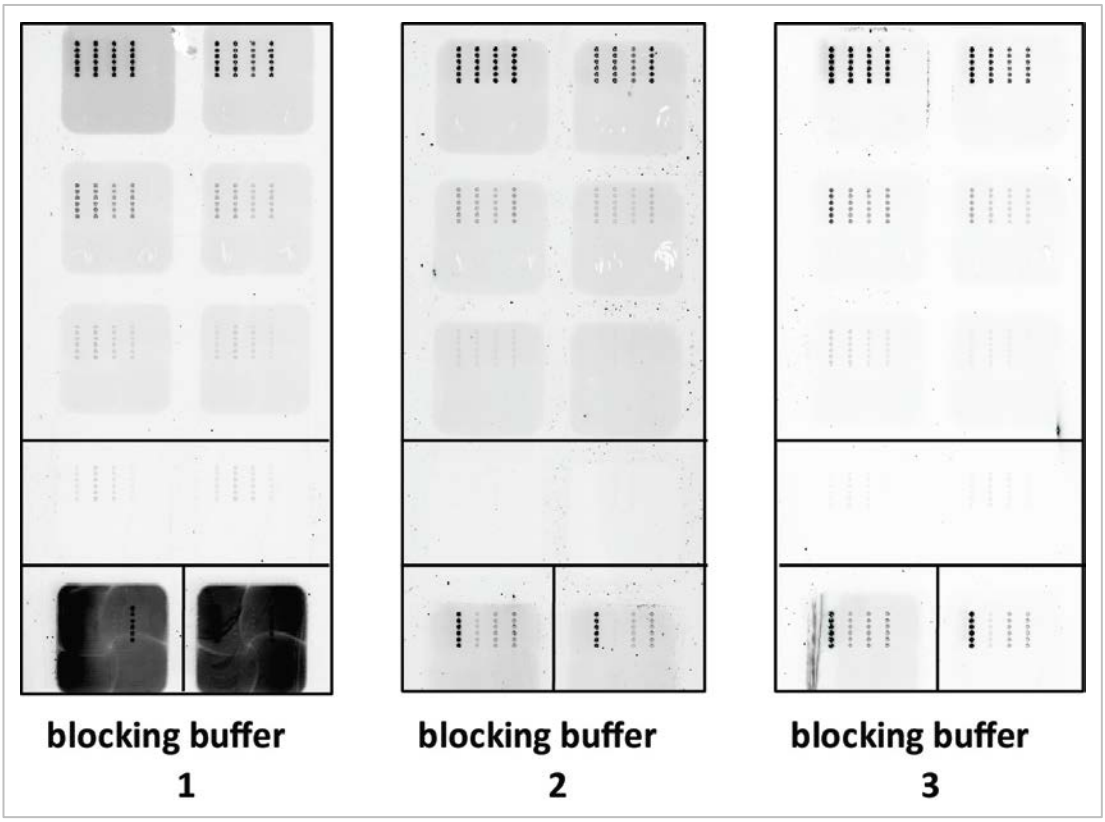


Figure 28: Comparison of three different block solutions. The upper part of each slide was incubated with a dilution series of standard proteins diluted in PBS with 20% FBS, the middle part with PBS, and the lower part with plasma samples of two individual patients in 1:5 dilution. Blocking buffer 1 = “Odyssey Blocking Buffer”; blocking buffer 2 = “milk based“; blocking buffer 3 = “Blocking Buffer for Fluorescent Western Blotting”. The best result was obtained with blocking buffer 3.

3.4.2 Test of multiplex capacity

The multiplex capacity of the 8-plex MIA antibody pairs was tested with two different assay set-ups. First, the cross reactivity of capture antibodies with different analytes was assessed by measuring three different kinds of samples. One sample contained all analytes serving as positive control, eight different samples with only one analyte (AREG, BTC, EGF, HBEGF, HGF, HRG, TGFα, or VEGF), and one sample with no analytes serving as negative control. The results of this "only one" test are shown in Figure 29. Positive signals at all capture antibody spots were detected for the sample containing all analytes and no positive signals were detected for the sample containing no analytes. For samples containing only one analyte, positive signals were detected at the matching capture antibody spots and not at other capture antibody spots.

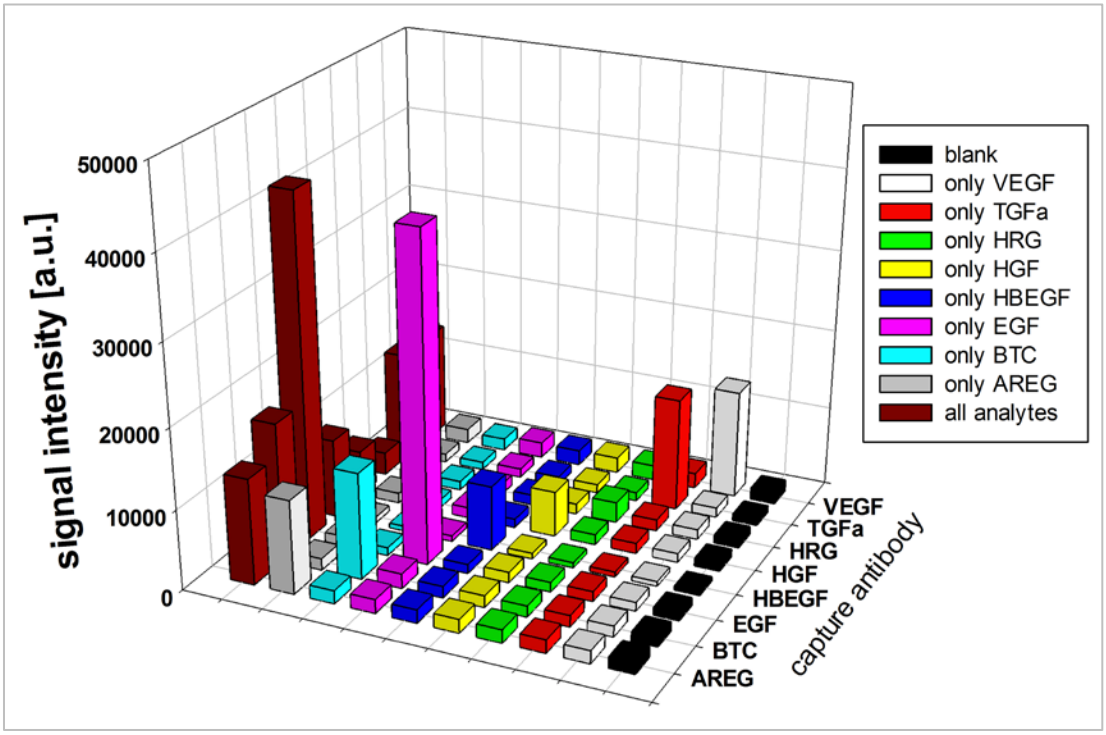


Figure 29: Multiplex capacity test "only one". Signal intensities (mean of triplicate measurements) at the respective capture antibody spots are shown for three different kinds of samples. The sample "all analytes" contained 1.5 ng/ml of AREG, BTC, HGF, HRG, and VEGF as well as 0.25 ng/ml of EGF, HBEGF, and TGFα. The sample "blank" contained none of the eight analytes. The different "only one" samples contained only one of the analytes with the aforementioned concentration.

Second, it was tested how the complete mixture of all antigens and detection antibodies contributes to the overall signal of an individual assay. Therefore three different kinds of samples were prepared. One sample contained all analytes serving as positive control, one sample with no analytes serving as negative control, and eight different samples containing “all minus one” analytes. The results of this “all minus one” test are shown in Figure 30. Removing the specific analyte from the 8-plex mixture reduced the signal intensity to the background level at the respective capture antibody spot.

These two multiplex capacity tests showed that the signal intensities measured at a capture antibody spot are specific for the respective analyte and are not caused by other components of the 8-plex MIA.

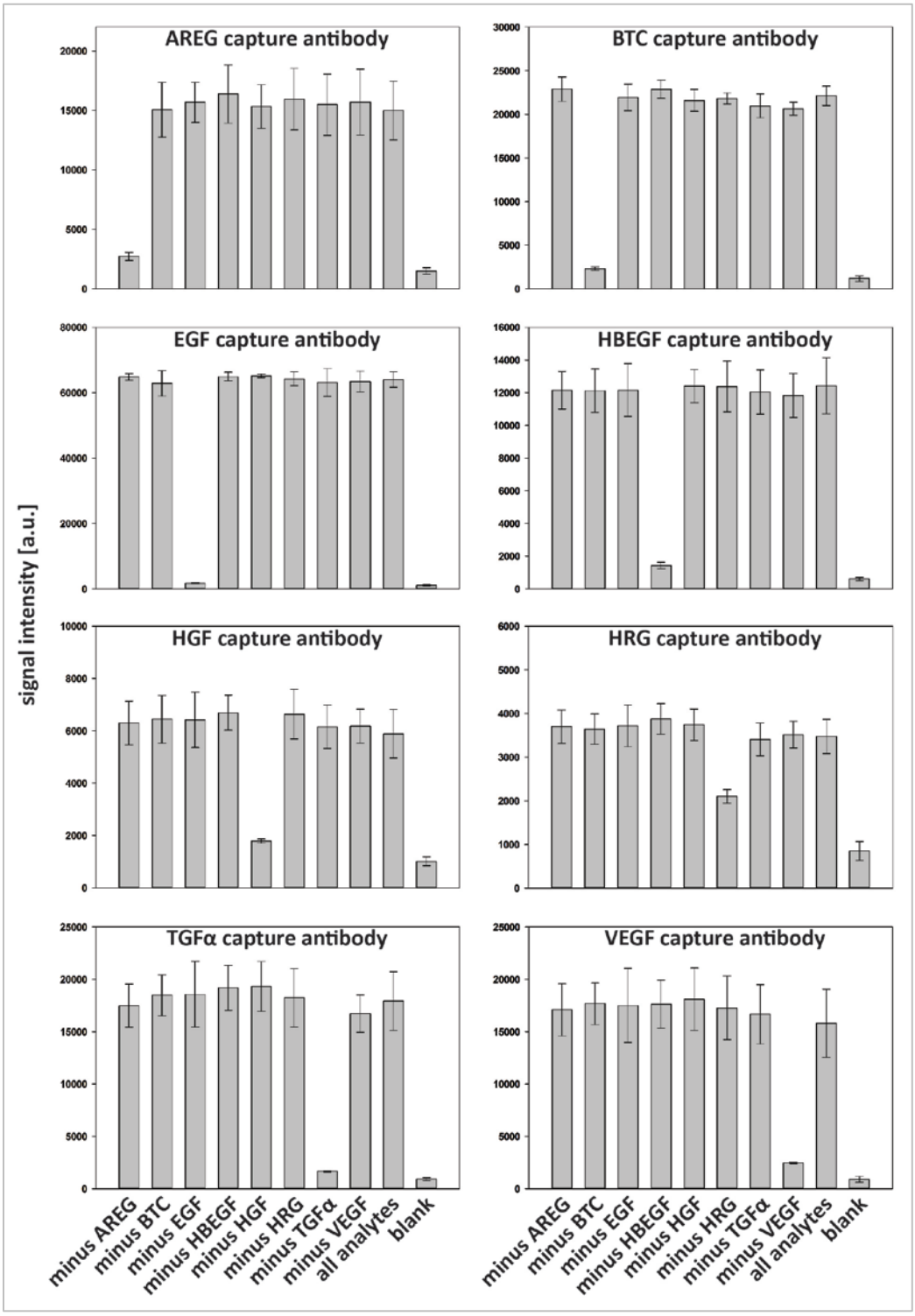


Figure 30: Multiplex capacity test "all minus one". Signal intensities (mean ± SD, n = 3) at the respective capture antibody spots are shown for three different kinds of samples. The sample "all analytes" contained 1.5 ng/ml of AREG, BTC, HGF, HRG, and VEGF as well as 0.25 ng/ml of EGF, HBEGF, and TGFα. The sample "blank" contained none of the eight analytes. The different "minus one" samples contained seven of the aforementioned analytes.

3.4.3 Development of the analysis software QuantProReloaded

The software QuantProReloaded was developed in cooperation with Anika Jöcker to facilitate the analysis of data from MIA experiments (Jocker et al. 2010) and is available for download at http://code.google.com/p/quantproreloaded/. QuantProReloaded, which is based on the ideas of QuantPro (Korf et al. 2008a), was completely redesigned to include two different functions combined with a user-friendly interface. The “accuracy analysis” function was used to calculate how accurately the analyte concentration could be determined for each single assay at a given concentration. Therefore, a re-estimation approach was employed by repeating the following procedure several times. The calibrator curve was calculated with 60% of the calibrator data points and the remaining data points were used to re-estimate the analyte concentration. The “measurement analysis” function is used for automated calibration curve fitting and subsequent calculation of analyte concentrations of the samples. In addition, this function calculates the lower limit of detection (LOD) for each single assay. The LOD was defined as the mean plus three times the standard deviation of the analyte free blank measurement.

3.4.4 Definition of a standard operating procedure for the 8-plex MIA

After development of the 8-plex MIA as described in the previous sections, a standard operating procedure (SOP) was defined. This SOP assures consistent and reproducible results of 8-plex MIA experiments. It covers the whole process from printing of capture antibody slides to calculation of analyte concentrations and includes also several quality control steps. In principle, this SOP is described in detail in chapter 2.2.8 and 2.2.9, but some special aspects will be highlighted in this section.

In total, 20 slides were printed at each individual capture antibody slide print run. The first and the last slide of each print run were used to check the quality of the print run. Therefore, spike-in samples with known concentration of recombinant human AREG, BTC, EGF, HBEGF, HGF, HRG, TGFα, and VEGF were measured in duplicate on each slide. The recovery rate of the spike-in controls and the variability between replicates was calculated for each slide and reported in the format shown in Figure 31. Only if the mean spike-in recovery rate was between 75% and 125% and the coefficient of variation (CV) of replicate measurements was less than 25%, the remaining slides of the print run were used for the measurement of the actual samples.

AREG [pg/ml]		BTC [pg/ml]		EGF [pg/ml]	
control_a	1477	control_a	1314	control_a	230
control_b	1751	control_b	1604	control_b	238
control_a recovery	98%	control_a recovery	88%	control_a recovery	92%
control_b recovery	117%	control_b recovery	107%	control_b recovery	95%
mean control	1614	mean control	1459	mean control	234
mean control recovery	108%	mean control recovery	97%	mean control recovery	94%
mean SD	194	mean SD	205	mean SD	6
control %CV	12%	control %CV	14%	control %CV	3%
HBEGF [pg/ml]		HGF [pg/ml]		HRG [pg/ml]	
control_a	256	control_a	1619	control_a	1746
control_b	281	control_b	1554	control_b	1897
control_a recovery	103%	control_a recovery	108%	control_a recovery	116%
control_b recovery	113%	control_b recovery	104%	control_b recovery	126%
mean control	269	mean control	1587	mean control	1821
mean control recovery	108%	mean control recovery	106%	mean control recovery	121%
mean SD	18	mean SD	46	mean SD	107
control %CV	7%	control %CV	3%	control %CV	6%
TGFa [pg/ml]		VEGF [pg/ml]		CV < 25%	
control_a	250	control_a	1623	25% ≤ CV ≤ 30%	
control_b	250	control_b	1520	CV > 30%	
control_a recovery	100%	control_a recovery	108%		
control_b recovery	100%	control_b recovery	101%	75% < recovery < 125%	
mean control	250	mean control	1571	70% ≤ recovery ≤ 75%	
mean control recovery	100%	mean control recovery	105%	125% ≤ recovery ≤ 130%	
mean SD	0	mean SD	73	recovery > 130%	
control %CV	0%	control %CV	5%	recovery < 70%	

Figure 31: Print run quality report form. The results of one slide are shown exemplarily. Control_a and control_b were spike-in samples with a mixture of recombinant proteins. In detail, 1.5 ng/ml of AREG, BTC, HGF, HRG, and VEGF as well as 0.25 ng/ml of EGF, HBEGF, and TGFα. The recovery rates for each analyte as well as the coefficient of variation (CV) for replicate measurements were reported.

Triplicate samples were always measured on three different slides, to avoid bias resulting from potential slide-effects (see Figure 11 for detailed sample layout, chapter 2.2.9). The CV of triplicate measurements was calculated for each of the eight analytes measured. If the CV was > 25% for one or several analytes, the measurement of this particular sample was repeated. The measurement of all samples was repeated, if the spike-in recovery of the internal control, which was measured on each silde, was not in the range of 80% to 120%. The upper limit of detection (ULOD) of each analyte was defined by the respective highest concentration used for the calibration curve. If a sample revealed a higher concentration than the respective ULOD for one or several analytes, the measurement of this sample was repeated with a higher dilution factor.

The comparability of 8-plex MIA data within one experiment but generated at different days and with slides from different print run batches was controlled by implementing the following steps. The LLOD, which was initially calculated for each analyte and calibration curve separately, was combined to an experiment specific LLOD. This experiment specific LLOD was defined as the mean of the individual LLOD. The experiment specific LLOD of the 8-plex MIA as derived for the "growth factors in plasma samples" experiment (details and results are described in section 3.6) are shown in Table 9.

Table 9: Experiment-specific lower limit of detection (LLOD) of the 8-plex MIA. The experiment-specific LLOD represents the median of 51 individual LLODs as derived for the "growth factors in plasma samples" experiment.

analyte	LLOD [pg/ml]
AREG	317
BTC	127
EGF	25
HBEGF	12
HGF	155
HRG	290
TGFα	5
VEGF	148

Different recombinant protein standard mix batches were used within one experiment. However each new standard mix batch was compared to the former. Therefore, spike-in samples of the old standard mix were prepared and the analyte concentration was measured at two different days using the calibration curve based on the new standard mix. By comparing results, correction factor was derived and applied to all samples measured with the new standard mix batch as reference.

In summary, a MIA for the quantification of eight different growth factors was successfully established. This MIA is compatible with tumor lysates and blood plasma samples and results of these experiments are described in the next two chapters. A more general SOP for the measurement of analytes in plasma samples using the MIA approach is described in the book “Protein Microarrays: Methods and Protocols” (Sonntag et al. 2011).

3.5 Measurement of growth factor concentrations in tumor samples

Growth factor concentrations in tumor lysates were determined using the 8-plex MIA. Lysates of 82 tumors were available for growth factor measurements in sufficient amount. All tumors were part of the initial 109 ERα-positive tumor samples analyzed by RPPA. The clinical characteristics of this patient cohort are summarized in Table 15 (appendix). AREG, HGF, and VEGF were detectable in over 10% of the tumor lysates. Results of these three growth factors were considered for further analysis. First, growth factor levels were compared with menopause status, tumor size, lymph node status, and histologic grade. In contrast to AREG, HGF as well as VEGF showed a correlation with histologic grade. High levels of VEGF were associated with histologic G3 tumor samples ($p < 0.01$, Kruskal–Wallis test) whereas high levels of HGF were associated with histologic G1 tumor samples ($p = 0.047$, Kruskal–Wallis test) as illustrated in Figure 32. No significant correlation with menopause status, tumor size, and lymph node status could be observed for any of the measured growth factors.

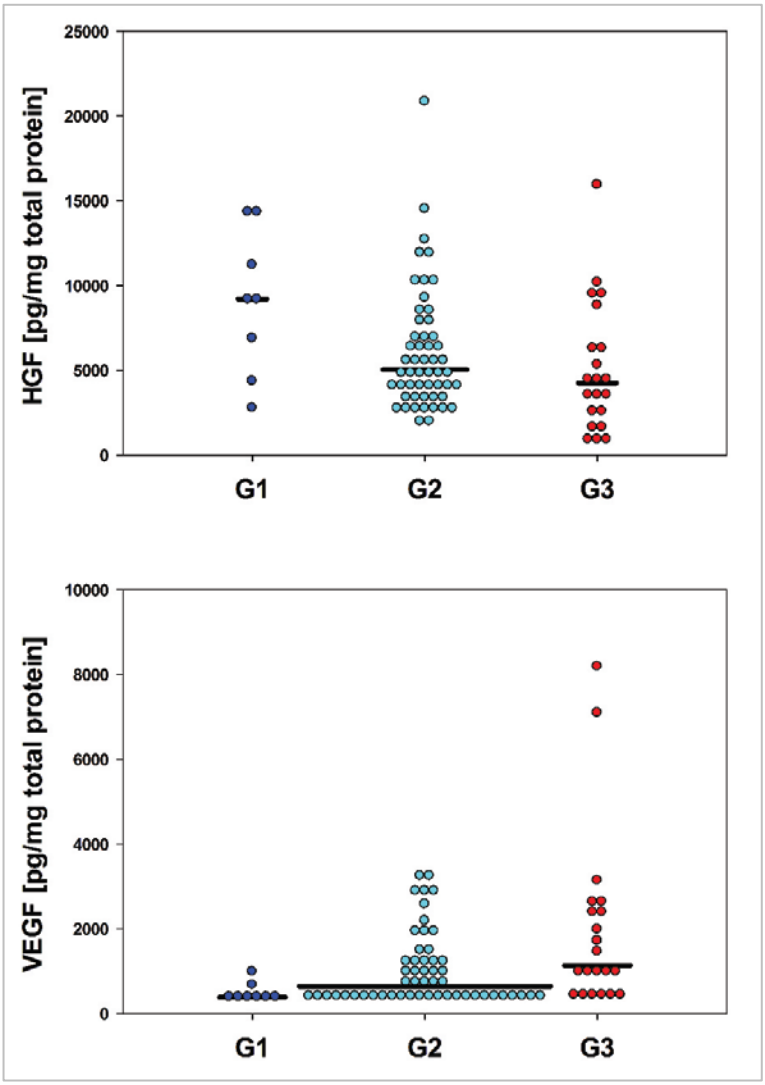


Figure 32: Comparison of growth factor concentrations with histologic grade. The HGF concentration of tumor lysates was significantly negative correlated with histologic grade whereas VEGF concentration was significantly positive correlated (Kruskal-Wallis test, $p = 0.047$ and $p = 0.01$, respectively).

In addition, AREG, HGF, and VEGF tumor lysate levels were compared between the newly defined R2LC risk groups. No difference between the low risk and the high risk group was observed in case of AREG. HGF measurements revealed a trend of lower levels in high risk tumors, although the result was not statistically significant. In contrast, VEGF was highly overexpressed in the R2LC high risk group compared to the low risk group (Figure 33). This difference between the R2LC low risk and high risk group was also apparent when analyzing the subgroup of histologic G2 tumor samples (Figure 34).

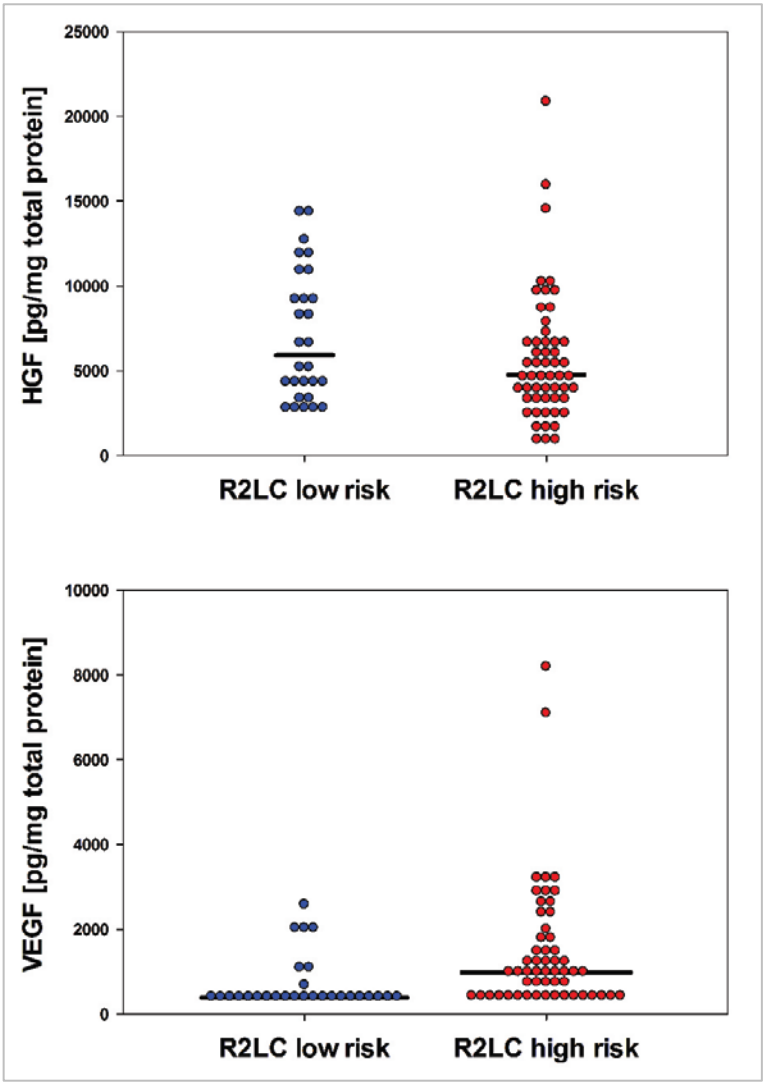


Figure 33: Comparison of growth factor concentrations with R2LC risk groups. A trend towards lower HGF levels in the R2LC high risk group was observed whereas VEGF concentration was significantly elevated in the R2LC high risk group (Wilcoxon rank sum test, $p = 0.121$ and $p < 0.001$, respectively).

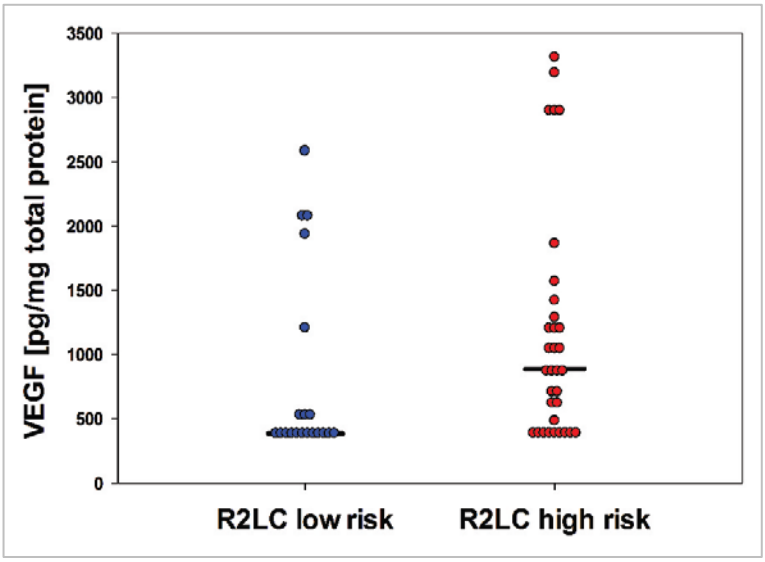


Figure 34: Comparison of VEGF levels between R2LC risk groups (only histologic G2). The VEGF concentration was significantly elevated in the R2LC high risk group compared to the R2LC low risk group (Wilcoxon rank sum test, p =0.03).

3.6 Measurement of growth factor concentrations in plasma samples

Growth factor concentrations in blood plasma samples of 282 individual patients were measured using the 8-plex MIA. Samples of patients with neoadjuvant treatment (n = 35) and further 30 samples of patients with various exclusion criteria (summarized in Table 16, appendix) were omitted for further analysis. Of the remaining sample set, 21 plasma samples were from patients with a benign tumor and 196 plasma samples from patients with invasive primary breast cancer. Patient characteristics for the group with invasive primary breast cancer are summarized in Table 17 (appendix).

In case of AREG, 63% of the plasma samples had an analyte concentration under the lower limit of quantification. For BTC, EGF, HBEGF, HGF, HRG, TGFα, and VEGF the analyte concentration was in 65%, 69%, 39%, 30%, 56%, 46%, and 87% of the samples under the lower limit of quantification, respectively. First, plasma growth factor levels were compared between groups of patients with different breast cancer subtypes. In detail, “luminal A like” (n = 92), “luminal B like” (n = 68), “triple-positive” (n = 7), “HER2-positive” (n = 3),

"triple-negative" (n = 21), and "benign" (n = 21). The criteria to approximate the intrinsic molecular subtypes of breast cancer using histopathological data are summarized in Table 10. No significant difference in concentration for any of the growth factors measured was observed for the different breast cancer subtypes.

Table 10: Surrogate definition of intrinsic molecular subtypes using histopathology data.

subtype	criteria
"luminal A like"	ERα and/or PR positive HER2 negative Ki-67 low (< 14%)
"luminal B like"	ERα and/or PR positive HER2 negative Ki-67 high (< 14%)
triple-positive	ERα and PR positive HER2 positive any Ki-67
HER2-positive	ERα and PR negative HER2 positive any Ki-67
triple-negative	ERα and PR negative HER2 negative any Ki-67

Next the routine clinical information of menopause status, tumor size, lymph node status, and histologic grade was compared with the plasma growth factor levels. The sample set was limited to patients with an overexpression of ERα (n = 168). Patient characteristics of this subset are summarized in Table 18 (appendix). Although growth factor concentrations had a huge variance (%CV 133 – 346) between individual patients, none of the growth factors tested showed a correlation with tumor size, lymph node status, or histologic grade. In case of HGF, higher levels were observed in the group of postmenopausal women compared to the group of premenopausal women (p = 0.001, Wilcoxon rank sum test).

Finally, the expression levels of AREG, HGF, and VEGF measured in tumor lysates and matching plasma samples (n = 74, only patients with ERα overexpression) were compared. However, no correlations between tumor lysate and plasma sample growth factor levels were noticed.

3.7 Pathway activation profiling of ERα-positive breast cancer tumors

Breast cancer is characterised by different intrinsic molecular subtypes representing biologically distinct disease entities. Especially the luminal intrinsic molecular subtype, characterized in general by overexpression of ERα, presents a very heterogeneous subtype. These differences of intrinsic biological features will not only affect the prognosis of the patients but also the response to various therapies. Although the genomic and transcriptomic variability of ERα-positive breast cancer was widely explored, studies on the functional proteomics level are still rare.

Therefore, pathway activation profiles of ERα-positive tumor samples (n = 109) were generated using RPPA. These pathway activation profiles covered components of signalling pathways (Figure 3) known to be implicated in cancer. The primary antibodies (n = 90) used for this analysis are summarized in Table 1 and further information on the target proteins and posttranslational modifications are summarized in Table 11.

Unsupervised two-way hierarchical cluster analysis based on Euclidean distance combined with complete linkage rule revealed four main clusters of tumor samples based on their target protein expression similarity (Figure 35). According to the color illustration in Figure 35, the clusters were named RPPA_orange, RPPA_blue, RPPA_green, and RPPA_purple. The RPPA_blue cluster (30 tumor samples) was characterized by high expression of almost all measured proteins and phosphoproteins whereas for the RPPA_green cluster (20 tumor samples) the expression levels were observed to be rather low. In case of the RPPA_orange cluster (19 tumor samples) and the RPPA_purple cluster (40 tumor samples), one half of the proteins and phosphoproteins was expressed at elevated levels whereas the other half was expressed at low levels, however vice versa.

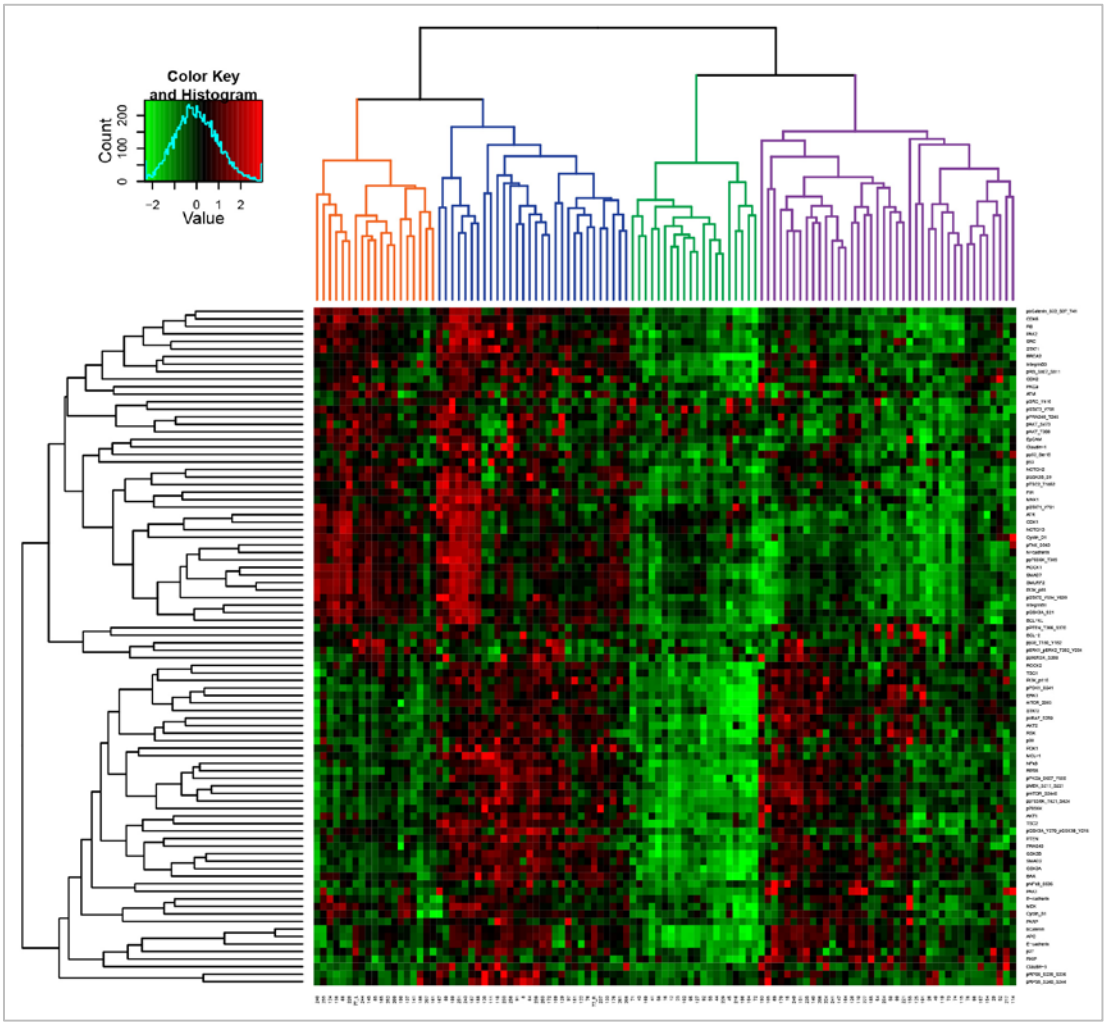


Figure 35: Hierarchical cluster analysis of ERα-positive tumor samples. Heat map illustrating an unsupervised two-way hierarchical cluster analysis based on Euclidean distance combined with complete linkage rule. The cluster analysis revealed four different main groups according to the expression pattern of signaling proteins (n = 90). The four clusters are highlighted in orange, blue, green, and purple. Color code of the heatmap: green = low expression, red = high expression.

The four different groups derived by the cluster analysis were next compared with clinical parameters (Figure 36). No significant relation with respect to menopause status, tumor size and lymph node status was observed for the four clusters (Chi-square test, p = 0.22, p = 0.274, p = 0.344, respectively). In case of histologic grade, significant differences in the distribution between the four cluster (Chi-square test, p = 0.002) were apparent with enrichment of histologic grade 3 tumor samples in the blue cluster.

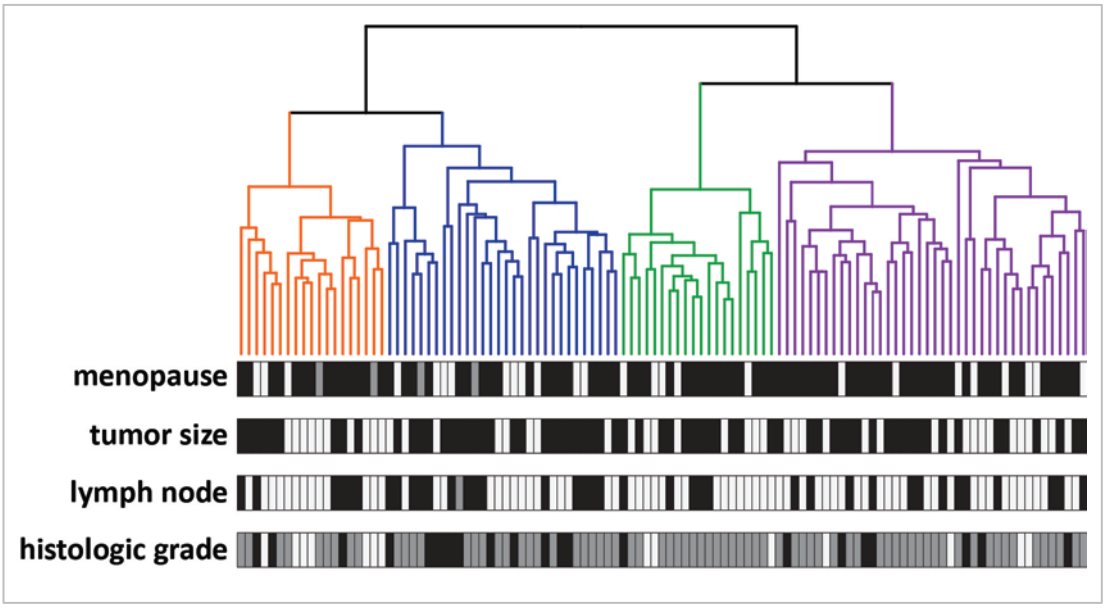


Figure 36: Hierarchical cluster analysis of ERα-positive tumor samples in comparison with clinical parameters. The dendrogram of the cluster analysis (Euclidean distance combined with complete linkage rule) covering 90 signaling proteins is shown in comparison with annotation of the tumor samples regarding menopause status, tumor size, lymph node status, and histologic grade. The four resulting main clusters are highlighted in orange, blue, green, and purple. Menopause status: white = pre-menopausal, grey = peri-menopausal, black = post-menopausal; tumor size: white = T1, black = T2/T3/T4; lymph node status: white = negative, black = positive; histologic grade: white = G1, grey = G2, black = G3.

The accumulation of higher grade tumor samples in the RPPA_blue group was also supported by comparison with the newly defined R2LC based risk classification (Figure 37). The blue cluster was highly enriched with tumor samples classified as R2LC high risk (Chi-square test, p < 0.001). In contrast, the green cluster was characterized by cases

classified as being at low risk for recurrence according to the R2LC based classification. The orange and the purple cluster presented rather a mixture of R2LC high and low risk cases, with tendency of more R2LC high risk cases in the RPPA_purple group and more R2LC low risk cases in the RPPA_orange group.

In addition, a trend to higher VEGF tumor lysates levels could be observed for the RPPA_blue group whereas rather low levels were apparent for the RPPA_green as well as for the RPPA_orange group. No differences with respect to AREG tumor lysate levels were apparent for the four clusters. The RPPA_orange group was characterized by higher HGF tumor lysate expression levels compared to the other three clusters (Figure 37).

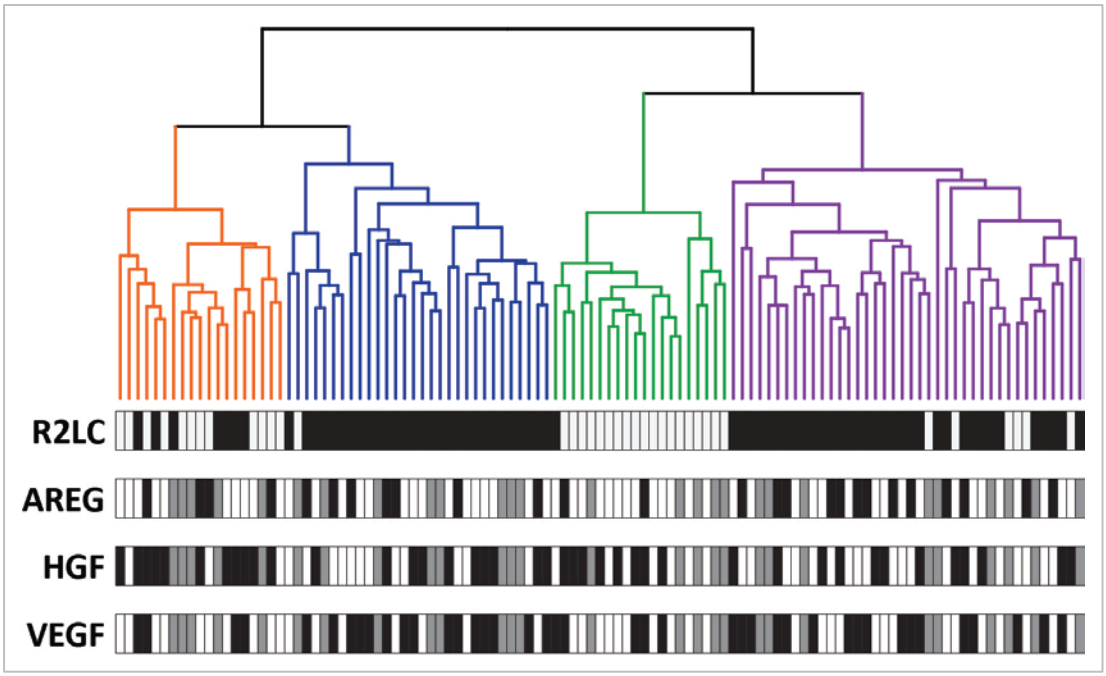


Figure 37: Hierarchical cluster analysis of ERα-positive tumor samples in comparison with experimental parameters. The dendrogram of the cluster analysis (Euclidean distance combined with complete linkage rule) covering 90 signaling proteins is shown in comparison with annotation of R2LC based risk classification as well as AREG, HGF, and VEGF tumor lysate levels. The four resulting main clusters are highlighted in orange, blue, green, and purple. R2LC: white = low risk, black = high risk; AREG, HGF, VEGF: white = low expression, black = high expression, grey = no growth factor expression measured. The median growth factor expression was defined as cut-off (AREG = 964 pg/mg total protein, HGF = 4995 pg/mg total protein, VEGF = 707 pg/mg total protein).

The RPPA_orange and the RPPA_purple group revealed an inverse expression pattern with respect to the different signaling proteins measured. All target proteins, which were significantly differential expressed between these two groups, are listed in Table 14 (appendix). Main differences were observed regarding TGFβ/SMAD signaling, JAK/STAT signaling, cell adhesion, control of G1/S cell cycle transition, PI3K/PTEN signaling and RAS/RAF signaling. For example, SMAD7 and SMURF2 were higher expressed in the RPPA_orange group whereas SMAD2 was lower expressed compared to the RPPA_purple group (Figure 38A). The RPPA_orange group was also characterized by elevated activation of JAK/STAT signaling with pSTAT1 (Y701), pSTAT3 (Y705), and pSTAT5 (Y694/Y699) being higher expressed in this group (Figure 38B).

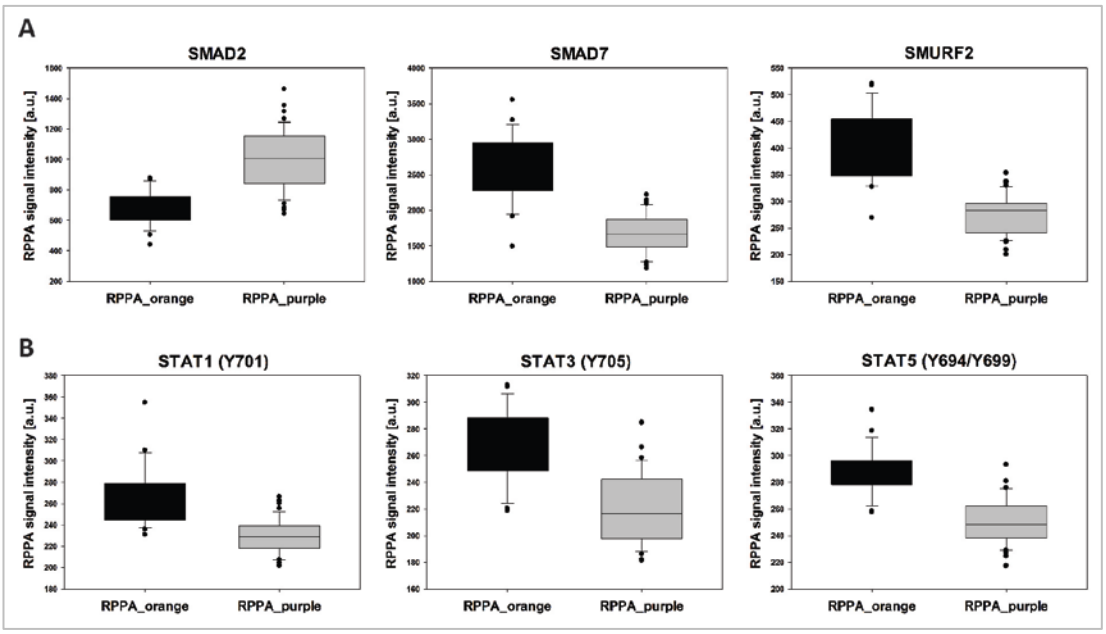


Figure 38: Differences between the RPPA_orange and RPPA purple group in case of TGFβ/SMAD and JAK/STAT signaling. A, The RPPA_purple group was characterized by high expression of SMAD2, but low expression of SMAD7 and SMURF7. B, The RPPA_orange group revealed high expression levels of pSTAT1 (Y701), pSTAT3 (Y705), and pSTAT5 (Y694/Y699) compared to the RPPA_purple group (Wilcoxon rank sum test, $p < 0.001$, respectively).

In addition, differences in PI3K/PTEN and RAS/RAF signaling were striking. Ten proteins/phosphoproteins being part of the two signaling pathways where significantly higher expressed in the RPPA_orange group, including pAKT (T308) and pAKT (S473). In contrast, nine proteins where significantly higher expressed in the RPPA_purple group, for example pcRAF (S259) and RKIP. All 19 differentially expressed proteins are summarized as correlation matrix in Figure 39.

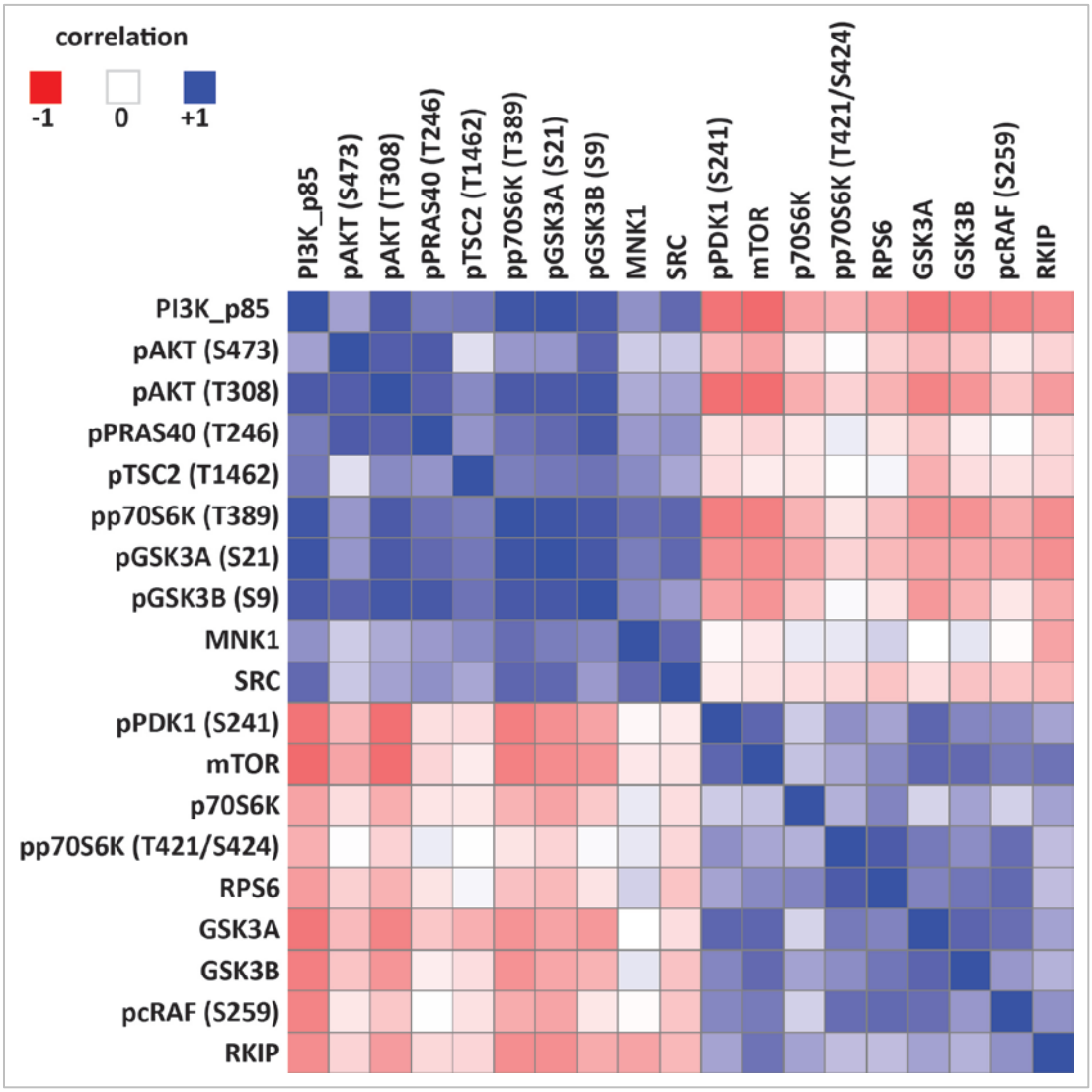


Figure 39: Correlation matrix of PI3K/PTEN and RAS/RAF signaling proteins. Signaling proteins were selected on basis of differential expression between the RPPA_orange and the RPPA_purple group. For this analysis only tumor samples of both groups (n = 59) were considered. Color code: positive correlation = blue, no correlation = white, negative correlation = red. The analysis was done with KNIME 2.5.0.

The RPPA_green group showed the most distinct differences compared to the other three groups. All cases of the RPPA_green group with exception of one case were classified by R2LC as low risk. In addition, the global signaling levels were very low for almost all proteins and phosphoproteins measured. Only a small cluster of phosphoproteins consisting of pSRC (Y416), pSTAT3 (Y705), pPRAS40 (T246), pAKT (S473), and pAKT (T308) showed slightly elevated expression levels (Figure 40). Based on histopathological classification, the analyzed tumor sample set was composed of 93 cases with invasive ductal carcinoma (IDC) and 16 cases with invasive lobular carcinoma (ILC). A high enrichment of ILC cases was observed in the RPPA_green group (Chi-square test, $p < 0.001$) as illustrated in Figure 40. This was in concordance with low levels of E-cadherin as a well known characteristic of ILC. In addition, APC and β-Catenin were also downregulated in ILC cases compared to IDC cases (Figure 41).

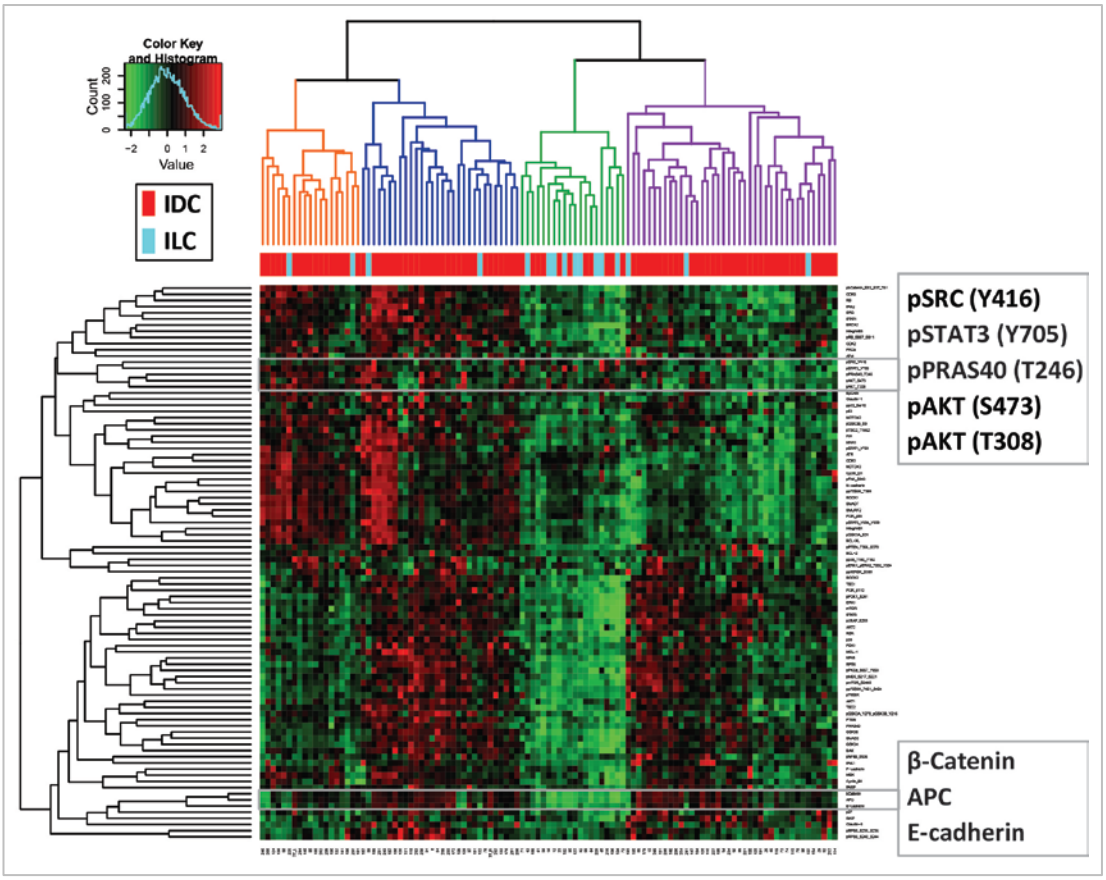


Figure 40: Hierarchical cluster analysis of ERα-positive tumor samples in comparison with breast cancer histology. ILC cases were highly enriched in the RPPA_green group (Chi-square test, $p < 0.001$). Two additional characteristics of the RPPA_green group are highlighted in addition. IDC = invasive ductal carcinoma; ILC = invasive lobular carcinoma.

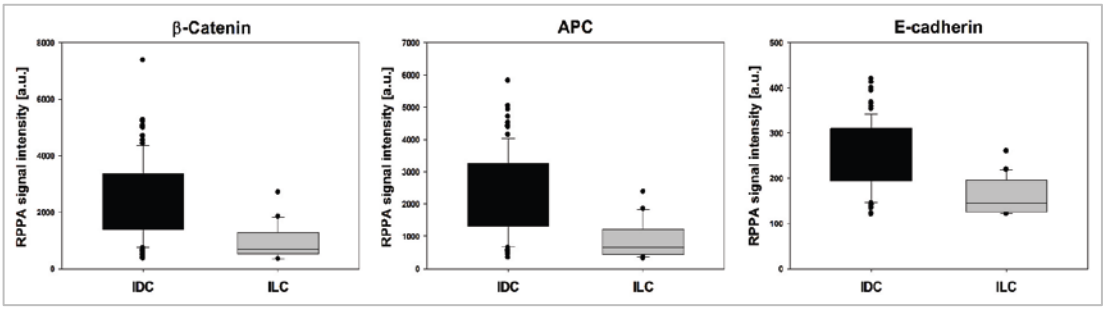


Figure 41: Protein abundance associated with invasive lobular carcinoma. β-Catenin, APC, and E-cadherin were significantly lower expressed in ILC cases compared to IDC cases (Wilcoxon rank sum test, $p < 0.001$, respectively). IDC = invasive ductal carcinoma; ILC = invasive lobular carcinoma.

4. Discussion

4.1 Technical aspects of RPPA based tumor profiling

RPPA have emerged over the last decade to a powerful technology in proteomics. Especially biomarker discovery studies benefit due to improved sample capacity and target quantification options compared to conventional methods. In contrast to blinded approaches such as two-dimensional gel electrophoresis and mass spectrometry, RPPA are considered a guided method, which requires presumptions regarding target proteins of interest. In line with other immunoassay methods, the quality of RPPA derived results is strongly dependent on the specificity and sensitivity of the antibodies used.

A common issue for all biomarker discovery studies based on tumor specimens is the sample quality. This includes dealing with intra-tumor heterogeneity aspects as well as sample stability influenced by ex vivo delay time in sample processing and subsequent preservation. How to approach and handle intra-tumor heterogeneity is conversely discussed in the RPPA community. On the one hand it is argued that performance of laser capture microdissection (LCM) is necessary in order to obtain pure samples containing only tumor cells thus avoiding the influence of other confounding cell populations on the results (Wulfkuhle et al. 2008). On the other hand concerns are raised regarding the quality of clinical material obtained by the labor-intensive LCM procedure (Gonzalez-Angulo et al. 2011). In addition, the use of non-microdissected material allows capturing the information contained in tumor cells as well as in the tumor microenvironment (Hennessy et al. 2010). For the RPPA based tumor profiling performed as part of this thesis, the second approach using non-microdissected samples was chosen. As prerequisite, all tumor specimens were reviewed by histopathology to contain more than 70% tumor cells. Depending on the size of the samples, they were further processed in two different ways (either cut with a scalpel or with a cryomicrotome) to obtain aliquots of each individual sample. These procedures allowed accounting for variations between different regions of the sample and for optimal usage of the finite tumor sample amount enabling various downstream processing protocols. The tumor specimens did not cluster based on the sample processing protocol proving that no bias was introduced by the application of two different approaches. A

future perspective to deal with variations introduced by intra-tumor heterogeneity could be the identification and use of a sophisticated marker panel able to quantify “on slide” the percentage of different cell types like tumor cells, infiltrating immune cells, fibroblasts, or endothelial cells.

The second important sample quality issue is preservation of tissue stability after excision of the tumor specimen. As RPPA based tumor profiling aims at providing a functional proteomic fingerprint of the tumor, proteases, kinases, and phosphatases have to be inhibited effectively and immediately. Studies of Espina et al. and Hennessy et al. for example highlight the importance of tissue preservation in the first hours after sample collection (Espina et al. 2008, Hennessy et al. 2010). In addition, tissue handling variability like time to snap-freezing has to be minimized. So far technologies for the assessment of protein and phosphoprotein preservation quality are lacking. However, with increasing knowledge of characteristic tumor fingerprints it is likely that marker combinations will be identified, which are able to give hints about sample quality. One observation for example across a large number of tumor specimens is that high levels of HER2 are most likely correlated with high levels of pHER2 (Y1248). This leads to the assumption, that a sample with high level of HER2 but no detectable level of pHER2 (Y1248) should be analyzed with caution as this phenomena indicates phosphoprotein instability (personal communication: Gordon B. Mills, University of Texas MD Anderson Cancer Center). According to this example, phosphoprotein quality of the breast cancer sample collection provided for this thesis was fine. A high correlation of HER2 and pHER2 (Y1248) expression levels was observed across the whole sample set.

Another technical aspect of RPPA based tumor profiling is the reproducibility of target protein expression measurements. This issue was addressed by assessment of five different proteins (ERα, PR, HER2, EGFR, and Ki-67) on slides printed and stained at different days. The one to one comparison revealed an excellent correlation for all five tested proteins. In addition, no systematic difference was introduced by the two individual printing runs of the identical tumor sample set. The respective proteins measured twice clustered together across all tumor samples and target proteins demonstrating the robustness of the RPPA approach.

4.2 RPPA based tumor profiling identifies known characteristics of breast cancer

Few biomarkers are assessed in the clinical routine to guide therapy decision for early breast cancer. This panel of biomarkers includes ERα, PR, and HER2. The expression levels of these biomarkers are measured by IHC and different scoring systems are used to classify patients as being either positive or negative for the respective biomarker (Hammond et al. 2010, Wolff et al. 2007). A comparison between the routine IHC classification and RPPA derived target protein expression levels of matching patient samples revealed an excellent positive correlation between these two methods for ERα and HER2. This is in line with studies of Hennessy et al. and Wulfkuhle et al. reporting a positive correlation between IHC and RPPA derived data in several independent breast cancer sample sets. In addition, these studies highlight, as also observed in the data presented here, the improved dynamic range obtained with RPPA, which could help to refine patient classification (Hennessy et al. 2010, Wulfkuhle et al. 2012). In case of PR, a positive correlation between the results derived with IHC and RPPA could be observed, however concordance was suboptimal for samples with low PR expression indicating sensitivity issues.

It is widely recognized that triple-negative breast cancer (TNBC) is a distinct disease entity with subtype specific characteristics (Metzger-Filho et al. 2012). Several proteins were identified by RPPA based tumor profiling to be highly characteristic for the group of TBNC patients. For example, protein abundance of EGFR, MET, Vimentin, P-cadherin, Ki-67, and CDK6 was significantly higher in TNBC tumor samples whereas GATA3, STARD10, pCytokeratin 8 (S23), Cytokeratin 18 as well as ERα and PR were significantly lower expressed. An association with TNBC was already reported for all of the aforementioned proteins. In general, TNBC tumor samples are characterized by a higher proliferation rate compared to other subtypes (Elsawaf and Sinn 2011). This was confirmed by RPPA revealing higher expression levels in the TNBC group for the cell proliferation marker Ki-67 as well as for CDK6 an important regulator of cell cycle progression (Gerdes et al. 1984, Meyerson and Harlow 1994). A negative correlation of EGFR and ERα expression was first shown in 1987 by Sainsbury et al. and EGFR overexpression was linked to poor prognosis (Sainsbury et al. 1987). In addition, P-cadherin and Vimentin levels inversely correlate with ERα expression and are associated with poor prognosis (Sousa et al. 2010, Peralta Soler et al. 1999).

Recently, high protein levels of MET were linked to the TNBC/basel-like subtype (Gastaldi et al. 2010). However another study by Raghav et al. could not identify a significant difference of MET expression between different breast cancer subtypes, although high levels of MET were correlated with worse prognosis (Raghav et al. 2012). GATA3 as well as Cytokeratin-8 and Cytokeratin-18 are established markers for the ERα-positive/luminal subtype (Hoch et al. 1999, Abd El-Rehim et al. 2004) and their expression was confirmed by RPPA as down-regulated in the TNBC group. Moreover, a recent study identified STARD10 (StAR-related lipid transfer START domain containing 10) as target of miR-661 playing an important role in regulation of epithelial to mesenchymal transition (EMT). As part of this study it was shown that loss of STARD10 gene expression was negatively correlated with the EMT-related basal like subtype, in line with the RPPA derived finding (Vetter et al. 2010).

In summary, known characteristics of breast cancer subtypes were reproduced using RPPA based tumor profiling and the functional proteomic fingerprints of the tumor samples provide an excellent basis to discover new breast cancer relevant aspects.

4.3 Risk classification of hormone receptor-positive breast cancer

Treatment of breast cancer patients with similar clinicopathological features can result in different outcomes regarding disease progression and survival. Over the last few years, gene expression profiling studies have improved the understanding of the molecular mechanisms associated with these very heterogeneous clinical outcomes (Perou et al. 2000, Sorlie et al. 2003, Sotiriou and Pusztai 2009). The luminal intrinsic molecular subtype, characterized in most cases by overexpression of hormone receptors, can be further divided into luminal A and luminal B associated with better and worse prognosis, respectively. This classification is crucial for therapy decision as patients of the luminal B subtype with higher risk for recurrence should be treated with chemo-endocrine therapy whereas patients of the luminal A subtype being at lower risk could be spared chemotherapy and its adverse side effects (Bedard and Cardoso 2011, Coates et al. 2012). Hence, to avoid over- or under-treatment of hormone receptor-positive patients, biomarkers are required which allow a precise definition of low and high risk breast cancer.

To differentiate between low and high risk tumors, the level of cell proliferation has emerged as a common theme, mainly supported by gene expression profiling data (Wirapati et al. 2008). This is in line with information provided by histologic grade, which is besides age, tumor size, and lymph node status a well established independent prognostic factor, combining information on tumor proliferation and differentiation (Elston and Ellis 1991). For patients whose tumors were characterized as histologic G1 or histologic G3, prognostic information is univocal, with a good prognosis for G1 and a poor prognosis for G3 patients. However, a considerable percentage of patients are classified as histologic G2 and in this instance histologic grading provides no helpful information for treatment decisions. Thus, the first main objective of this thesis was the identification of a robust protein biomarker signature able to facilitate risk classification of hormone receptor-positive breast cancer patients, especially of those diagnosed with histologic G2.

The biomarker selection workflow presented in this thesis was based on the idea of using quantitative protein expression data of tumor samples classified as histologic G1 and histologic G3, which served as surrogates for the low and high risk group, to identify those proteins being differentially expressed between the two extreme groups. Bioinformatics offers many different methods to solve two-group classification problems in high-

throughput data sets. However, no approach clearly outperforms any other algorithm at once in terms of prediction accuracy, feature selection stability, and biological relevance (Haury et al. 2011, Cun and Frohlich 2012). Therefore, a new biomarker selection workflow, named *bootfs*, was developed, which combines a bootstrap approach, to simulate patient variability and enhance biomarker selection performance, with three different classification methods, namely SCAD-SVM, RF-Boruta, and PAM (Zhang et al. 2006, Kursa and Rudnicki 2010, Tibshirani et al. 2002). Only those target proteins selected by all three classification algorithms at once in a particular bootstrap data set entered the final biomarker ranking which reflected the selection frequency of each individual protein across all bootstrap data sets. With this approach, it seems to be more likely to identify not only the perfect biomarkers signature for a particular sample set used for the selection process, but to identify a signature consisting of robust biomarkers.

Ki-67, TOP2A, and PCNA were listed consistently by the *bootfs* feature selection workflow among the top 10 hits of proteins being important for the differentiation of low and high risk tumors. All three proteins are well-known cell proliferation markers and confirm cell proliferation rate as the most relevant discriminator between low and high risk tumors (Gerdes et al. 1984, Boege et al. 1995, Takasaki et al. 1981). Cell proliferation rate presents also the common driving force behind the prognostic information provided by several published gene expression signatures (Sotiriou and Pusztai 2009), for example the 21-gene signature known as Oncotype Dx® (Sparano and Paik 2008) or the 70-gene signature known as MammaPrint® (Cardoso et al. 2008). However, the three top hits to classify low and high risk tumors were identified as caveolin-1, NDKA, and RPS6 by the *bootfs* selection workflow. Caveolin-1 revealed a higher expression in histologic G1 tumors, whereas NDKA and RPS6 were more strongly expressed in histologic G3 tumors.

All three top hit proteins play a role in diverse biological processes. NDKA, for example, catalyzes the transphosphorylation of γ-phosphates from deoxynucleoside triphosphates to deoxynucleoside diphosphates to supply cells with nucleotides other than ATP (Agarwal et al. 1978). Besides cell proliferation, NDKA is involved in cell differentiation, chromosomal stability, and signal transduction (Cipollini et al. 1997, Lombardi et al. 2000, Conery et al. 2010, Otero 2000). Although NDKA (or NM23-H1) was initially identified by Steeg et al. in 1988 as a gene being downregulated in murine melanoma cell lines with high metastatic potential (Steeg et al. 1988), contradicting results have since then been reported for this

gene in other tumor entities. For example, high levels of NDKA expression were linked with aggressive types of prostate cancer, neuroblastoma, and pancreatic cancer (Igawa et al. 1994, Garcia et al. 2012, Takadate et al. 2012). The results presented in this thesis, strengthen the hypothesis that NDKA could also be a valuable marker for the identification of high risk breast cancer patients. In detail, NDKA was found highly expressed in histologic G3 tumors as identified by RPPA and confirmed by Western blot as well as by IHC of matched tumor samples (personal communication: Sebastian Aulmann, University of Heidelberg). In addition, protein and mRNA expression of NDKA was highly correlated. The analysis of the breast cancer gene expression data set provided by Curtis et al. confirmed a positive correlation of NDKA expression levels and high tumor grading. Along with several other ribosomal proteins, RPS6 is part of the ribosomal 40S subunit controlling protein synthesis rate and cell size during cell division and differentiation (Meyuhas 2008). RPPA based tumor profiling identified RPS6 as being highly expressed in histologic G3 tumor samples and this finding was confirmed by Western blot. However, RPS6 protein expression was not correlated with mRNA expression, in line with a previous report (Hennessy et al. 2010) indicating regulation of RPS6 at the posttranscriptional level. In contrast to NDKA and RPS6, caveolin-1 was strongly expressed in histologic G1 tumor samples and a positive correlation between protein and mRNA levels was observed. The differential expression of caveolin-1 in low and high grade tumors was confirmed by analysis of the Curtis et al. gene expression profiling data. Caveolin-1 is the main component of caveolae, which are specialized lipid rafts, serving for example as molecular hubs by modulating the activity of signaling pathways. In the context of breast cancer, loss of caveolin-1 in cancer associated fibroblasts results in an activated tumor microenvironment and is linked to poor clinical outcome (Sloan et al. 2004, Witkiewicz et al. 2009, Sotgia et al. 2012). This is in line with IHC data obtained for matching tumor samples showing that caveolin-1 expression was mainly detected in the tumor microenvironment and that loss of caveolin-1 expression was observed for higher grade tumor samples (personal communication: Sebastian Aulmann, University of Heidelberg).

In contrast to various breast cancer gene expression profiling studies (Ringner et al. 2011), data of only one breast cancer RPPA study is publicly available so far (Koboldt et al. 2012). Protein expression levels of caveolin-1 and RPS6, but not NDKA, were measured in this study across 403 breast cancer samples comprising all intrinsic molecular subtypes. High caveolin-1 expression levels were associated with the luminal A molecular subtype,

whereas RPS6 was highly expressed in samples of the luminal B molecular subtype, in line with the data presented in this thesis. Interestingly, this "The Cancer Genome Atlas Network" study describes a novel protein expression defined breast cancer subgroup characterized mainly by stromal/microenvironmental components produced for example by cancer associated fibroblasts. This subgroup termed "reactive I" consisted primarily of a subset of luminal A tumors as defined by mRNA profiling and was characterized by high expression of caveolin-1. In addition, the authors note that no marked difference in percentage of tumor cell content of the "reactive I" subgroup compared to tumor samples of other subtypes was observed as assessed by SNP array analysis or pathological examination (Koboldt et al. 2012).

Caveolin-1, NDKA, and RPS6 followed by Ki-67 were the most important proteins, as indicated by the *bootfs* selection workflow ranking, to discriminate between histologic G1 and G3 tumor samples. Protein expression levels of those four proteins were subsequently assessed in histologic G2 tumor samples and compared to the expression levels obtained for histologic G1 and G3 tumors. This comparison revealed that histologic G2 tumors covered the full expression range of the four selected proteins. In addition, two-way hierarchical cluster analysis resulted not in an intermediate or separate cluster for histologic G2 tumors. This finding lead to the assumption that histologic G2 patients with high level expression of NDKA, RPS6 or Ki-67 as well as low level expression of caveolin-1 are at high risk for relapse as their protein biomarker profile is highly similar to that of histologic G3 patients. A similar phenomenon was already reported on gene expression level by two different research groups (Ma et al. 2003, Sotiriou et al. 2006). Especially Sotiriou and colleagues could impressively show that a gene expression signature consisting of 97 genes termed *genomic grade index* (GGI) was able to reclassify histologic G2 tumors into groups of either low or high risk of recurrence (Metzger Filho et al. 2011).

In order to assign single histologic G2 samples as being either at low or high risk according to the identified surrogate expression profile consisting of caveolin-1, NDKA, RPS6, and Ki-67, a risk classification score named R2LC was developed. This R2LC score is individually calculated for each sample by weighted combination of the respective protein expression levels. The performance of R2LC to classify the extreme groups of histologic G1 and histologic G3 samples into low and high risk was tested by ROC-analysis (5-fold cross validation with 10 repeats) and resulted in a satisfactory AUC of 0.987. However, it has to

be noted that this performance test was done using the identical sample set as used for identification of the biomarker panel. A true validation of the R2LC classification performance requires an independent sample set. In addition, an independent sample set with associated disease free survival information will be necessary to evaluate the performance of R2LC for correct reclassification of histologic G2 tumor samples into low and high risk groups of cancer recurrence.

In the meantime, the R2LC based risk classification of the tumor sample set used in this thesis was compared to the associated standard clinical parameters. A significant proportion of lymph node negative patients were classified as R2LC low risk, whereas no correlation with age or tumor size could be observed. To complement and improve standard prognostic factors, several alternative methods based on gene expression analysis have been developed by various research groups over the recent years and some of them are currently tested in clinical trials (Cardoso et al. 2008, Sparano and Paik 2008, Parker et al. 2009, Liedtke et al. 2009, Dubsky et al. 2013). The *genefu* R package (Haibe-Kains et al. 2011) enables the easy application of some of these published gene expression signatures on in-house generated data sets. However it has to be kept in mind that transferring gene expression signature information from one experimental set-up to another experimental set-up is not trivial. Confounding factors include differences in sample preparation protocols, gene expression profiling methods as well as sample set size and composition. Nevertheless, GGI and PAM50 classification for the subset of samples with available mRNA expression profiling data was computed and compared to the RPPA derived R2LC risk classification. In general a good concordance was observed between R2LC based risk classification and the other two gene expression based classification methods. However, a trend for R2LC classifying more patients as high risk compared to the other two methods was obvious.

Although gene expression signatures could provide valuable prognostic information for therapy decision making, it is not always feasible to obtain this information on a routine basis due to high costs and required technical equipment. Therefore the assessment of Ki-67 expression using IHC was recommended by the *St Gallen international expert consensus on the primary therapy of early breast cancer 2011* as convenient approximation for risk classification of hormone receptor-positive breast cancer (Goldhirsch et al. 2011). As cut-off criteria 14% nuclear Ki-67 staining was proposed as derived by comparison with

PAM50 classification as gold standard (Cheang et al. 2009). So far, established guidelines exist only for IHC based determination of estrogen and progesterone receptor status as well as HER2 receptor status (Hammond et al. 2010, Wolff et al. 2007). The development of a guideline for the assessment of Ki-67 by IHC is in progress, but seems to present substantial challenges (Viale et al. 2008, Dowsett et al. 2011). For example, a recent study points out that IHC Ki-67 quantification suffers from high inter- and intra-observer variability especially in moderately differentiated breast carcinomas (Varga et al. 2012) underlining that options to translate newly identified biomarkers into IHC assays for the clinic routine are currently limited due to technical limitations (Leong 2004). Moreover, if a candidate biomarker reveals a complex staining pattern it is not suitable for IHC. RPPA in contrast provide an unbiased quantitative readout with high dynamic range for the biomarkers of interest and can therefore overcome limitations as faced by IHC. In addition, only small amounts of tumor material are required for the measurement of several target proteins in parallel, thus allowing for the measurement of multiplex biomarker signatures. For these reasons, the four biomarker signature including Ki-67 as identified in this thesis presents, in combination with RPPA, an attractive alternative to traditional IHC due to the accurate and quantitative readout.

Although snap-frozen tumor samples were used for the identification of the biomarker signature, it seems likely that the sample preparation protocol could be adapted to the use of formalin-fixed paraffin-embedded (FFPE) tumor tissue, which is commonly used as standard for tissue fixation enabling storage at room temperature. Recently, different FFPE protein extraction protocols were introduced showing the compatibility with downstream RPPA analysis for the majority of the tested proteins so far (Berg et al. 2010, Guo et al. 2012). Another option would be the use of an alternative fixation method as proposed by Mueller et al., which was developed for the multipurpose of tissue morphology preservation, nucleic acid stabilization as well as protein and phosphoprotein stabilization (Mueller et al. 2011).

It is obvious, that the method for the assessment of the four biomarker panel consisting of caveolin-1, NDKA, RPS6, and Ki-67, has to be adapted to be compatible with the analysis of single patient samples as required in the day-to-day routine. Traditionally, RPPA are designed for the relative quantification of target proteins across large sample set. However, identification and validation of appropriate reference material, for example dilution series

of cell line pools or recombinant proteins with known concentration, could enable the development of a true “single sample“ RPPA. In this context, the coefficients of the R2LC risk classification score as presented in this initial study have to be adapted to reflect the optimal cut-off criteria.

4.4 Tumor lysate and plasma levels of growth factors in breast cancer

Deregulated growth factor expression plays an important role in development and progression of breast cancer (Hynes and Watson 2010). Especially autocrine production of EGFR, MET, and VEGFR family member ligands by tumor cells had been frequently reported, besides paracrine contribution from tumor stroma (Kenny and Bissell 2007, Willmarth et al. 2009, Wilson et al. 2011, Mueller et al. 2012, Perrot-Applanat and Di Benedetto 2012). Within the scope of this thesis, a MIA was developed to evaluate the potential of a selected set of growth factors to serve as biomarkers and to elucidate the expression patterns of these growth factors in the context of breast cancer. The selected set of growth factors comprised six ligands of EGFR family members, in detail AREG, BTC, EGF, HBEGF, HRG, and TGFα. In addition, HGF and VEGF, the ligands of MET and VEGFR, can be quantified with this 8-plex MIA.

A multiplex immunoassay approach was chosen for the measurement of these growth factors in order to maximize the throughput using a limited amount of precious breast cancer tumor lysate and blood plasma. The principles of the MIA procedure for multiplex quantification of analytes had been already established at the Division of Molecular Genome Analysis (DKFZ) (Korf et al. 2008a), however only for the assessment of a few signaling proteins in cell culture samples. Hence, a completely new MIA workflow was developed and adapted to the measurement of eight different growth factors in tumor lysate and plasma samples. Primarily the use of plasma samples, presenting a highly complex biological matrix, made changes in the established workflow necessary. The blocking procedure had to be improved as well as the diluent for the recombinant standard protein mix in order to assemble a matrix comparable to human plasma. In this context, it was observed that matrix effects should not be underestimated as a comparison of two different diluents (1% BSA in PBS vs. 20% FBS in PBS) revealed marked differences in signal intensities obtained for the same concentration of recombinant standard protein. This fact has to be kept in mind if results are compared across different assay platforms and presents most likely one reason why comparison between different commercially available multiplex ELISA kits have yielded rather inconclusive results so far (Lash et al. 2006, Toedter et al. 2008).

Another important prerequisite for reliable MIA experiments was the establishment of a standard operating procedure covering the whole process from printing of capture antibody slides to calculation of analyte concentrations as well as several quality control steps. Particularly if numerous samples are analyzed over an extended time frame, assay conditions have to be kept constant in order to obtain comparable results. Simultaneously to the development of the 8-plex MIA SOP, a software program, named QuantProReloaded, was designed to enable a semi-automated analysis of data from MIA experiments (Jocker et al. 2010). In contrast to other multiplex ELISA analysis software, like the commercially available SearchLight®Plus or the open source available ProMAT (Backen et al. 2009, White et al. 2006), QuantProReloaded was tailored to the specific MIA layout and combined with a user-friendly interface.

Out of the eight growth factors measured in tumor lysates of hormone receptor-positive breast cancer patients, three (AREG, HGF, VEGF) were detectable in over 10% of the samples. Subsequent comparison with associated clinical data revealed no significant correlation with menopause status, tumor size, and lymph node status for any of these three growth factors. In contrast to AREG, HGF and VEGF revealed a correlation with histologic grade. High levels of VEGF were strongly associated with high grade tumor whereas high levels of HGF showed a weak association with low grade tumors. By comparison with the newly defined R2LC risk groups only the association of high risk tumors with high VEGF levels was confirmed. This finding is in line with reports of several other research groups about tumor lysate VEGF levels and their association with poor disease free and poor overall survival (Linderholm et al. 1998, Desruisseau et al. 2004, Konecny et al. 2004).

Blood is a preferred source for biomarkers as it reflects the various physiological and pathological states of an individual. In addition, blood can be sampled in a less invasive way, which makes it also compatible with outpatient settings. Therefore, the eight growth factors were also measured in blood samples obtained of breast cancer patients and of patients with benign tumor at primary diagnosis in order to evaluate their potential as biomarkers. The choice of using either blood serum or plasma as source for biomarkers strongly depends on the analytes of interest. It is known for example that VEGF cannot be reliable measured in serum samples as VEGF is bound to thrombocytes and is released into the serum during the clotting process. Therefore, VEGF levels of serum samples rather

reflect the clotting process than an actual pathologic state (Webb et al. 1998, Banks et al. 1998). Consequently, plasma samples were used in this study. Depending on the particular analyte, 13% - 70% of the samples revealed an analyte concentration higher than the respective lower limit of quantification. First, growth factor levels were compared between different breast cancer subtypes ("luminal A like", "luminal B like", triple-positive, HER2-positive, triple-negative) and benign tumors. However, no differential expression of the eight different growth factors was observed among breast cancer subtypes and benign disease.

Furthermore, the routine clinical information of hormone receptor-positive breast cancer patients was compared with the respective plasma growth factor levels. None of the eight different growth factors revealed an association with tumor size, lymph node status, or histologic grade. In case of HGF, higher concentrations were observed in the group of postmenopausal women compared to the group of premenopausal women. The association of age and HGF plasma concentration was already reported by Yamamoto et al., who described in addition an independent correlation of high HGF levels and atherosclerotic disease (Yamamoto et al. 2001). Although blood is a preferred source for biomarkers due to the ease of sampling and the possibility of disease monitoring over time, the probability of confounding factors by secondary diseases is much higher compared to the assessment of tumor tissue. Accordingly, functional proteomic analysis is only feasible using tumor tissue and analysis of plasma should be considered rather as dysfunctional proteomics. Therefore it is not surprising that no correlation between expression levels of AREG, HGF, and VEGF measured in tumor lysates and corresponding plasma samples could be observed.

Nevertheless, the assessment of growth factor concentrations in plasma is still promising for the identification of biomarkers used for prediction of therapy response or resistance, especially in the field of targeted breast cancer therapeutics. This hypothesis is supported by the huge variance of growth factor concentrations observed in this study between individual patients. Indeed, high plasma VEGF levels have been reported to be a predictive marker for the therapy with bevacizumab in HER2-negative metastatic breast cancer, whereas no correlation with clinical characteristics like disease-free survival or metastatic burden were observed (Miles et al. 2013).

4.5 Pathway activation patterns towards definition of pathway-specific and patient-tailored therapy

Breast cancer development and progression is due to inherited and acquired genomic aberrations. The underlying genomic and transcriptomic variability of breast cancer tumors was recently highlighted by Curtis et al. in a comprehensive analysis of over 2,000 primary breast cancer specimens. Using integrative clustering of copy number and gene expression data they refined the established intrinsic molecular subtypes. The majority of the newly defined subgroups were exclusively associated with luminal breast cancer emphasizing the large heterogeneity of this subtype (Curtis et al. 2012). However on the functional level, breast cancer is not only a genomic but mainly a proteomic disease. Since genomic and transcriptomic profiling alone cannot sufficiently predict protein pathway activation, it is important to explore and define the heterogeneity of luminal breast cancer on the signaling pathway activation level. Especially as protein signaling pathways present the direct targets of new classes of therapeutics like tyrosine kinase inhibitors and monoclonal antibodies. In this context, pathway activation profiling will not only be important to identify new drug targets or drug combinations, but also to discover predictive biomarkers defining subgroups of patients most likely to respond to a already available targeted therapeutics.

The pathway activation profiling of hormone receptor-positive breast cancer, presented in this thesis, covered the major signaling pathways known to be involved in cancer. Unsupervised two-way hierarchical cluster analysis revealed four main clusters of tumor samples based on their signaling protein expression similarity.

The RPPA_blue cluster was characterized by high expression of almost all proteins analyzed and a high enrichment of histologic G3 tumor samples. This high degree of dedifferentiation was also supported by the newly defined R2LC based risk classification. All tumors, except of one in this cluster, were classified as R2LC high risk. In addition, VEGF tumor lysate levels were significantly higher in samples of this cluster compared to the others, hinting also to an association of high cancer recurrence risk in this subgroup (Linderholm et al. 1998, Desruisseau et al. 2004, Konecny et al. 2004). In summary, tumors of the RPPA_blue cluster combined features of highly aggressive tumors and present most likely the group with the worst prognosis among hormone receptor-positive breast cancer. Since many signaling

pathways were highly active in this group it is unlikely that targeting only one pathway will be sufficient. It will be important, especially for this subgroup, to identify drug combinations to efficiently prevent tumor recurrence or progression.

The RPPA_orange and the RPPA_purple cluster were characterized by an inverse expression pattern with respect to several different signaling proteins analyzed. Main differences were observed for example regarding TGFβ/SMAD signaling, JAK/STAT signaling, and PI3K/PTEN signaling. In case of TGFβ/SMAD signaling, SMAD7 and SMURF2 were lower expressed in the RPPA_purple cluster whereas SMAD2 was higher expressed compared to the RPPA_orange cluster. Coexpression of inhibitory SMAD7 and the ubiquitin ligase SMURF2 leads to degradation of SMAD2 as well as TGFβ receptor and consequently to inhibition of TGFβ/SMAD signaling (Hayashi et al. 1997, Kavsak et al. 2000, Lin et al. 2000). In this context, the observed differential expression suggests that TGFβ/SMAD signaling was active in tumors of the RPPA_purple cluster but inhibited in tumors of the RPPA_orange cluster. The RPPA_orange group was further characterized by elevated activation of JAK/STAT signaling with pSTAT3 (Y705) and pSTAT5 (Y694/Y699) higher expressed compared to tumors of the RPPA_purple cluster. High activation of STAT3 in breast cancer has been already reported (Watson and Miller 1995), although STAT3 is known to be involved in initiation of apoptosis during involution of the normal mammary gland (Chapman et al. 1999). The mediation of immune evasion, by production of factors inhibiting multiple proinflammatory cytokines as well as dendritic cell maturation, was proposed as possible contribution of activated STAT3 to tumor progression (Wang et al. 2004). Although STAT3 and STAT5 have been described to have opposing functions in the normal mammary gland, simultaneous high activity of STAT3 and STAT5, as characteristic for the RPPA_orange group, has been observed in a subset of breast cancer tumors and was linked to more differentiated tumors compared to tumors with mainly STAT3 activity (Walker et al. 2009). This is in line with the observation that the percentage of histologic G1 tumors was higher in the RPPA_orange group and was also confirmed by enrichment of R2LC low risk cases in this group. Interestingly, tumors of the RPPA_orange cluster revealed the highest HGF expression compared to all other clusters. HGF as ligand of the MET receptor is known to be mainly involved in promoting invasion and metastasis (Birchmeier et al. 2003). At first, the occurrence of high HGF levels in a subgroup of tumors with rather good differentiation and low proliferation rate seems paradox. However, in contrast to other breast cancer subtypes, luminal breast cancer is associated with late recurrence (> 5 years after

diagnosis) and so far little is known about the underlying biological mechanisms for this phenomenon (Saphner et al. 1996, Dignam et al. 2009). Treatment of hormone receptor-positive breast cancer with adjuvant chemotherapy in addition to endocrine therapy is beneficial for a subset of patients, but seems to have little impact on the risk of late recurrence as most of the benefit occurs during the first years after diagnosis (Early Breast Cancer Trialists' Collaborative Group 2005, Lin and Winer 2008). In this context, chemotherapy responsiveness has been shown to be linked to high cell proliferation and reflects the inability of current gene expression signatures, which mainly rely on markers for cell proliferation, to predict recurrence beyond 5 years of diagnosis (Esserman et al. 2011). It can now be speculated that tumors assembling characteristics of the RPPA_orange group, with good differentiation, low proliferation, but high levels of HGF, represent exactly those cases with late recurrence risk. As already discussed, those tumors are not likely to respond well to chemotherapy and therefore other therapy options are needed. Since the RPPA_orange group was characterized by high activation of PI3K/PTEN signaling, a treatment alternative could be the use of mTOR inihibitors like everolimus, which showed already promising results in clinical trials (Baselga et al. 2009, Baselga et al. 2012a).

The RPPA_green cluster revealed the most distinct differences compared to the other three groups. This cluster was characterized by weak signaling activity across almost all analyzed proteins. Only a small cluster of phosphoproteins, consisting of pSRC (Y416), pSTAT3 (Y705), pPRAS40 (T246), pAKT (S473), and pAKT (T308), showed slightly elevated expression levels. Interestingly, a high enrichment of ILC cases was observed within the RPPA_green cluster. ILC is after IDC the most common histologic type of breast cancer with a prevalence of 5 - 15% and distinct morphological as well as pathological characteristics (Weigelt and Reis-Filho 2009). ILC is characterized by small round cells with scant cytoplasm, which invade the surrounding stroma in a non-destructive manner. Based on this special growth pattern, ILC often fails to form distinct masses complicating early diagnosis by palpation or mammography (Krecke and Gisvold 1993). The best established biomarker to distinguish between IDC and ILC is E-cadherin, with a low expression level of this cell adhesion protein being characteristic for ILC (Berx et al. 1995, Weigelt et al. 2010). The finding of E-cadherin loss in ILC is in line with the RPPA derived data and concordant with the simultaneous observed downregulation of APC and β-catenin for ILC cases and in general for tumors of the RPPA_green cluster (Sarrio et al. 2003, De Leeuw et al. 1997). Moreover, the RPPA derived pathway activation profiling strengthens the distinct biologic characteristics of ILC

or ILC-like cases of the RPPA_green cluster compared to others. ILC is commonly associated with older patient age, larger tumors, increased ERα expression, and a rather low proliferation rate (Arpino et al. 2004, Pestalozzi et al. 2008). Accordingly, it was reported that ILC rarely achieve pathological complete response after neoadjuvant chemotherapy reflecting again the link of low proliferation and chemotherapy unresponsiveness (Cristofanilli et al. 2005, Tubiana-Hulin et al. 2006, Farese and Aebi 2008). This characteristic was supported by the RPPA based risk classification presented in this thesis. All tumors of the RPPA_green cluster except of one were classified as R2LC low risk although this cluster was highly enriched with histologic G2 tumors.

In summary, the RPPA based pathway activation profiling of hormone receptor-positive breast cancer presented in this thesis provides for the first time a comprehensive snapshot of the heterogeneity of this subtype on the proteomic level. The four different subgroups, defined on the basis of differential expression of 90 signaling proteins, showed each unique characteristics, which were also related to established clinicopathologic features. It will be exciting to see whether these subgroups can be reproduced in further independent tumor sample sets. In addition, the selection of the 90 signaling proteins should be refined and extended in the future to cover additional components of signaling pathways. For example, receptor tyrosine kinases or other membrane bound proteins, which present ideal targets of monoclonal antibody therapeutics. In addition, the approach could be also extended to the detection of posttranslational modifications other than phosphorylation. Of course these efforts will be always limited by the availability of highly target specific antibodies. In this context, alternative binders based on non-immunoglobulin based protein scaffolds could provide an attractive replacement (Boersma and Pluckthun 2011, Lofblom et al. 2011). Finally, similar to the efforts in next-generation sequencing to separate the driver from the passenger mutations (Stephens et al. 2012), it will be necessary to define the “driver” signaling hubs on the proteomic level. These “driver” signaling proteins will present the ideal predictive biomarkers as well as the ideal drug targets for the personalized treatment of breast cancer patients.

References

Abd El-Rehim D.M., Pinder S.E., Paish C.E., Bell J., Blamey R.W., Robertson J.F., Nicholson R.I., Ellis I.O. (2004). Expression of luminal and basal cytokeratins in human breast carcinoma. *J Pathol* 203**:** 661-671.

Adams C.W., Allison D.E., Flagella K., Presta L., Clarke J., Dybdal N., McKeever K., Sliwkowski M.X. (2006). Humanization of a recombinant monoclonal antibody to produce a therapeutic HER dimerization inhibitor, pertuzumab. *Cancer immunology, immunotherapy : CII* 55**:** 717-727.

Agarwal R.P., Robison B., Parks R.E., Jr. (1978). Nucleoside diphosphokinase from human erythrocytes. *Methods in enzymology* 51**:** 376-386.

Anders C.K., Hsu D.S., Broadwater G., Acharya C.R., Foekens J.A., Zhang Y., Wang Y., Marcom P.K., Marks J.R., Febbo P.G., Nevins J.R., Potti A., Blackwell K.L. (2008). Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression. *J Clin Oncol* 26**:** 3324-3330.

Arpino G., Bardou V.J., Clark G.M., Elledge R.M. (2004). Infiltrating lobular carcinoma of the breast: tumor characteristics and clinical outcome. *Breast Cancer Res* 6**:** R149-156.

Backen A.C., Cummings J., Mitchell C., Jayson G., Ward T.H., Dive C. (2009). 'Fit-for-purpose' validation of SearchLight multiplex ELISAs of angiogenesis for clinical trial use. *J Immunol Methods* 342**:** 106-114.

Banks R.E., Forbes M.A., Kinsey S.E., Stanley A., Ingham E., Walters C., Selby P.J. (1998). Release of the angiogenic cytokine vascular endothelial growth factor (VEGF) from platelets: significance for VEGF measurements and cancer biology. *Br J Cancer* 77**:** 956-964.

Baselga J., Campone M., Piccart M., Burris H.A., 3rd, Rugo H.S., Sahmoud T., Noguchi S., Gnant M., Pritchard K.I., Lebrun F., Beck J.T., Ito Y., Yardley D., Deleu I., Perez A., Bachelot T., Vittori L., Xu Z., Mukhopadhyay P., Lebwohl D., Hortobagyi G.N. (2012a). Everolimus in postmenopausal hormone-receptor-positive advanced breast cancer. *N Engl J Med* 366**:** 520-529.

Baselga J., Cortes J., Kim S.B., Im S.A., Hegg R., Im Y.H., Roman L., Pedrini J.L., Pienkowski T., Knott A., Clark E., Benyunes M.C., Ross G., Swain S.M. (2012b). Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. *N Engl J Med* 366**:** 109-119.

Baselga J., Perez E.A., Pienkowski T., Bell R. (2006). Adjuvant trastuzumab: a milestone in the treatment of HER-2-positive early breast cancer. *Oncologist* 11 Suppl 1**:** 4-12.

Baselga J., Semiglazov V., van Dam P., Manikhas A., Bellet M., Mayordomo J., Campone M., Kubista E., Greil R., Bianchi G., Steinseifer J., Molloy B., Tokaji E., Gardner H., Phillips P., Stumm M., Lane H.A., Dixon J.M., Jonat W., Rugo H.S. (2009). Phase II randomized study of neoadjuvant everolimus plus letrozole compared with placebo plus letrozole in patients with estrogen receptor-positive breast cancer. *J Clin Oncol* 27**:** 2630-2637.

Bauer K.R., Brown M., Cress R.D., Parise C.A., Caggiano V. (2007). Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. *Cancer* 109**:** 1721-1728.

Beatson G.T. (1896). On the treatment if inoperable cases if carcinoma of the mamma: suggestions for a new method of treatment, with illustrative cases. *The Lancet* 2**:** 162-165.

Bedard P.L., Cardoso F. (2011). Can some patients avoid adjuvant chemotherapy for early-stage breast cancer? *Nat Rev Clin Oncol* 8**:** 272-279.

Benninghoff A., Drenckhahn D. (2004). *Anatomie*, vol. 16. Urban & Fischer Verlag/Elsevier GmbH.

Berg D., Hipp S., Malinowsky K., Bollner C., Becker K.F. (2010). Molecular profiling of signalling pathways in formalin-fixed and paraffin-embedded cancer tissues. *Eur J Cancer* 46**:** 47-55.

Berx G., Cleton-Jansen A.M., Nollet F., de Leeuw W.J., van de Vijver M., Cornelisse C., van Roy F. (1995). E-cadherin is a tumour/invasion suppressor gene mutated in human lobular breast cancers. *EMBO J* 14**:** 6107-6115.

Biomarkers Definitions Working Group (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 69**:** 89-95.

Birchmeier C., Birchmeier W., Gherardi E., Vande Woude G.F. (2003). Met, metastasis, motility and more. *Nat Rev Mol Cell Biol* 4**:** 915-925.

Boege F., Andersen A., Jensen S., Zeidler R., Kreipe H. (1995). Proliferation-associated nuclear antigen Ki-S1 is identical with topoisomerase II alpha. Delineation of a carboxy-terminal epitope with peptide antibodies. *Am J Pathol* 146**:** 1302-1308.

Boersma Y.L., Pluckthun A. (2011). DARPins and other repeat protein scaffolds: advances in engineering and applications. *Current opinion in biotechnology* 22**:** 849-857.

Bolstad B.M., Irizarry R.A., Astrand M., Speed T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19**:** 185-193.

Brase J.C., Johannes M., Schlomm T., Falth M., Haese A., Steuber T., Beissbarth T., Kuner R., Sultmann H. (2011). Circulating miRNAs are correlated with tumor progression in prostate cancer. *Int J Cancer* 128**:** 608-616.

Brisken C., O'Malley B. (2010). Hormone action in the mammary gland. *Cold Spring Harb Perspect Biol* 2**:** a003178.

Burstein H.J., Polyak K., Wong J.S., Lester S.C., Kaelin C.M. (2004). Ductal carcinoma in situ of the breast. *N Engl J Med* 350**:** 1430-1441.

Burstein H.J., Prestrud A.A., Seidenfeld J., Anderson H., Buchholz T.A., Davidson N.E., Gelmon K.E., Giordano S.H., Hudis C.A., Malin J., Mamounas E.P., Rowden D., Solky A.J., Sowers M.R., Stearns V., Winer E.P., Somerfield M.R., Griggs J.J. (2010). American Society of Clinical Oncology clinical practice guideline: update on adjuvant endocrine therapy for women with hormone receptor-positive breast cancer. *J Clin Oncol* 28**:** 3784-3796.

Cardoso F., Van't Veer L., Rutgers E., Loi S., Mook S., Piccart-Gebhart M.J. (2008). Clinical application of the 70-gene profile: the MINDACT trial. *J Clin Oncol* 26**:** 729-735.

Carey L.A., Rugo H.S., Marcom P.K., Mayer E.L., Esteva F.J., Ma C.X., Liu M.C., Storniolo A.M., Rimawi M.F., Forero-Torres A., Wolff A.C., Hobday T.J., Ivanova A., Chiu W.K., Ferraro M., Burrows E., Bernard P.S., Hoadley K.A., Perou C.M., Winer E.P. (2012). TBCRC 001: randomized phase II study of cetuximab in combination with carboplatin in stage IV triple-negative breast cancer. *J Clin Oncol* 30**:** 2615-2623.

Castro F., Dirks W.G., Fahnrich S., Hotz-Wagenblatt A., Pawlita M., Schmitt M. (2013). High-throughput SNP-based authentication of human cell lines. *Int J Cancer* 132**:** 308-314.

Chapman P.B., Hauschild A., Robert C., Haanen J.B., Ascierto P., Larkin J., Dummer R., Garbe C., Testori A., Maio M., Hogg D., Lorigan P., Lebbe C., Jouary T., Schadendorf D., Ribas A., O'Day S.J., Sosman J.A., Kirkwood J.M., Eggermont A.M., Dreno B., Nolop K., Li J., Nelson B., Hou J., Lee R.J., Flaherty K.T., McArthur G.A. (2011). Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med* 364**:** 2507-2516.

Chapman R.S., Lourenco P.C., Tonner E., Flint D.J., Selbert S., Takeda K., Akira S., Clarke A.R., Watson C.J. (1999). Suppression of epithelial apoptosis and delayed mammary gland involution in mice with a conditional knockout of Stat3. *Genes & development* 13**:** 2604-2616.

Cheang M.C., Chia S.K., Voduc D., Gao D., Leung S., Snider J., Watson M., Davies S., Bernard P.S., Parker J.S., Perou C.M., Ellis M.J., Nielsen T.O. (2009). Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *J Natl Cancer Inst* 101**:** 736-750.

Cipollini G., Berti A., Fiore L., Rainaldi G., Basolo F., Merlo G., Bevilacqua G., Caligo M.A. (1997). Down-regulation of the nm23.h1 gene inhibits cell proliferation. *Int J Cancer* 73**:** 297-302.

Clynes R.A., Towers T.L., Presta L.G., Ravetch J.V. (2000). Inhibitory Fc receptors modulate in vivo cytotoxicity against tumor targets. *Nat Med* 6**:** 443-446.

Coates A.S., Colleoni M., Goldhirsch A. (2012). Is adjuvant chemotherapy useful for women with luminal a breast cancer? *J Clin Oncol* 30**:** 1260-1263.

Cole M.P., Jones C.T., Todd I.D. (1971). A new anti-oestrogenic agent in late breast cancer. An early clinical appraisal of ICI46474. *Br J Cancer* 25**:** 270-275.

Conery A.R., Sever S., Harlow E. (2010). Nucleoside diphosphate kinase Nm23-H1 regulates chromosomal stability by activating the GTPase dynamin during cytokinesis. *Proc Natl Acad Sci U S A* 107**:** 15461-15466.

Cristofanilli M., Gonzalez-Angulo A., Sneige N., Kau S.W., Broglio K., Theriault R.L., Valero V., Buzdar A.U., Kuerer H., Buccholz T.A., Hortobagyi G.N. (2005). Invasive lobular carcinoma classic type: response to primary chemotherapy and survival outcomes. *J Clin Oncol* 23**:** 41-48.

Cun Y., Frohlich H.F. (2012). Prognostic gene signatures for patient stratification in breast cancer - accuracy, stability and interpretability of gene selection approaches using prior knowledge on protein-protein interactions. *BMC Bioinformatics* 13**:** 69.

Curtis C., Shah S.P., Chin S.F., Turashvili G., Rueda O.M., Dunning M.J., Speed D., Lynch A.G., Samarajiwa S., Yuan Y., Graf S., Ha G., Haffari G., Bashashati A., Russell R., McKinney S., Langerod A., Green A., Provenzano E., Wishart G., Pinder S., Watson P., Markowetz F., Murphy L., Ellis I., Purushotham A., Borresen-Dale A.L., Brenton J.D., Tavare S., Caldas C., Aparicio S. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486**:** 346-352.

Cuzick J., Sestak I., Baum M., Buzdar A., Howell A., Dowsett M., Forbes J.F. (2010). Effect of anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: 10-year analysis of the ATAC trial. *Lancet Oncol* 11**:** 1135-1141.

Davies C., Godwin J., Gray R., Clarke M., Cutter D., Darby S., McGale P., Pan H.C., Taylor C., Wang Y.C., Dowsett M., Ingle J., Peto R. (2011). Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet* 378**:** 771-784.

de Cupis A., Schettini G., Favoni R.E. (1999). New vs old fashioned oestradiol antagonists in mammary carcinoma: 'in vitro' and 'in vivo' pharmacological approaches. *Pharmacological research : the official journal of the Italian Pharmacological Society* 39**:** 335-344.

De Leeuw W.J., Berx G., Vos C.B., Peterse J.L., Van de Vijver M.J., Litvinov S., Van Roy F., Cornelisse C.J., Cleton-Jansen A.M. (1997). Simultaneous loss of E-cadherin and catenins in invasive lobular breast cancer and lobular carcinoma in situ. *J Pathol* 183**:** 404-411.

Desruisseau S., Palmari J., Giusti C., Romain S., Martin P.M., Berthois Y. (2004). Clinical relevance of amphiregulin and VEGF in primary breast cancers. *Int J Cancer* 111**:** 733-740.

Dhakal H.P., Naume B., Synnestvedt M., Borgen E., Kaaresen R., Schlichting E., Wiedswang G., Bassarova A., Holm R., Giercksky K.E., Nesland J.M. (2012). Expression of vascular endothelial growth factor and vascular endothelial growth factor receptors 1 and 2 in invasive breast carcinoma: prognostic significance and relationship with markers for aggressiveness. *Histopathology* 61**:** 350-364.

Dignam J.J., Dukic V., Anderson S.J., Mamounas E.P., Wickerham D.L., Wolmark N. (2009). Hazard of recurrence and adjuvant treatment effects over time in lymph node-negative breast cancer. *Breast Cancer Res Treat* 116**:** 595-602.

Dowsett M., Jones A., Johnston S.R., Jacobs S., Trunet P., Smith I.E. (1995). In vivo measurement of aromatase inhibition by letrozole (CGS 20267) in postmenopausal patients with breast cancer. *Clin Cancer Res* 1**:** 1511-1515.

Dowsett M., Nielsen T.O., A'Hern R., Bartlett J., Coombes R.C., Cuzick J., Ellis M., Henry N.L., Hugh J.C., Lively T., McShane L., Paik S., Penault-Llorca F., Prudkin L., Regan M., Salter J., Sotiriou C., Smith I.E., Viale G., Zujewski J.A., Hayes D.F. (2011). Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. *J Natl Cancer Inst* 103**:** 1656-1664.

Dubsky P., Filipits M., Jakesz R., Rudas M., Singer C.F., Greil R., Dietze O., Luisser I., Klug E., Sedivy R., Bachner M., Mayr D., Schmidt M., Gehrmann M.C., Petry C., Weber K.E., Kronenwett R., Brase J.C., Gnant M. (2013). EndoPredict improves the prognostic classification derived from common clinical guidelines in ER-positive, HER2-negative early breast cancer. *Ann Oncol* 24**:** 640-647.

Early Breast Cancer Trialists' Collaborative Group (2005). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 365**:** 1687-1717.

Eckstein N., Servan K., Girard L., Cai D., von Jonquieres G., Jaehde U., Kassack M.U., Gazdar A.F., Minna J.D., Royer H.D. (2008). Epidermal growth factor receptor pathway analysis identifies amphiregulin as a key factor for cisplatin resistance of human breast cancer cells. *J Biol Chem* 283**:** 739-750.

Ekins R.P. (1989). Multi-analyte immunoassay. *Journal of pharmaceutical and biomedical analysis* 7**:** 155-168.

Ellis I.O., Schnitt S.J., Sastre-Garau X., Bussolati G., Tavassoli F.A., Eusebi V., Peterse J.L., Mukai K., Tabar L., Jacquemier J., Cornelisse C.J., Sasco A.J., Kaaks R., Pisani P., Goldgar D.E., Devilee P., Cleton-Jansen M.J., Borresen-Dale A.L., Van 't Veer L.J., Sapino A. (2003). Pathology and Genetics of Tumors of the Breast and Female Genital Organs. In: Kleihues P, Sobin LH (eds). *World Health Organization Classification of Tumours*.

Elsawaf Z., Sinn H.P. (2011). Triple-Negative Breast Cancer: Clinical and Histological Correlations. *Breast Care (Basel)* 6**:** 273-278.

Elston C.W., Ellis I.O. (1991). 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**:** 403-410.

Enmark E., Pelto-Huikko M., Grandien K., Lagercrantz S., Lagercrantz J., Fried G., Nordenskjold M., Gustafsson J.A. (1997). Human estrogen receptor beta-gene structure, chromosomal localization, and expression pattern. *The Journal of clinical endocrinology and metabolism* 82**:** 4258-4265.

Espina V., Edmiston K.H., Heiby M., Pierobon M., Sciro M., Merritt B., Banks S., Deng J., VanMeter A.J., Geho D.H., Pastore L., Sennesh J., Petricoin E.F., 3rd, Liotta L.A. (2008). A portrait of tissue phosphoprotein stability in the clinical tissue procurement process. *Mol Cell Proteomics* 7**:** 1998-2018.

Esserman L.J., Moore D.H., Tsing P.J., Chu P.W., Yau C., Ozanne E., Chung R.E., Tandon V.J., Park J.W., Baehner F.L., Kreps S., Tutt A.N., Gillett C.E., Benz C.C. (2011). Biologic markers determine both the risk and the timing of recurrence in breast cancer. *Breast Cancer Res Treat* 129**:** 607-616.

Fackenthal J.D., Olopade O.I. (2007). Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations. *Nat Rev Cancer* 7**:** 937-948.

Farese S.A., Aebi S. (2008). Infiltrating lobular carcinoma of the breast: systemic treatment. *Breast Dis* 30**:** 45-52.

Ferlay J., Shin H.R., Bray F., Forman D., Mathers C., Parkin D.M. (2010). GLOBOCAN 2008 v2.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10. *Lyon, France: International Agency for Research on Cancer; 2010 Available from: http://globocaniarcfr*.

Frank R., Hargreaves R. (2003). Clinical biomarkers in drug discovery and development. *Nature reviews Drug discovery* 2**:** 566-580.

Fulton R.J., McDade R.L., Smith P.L., Kienker L.J., Kettman J.R., Jr. (1997). Advanced multiplexed analysis with the FlowMetrix system. *Clin Chem* 43**:** 1749-1756.

Garcia I., Mayol G., Rios J., Domenech G., Cheung N.K., Oberthuer A., Fischer M., Maris J.M., Brodeur G.M., Hero B., Rodriguez E., Sunol M., Galvan P., de Torres C., Mora J., Lavarino C. (2012). A three-gene expression signature model for risk stratification of patients with neuroblastoma. *Clin Cancer Res* 18**:** 2012-2023.

Garrett T.P., McKern N.M., Lou M., Elleman T.C., Adams T.E., Lovrecz G.O., Kofler M., Jorissen R.N., Nice E.C., Burgess A.W., Ward C.W. (2003). The crystal structure of a truncated ErbB2 ectodomain reveals an active conformation, poised to interact with other ErbB receptors. *Mol Cell* 11**:** 495-505.

Gastaldi S., Comoglio P.M., Trusolino L. (2010). The Met oncogene and basal-like breast cancer: another culprit to watch out for? *Breast Cancer Res* 12**:** 208.

Gerdes J., Lemke H., Baisch H., Wacker H.H., Schwab U., Stein H. (1984). Cell cycle analysis of a cell proliferation-associated human nuclear antigen defined by the monoclonal antibody Ki-67. *J Immunol* 133**:** 1710-1715.

Geyer C.E., Forster J., Lindquist D., Chan S., Romieu C.G., Pienkowski T., Jagiello-Gruszfeld A., Crown J., Chan A., Kaufman B., Skarlos D., Campone M., Davidson N., Berger M., Oliva C., Rubin S.D., Stein S., Cameron D. (2006). Lapatinib plus capecitabine for HER2-positive advanced breast cancer. *N Engl J Med* 355**:** 2733-2743.

Goldhirsch A., Ingle J.N., Gelber R.D., Coates A.S., Thurlimann B., Senn H.J. (2009). Thresholds for therapies: highlights of the St Gallen International Expert Consensus on the primary therapy of early breast cancer 2009. *Ann Oncol* 20**:** 1319-1329.

Goldhirsch A., Wood W.C., Coates A.S., Gelber R.D., Thurlimann B., Senn H.J. (2011). Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol* 22**:** 1736-1747.

Gonzalez-Angulo A.M., Hennessy B.T., Meric-Bernstam F., Sahin A., Liu W., Ju Z., Carey M.S., Myhre S., Speers C., Deng L., Broaddus R., Lluch A., Aparicio S., Brown P., Pusztai L., Symmans W.F., Alsner J., Overgaard J., Borresen-Dale A.L., Hortobagyi G.N., Coombes K.R., Mills G.B. (2011). Functional proteomics can define prognosis and predict pathologic complete response in patients with breast cancer. *Clin Proteomics* 8**:** 11.

Gonzalez R.M., Seurynck-Servoss S.L., Crowley S.A., Brown M., Omenn G.S., Hayes D.F., Zangar R.C. (2008). Development and validation of sandwich ELISA microarrays with minimal assay interference. *J Proteome Res* 7**:** 2406-2414.

Gottlicher M., Heck S., Herrlich P. (1998). Transcriptional cross-talk, the second mode of steroid hormone receptor action. *J Mol Med (Berl)* 76**:** 480-489.

Green S., Walter P., Kumar V., Krust A., Bornert J.M., Argos P., Chambon P. (1986). Human oestrogen receptor cDNA: sequence, expression and homology to v-erb-A. *Nature* 320**:** 134-139.

Grubb R.L., Deng J., Pinto P.A., Mohler J.L., Chinnaiyan A., Rubin M., Linehan W.M., Liotta L.A., Petricoin E.F., Wulfkuhle J.D. (2009). Pathway biomarker profiling of localized and metastatic human prostate cancer reveal metastatic and prognostic signatures. *J Proteome Res* 8**:** 3044-3054.

Guo H., Liu W., Ju Z., Tamboli P., Jonasch E., Mills G.B., Lu Y., Hennessy B.T., Tsavachidou D. (2012). An efficient procedure for protein extraction from formalin-fixed, paraffin-embedded tissues for reverse phase protein arrays. *Proteome Sci* 10**:** 56.

Haibe-Kains B., Schroeder M., Culhane A., Bontempi G., Sotiriou C., Quackenbush J.F. (2011). genefu: Relevant Functions for Gene Expression Analysis, Especially in Breast Cancer. *R package version 121*.

Hall J.M., Lee M.K., Newman B., Morrow J.E., Anderson L.A., Huey B., King M.C. (1990). Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* 250**:** 1684-1689.

Hammond M.E., Hayes D.F., Dowsett M., Allred D.C., Hagerty K.L., Badve S., Fitzgibbons P.L., Francis G., Goldstein N.S., Hayes M., Hicks D.G., Lester S., Love R., Mangu P.B., McShane L., Miller K., Osborne C.K., Paik S., Perlmutter J., Rhodes A., Sasano H., Schwartz J.N., Sweep F.C., Taube S., Torlakovic E.E., Valenstein P., Viale G., Visscher D., Wheeler T., Williams R.B., Wittliff J.L., Wolff A.C. (2010). American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J Clin Oncol* 28**:** 2784-2795.

Hanahan D., Weinberg R.A. (2000). The hallmarks of cancer. *Cell* 100**:** 57-70.

Hanahan D., Weinberg R.A. (2011). Hallmarks of cancer: the next generation. *Cell* 144**:** 646-674.

Harvey J.M., Clark G.M., Osborne C.K., Allred D.C. (1999). 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**:** 1474-1481.

Haury A.C., Gestraud P., Vert J.P. (2011). The influence of feature selection methods on accuracy, stability and interpretability of molecular signatures. *PLoS One* 6**:** e28210.

Hayashi H., Abdollah S., Qiu Y., Cai J., Xu Y.Y., Grinnell B.W., Richardson M.A., Topper J.N., Gimbrone M.A., Jr., Wrana J.L., Falb D. (1997). The MAD-related protein Smad7 associates with the TGFbeta receptor and functions as an antagonist of TGFbeta signaling. *Cell* 89**:** 1165-1173.

Heil J., Gondos A., Rauch G., Marme F., Rom J., Golatta M., Junkermann H., Sinn P., Aulmann S., Debus J., Hof H., Schutz F., Brenner H., Sohn C., Schneeweiss A. (2012). Outcome analysis of patients with primary breast cancer initially treated at a certified academic breast unit. *Breast* 21**:** 303-308.

Henjes F., Bender C., von der Heyde S., Braun L., Mannsperger H.A., Schmidt C., Wiemann S., Hasmann M., Aulmann S., Beissbarth T., Korf U. (2012). Strong EGFR signaling in cell line models of ERBB2-amplified breast cancer attenuates response towards ERBB2-targeting drugs. *Oncogenesis* 1**:** e16.

Hennessy B.T., Lu Y., Gonzalez-Angulo A.M., Carey M.S., Myhre S., Ju Z., Davies M.A., Liu W., Coombes K., Meric-Bernstam F., Bedrosian I., McGahren M., Agarwal R., Zhang F., Overgaard J., Alsner J., Neve R.M., Kuo W.L., Gray J.W., Borresen-Dale A.L., Mills G.B. (2010). A Technical Assessment of the Utility of Reverse Phase Protein Arrays for the Study of the Functional Proteome in Non-microdissected Human Breast Cancers. *Clin Proteomics* 6**:** 129-151.

Hoch R.V., Thompson D.A., Baker R.J., Weigel R.J. (1999). GATA-3 is expressed in association with estrogen receptor in breast cancer. *Int J Cancer* 84**:** 122-128.

Hynes N.E., Watson C.J. (2010). Mammary gland growth factors: roles in normal development and in cancer. *Cold Spring Harb Perspect Biol* 2**:** a003186.

Iadevaia S., Lu Y., Morales F.C., Mills G.B., Ram P.T. (2010). Identification of optimal drug combinations targeting cellular networks: integrating phospho-proteomics and computational network analysis. *Cancer Res* 70**:** 6704-6714.

Igawa M., Rukstalis D.B., Tanabe T., Chodak G.W. (1994). High levels of nm23 expression are related to cell proliferation in human prostate cancer. *Cancer Res* 54**:** 1313-1318.

Ingvarsson J., Larsson A., Sjoholm A.G., Truedsson L., Jansson B., Borrebaeck C.A., Wingren C. (2007). Design of recombinant antibody microarrays for serum protein profiling: targeting of complement proteins. *J Proteome Res* 6**:** 3527-3536.

Iqbal J., Thike A.A., Cheok P.Y., Tse G.M., Tan P.H. (2012). Insulin growth factor receptor-1 expression and loss of PTEN protein predict early recurrence in triple-negative breast cancer. *Histopathology*.

Jasin M. (2002). Homologous repair of DNA damage and tumorigenesis: the BRCA connection. *Oncogene* 21**:** 8981-8993.

Jocker A., Sonntag J., Henjes F., Gotschel F., Tresch A., Beissbarth T., Wiemann S., Korf U. (2010). QuantProReloaded: quantitative analysis of microspot immunoassays. *Bioinformatics* 26**:** 2480-2481.

Kahlert S., Nuedling S., van Eickels M., Vetter H., Meyer R., Grohe C. (2000). Estrogen receptor alpha rapidly activates the IGF-1 receptor pathway. *J Biol Chem* 275**:** 18447-18453.

Kantelhardt E.J., Vetter M., Schmidt M., Veyret C., Augustin D., Hanf V., Meisner C., Paepke D., Schmitt M., Sweep F., von Minckwitz G., Martin P.M., Jaenicke F., Thomssen C., Harbeck N. (2011). Prospective evaluation of prognostic factors uPA/PAI-1 in node-negative breast cancer: phase III NNBC3-Europe trial (AGO, GBG, EORTC-PBG) comparing 6xFEC versus 3xFEC/3xDocetaxel. *BMC Cancer* 11**:** 140.

Karapetis C.S., Khambata-Ford S., Jonker D.J., O'Callaghan C.J., Tu D., Tebbutt N.C., Simes R.J., Chalchal H., Shapiro J.D., Robitaille S., Price T.J., Shepherd L., Au H.J., Langer C., Moore M.J., Zalcberg J.R. (2008). K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med* 359**:** 1757-1765.

Kaufmann M., von Minckwitz G., Bear H.D., Buzdar A., McGale P., Bonnefoi H., Colleoni M., Denkert C., Eiermann W., Jackesz R., Makris A., Miller W., Pierga J.Y., Semiglazov V., Schneeweiss A., Souchon R., Stearns V., Untch M., Loibl S. (2007). Recommendations from an international expert panel on the use of neoadjuvant (primary) systemic treatment of operable breast cancer: new perspectives 2006. *Ann Oncol* 18**:** 1927-1934.

Kavsak P., Rasmussen R.K., Causing C.G., Bonni S., Zhu H., Thomsen G.H., Wrana J.L. (2000). Smad7 binds to Smurf2 to form an E3 ubiquitin ligase that targets the TGF beta receptor for degradation. *Mol Cell* 6**:** 1365-1375.

Kenny P.A., Bissell M.J. (2007). Targeting TACE-dependent EGFR ligand shedding in breast cancer. *J Clin Invest* 117**:** 337-345.

Koboldt D.C., Fulton R.S., McLellan M.D., Schmidt H., Kalicki-Veizer J., McMichael J.F., Fulton L.L., Dooling D.J., Ding L., Mardis E.R., Wilson R.K., Ally A., Balasundaram M., Butterfield Y.S., Carlsen R., Carter C., Chu A., Chuah E., Chun H.J., Coope R.J., Dhalla N., Guin R., Hirst C., Hirst M., Holt R.A., Lee D., Li H.I., Mayo M., Moore R.A., Mungall A.J., Pleasance E., Gordon Robertson A., Schein J.E., Shafiei A., Sipahimalani P., Slobodan J.R., Stoll D., Tam A., Thiessen N., Varhol R.J., Wye N., Zeng T., Zhao Y., Birol I., Jones S.J., Marra M.A., Cherniack A.D., Saksena G., Onofrio R.C., Pho N.H., Carter S.L., Schumacher S.E., Tabak B., Hernandez B., Gentry J., Nguyen H., Crenshaw A., Ardlie K., Beroukhim R., Winckler W., Getz G., Gabriel S.B., Meyerson M., Chin L., Park P.J., Kucherlapati R., Hoadley K.A., Todd Auman J., Fan C., Turman Y.J., Shi Y., Li L., Topal M.D., He X., Chao H.H., Prat A., Silva G.O., Iglesia M.D., Zhao W., Usary J., Berg J.S., Adams M., Brooker J., Wu J., Gulabani A., Bodenheimer T., Hoyle A.P., Simons J.V., Soloway M.G., Mose L.E., Jefferys S.R., Balu S., Parker J.S., Neil Hayes D., Perou C.M., Malik S., Mahurkar S., Shen H., Weisenberger D.J., Triche Jr T., Lai P.H., Bootwalla M.S., Maglinte D.T., Berman B.P., Van Den Berg D.J., Baylin S.B., Laird P.W., Creighton C.J., Donehower L.A., Noble M., Voet D., Gehlenborg N., Dicara D., Zhang J., Zhang H., Wu C.J., Yingchun Liu S., Lawrence M.S., Zou L., Sivachenko A., Lin P., Stojanov P., Jing R., Cho J., Sinha R., Park R.W., Nazaire M.D., Robinson J., Thorvaldsdottir H., Mesirov J., Reynolds S., Kreisberg R.B., Bernard B., Bressler R., Erkkila T., Lin J., Thorsson V., Zhang W., Shmulevich I., Ciriello G., Weinhold N., Schultz N., Gao J., Cerami E., Gross B., Jacobsen A., Arman Aksoy B., Antipin Y., Reva B., Shen R., Taylor B.S., Ladanyi M., Sander C., Anur P., Spellman P.T., Lu Y., Liu W., Verhaak R.R., Mills G.B., Akbani R., Zhang N., Broom B.M., Casasent T.D., Wakefield C., Unruh A.K., Baggerly K., Coombes K., Weinstein J.N., Haussler D., Benz C.C., Stuart J.M., Benz S.C., Zhu J., Szeto C.C., Scott G.K., Yau C., Paull E.O., Carlin D., Wong C., Sokolov A., Thusberg J., Mooney S., Ng S., Goldstein T.C., Ellrott K., Grifford M., Wilks C., Ma S., Craft B., Yan C., Hu Y., Meerzaman D., Gastier-Foster J.M., Bowen J., Ramirez N.C., Black A.D., Xpath Error Unknown Variable Tname R.E., White P., Zmuda E.J., Frick J., Lichtenberg T.M., Brookens R., George M.M., Gerken M.A., Harper H.A., Leraas K.M., Wise L.J., Tabler T.R., McAllister C., Barr T., Hart-Kothari M., Tarvin K., Saller C., Sandusky G., Mitchell C., Iacocca M.V., Brown J., Rabeno B., Czerwinski C., Petrelli N., Dolzhansky O., Abramov M., Voronina O., Potapova O., Marks J.R., Suchorska W.M., Murawa D., Kycler W., Ibbs M., Korski K., Spychala A., Murawa P., Brzezinski J.J., Perz H., Lazniak R., Teresiak M., Tatka H., Leporowska E., Bogusz-Czerniewicz M., Malicki J., Mackiewicz A., Wiznerowicz M., Van Le X., Kohl B., Viet Tien N., Thorp R., Van Bang N., Sussman H., Duc Phu B., Hajek R., Phi Hung N., Viet The Phuong T., Quyet Thang H., Zaki Khan K., Penny R., Mallery D., Curley E., Shelton C., Yena P., Ingle J.N., Couch F.J., Lingle W.L., King T.A., Maria Gonzalez-Angulo A., Dyer M.D., Liu S., Meng X., Patangan M., Waldman F., Stoppler H., Kimryn Rathmell W., Thorne L., Huang M., Boice L., Hill A., Morrison C., Gaudioso C., Bshara W., Daily K., Egea S.C., Pegram M.D., Gomez-Fernandez C., Dhir R., Bhargava R., Brufsky A., Shriver C.D., Hooke J.A., Leigh Campbell J., Mural R.J., Hu H., Somiari S., Larson C., Deyarmin B., Kvecher L., Kovatich A.J., Ellis M.J., Stricker T., White K., Olopade O., Luo C., Chen Y., Bose R., Chang L.W., Beck A.H., Pihl T., Jensen M., Sfeir R., Kahn A., Kothiyal P., Wang Z., Snyder E., Pontius J., Ayala B., Backus M., Walton J., Baboud J., Berton D., Nicholls M., Srinivasan D., Raman R., Girshik S., Kigonya P., Alonso S., Sanbhadti R., Barletta S., Pot D., Sheth M., Demchok J.A., Mills Shaw K.R., Yang L., Eley G., Ferguson M.L., Tarnuzzer R.W., Dillon L.A., Buetow K., Fielding P., Ozenberger B.A., Guyer M.S., Sofia H.J., Palchik J.D. (2012). Comprehensive molecular portraits of human breast tumours. *Nature* 490**:** 61-70.

Konecny G.E., Meng Y.G., Untch M., Wang H.J., Bauerfeind I., Epstein M., Stieber P., Vernes J.M., Gutierrez J., Hong K., Beryt M., Hepp H., Slamon D.J., Pegram M.D. (2004). Association between HER-2/neu and vascular endothelial growth factor expression predicts clinical outcome in primary breast cancer patients. *Clin Cancer Res* 10**:** 1706-1716.

Korf U., Derdak S., Tresch A., Henjes F., Schumacher S., Schmidt C., Hahn B., Lehmann W.D., Poustka A., Beissbarth T., Klingmuller U. (2008a). Quantitative protein microarrays for time-resolved measurements of protein phosphorylation. *Proteomics* 8**:** 4603-4612.

Korf U., Henjes F., Schmidt C., Tresch A., Mannsperger H., Lobke C., Beissbarth T., Poustka A. (2008b). Antibody microarrays as an experimental platform for the analysis of signal transduction networks. *Adv Biochem Eng Biotechnol* 110**:** 153-175.

Kornblau S.M., Singh N., Qiu Y., Chen W., Zhang N., Coombes K.R. (2010). Highly phosphorylated FOXO3A is an adverse prognostic factor in acute myeloid leukemia. *Clin Cancer Res* 16**:** 1865-1874.

Krecke K.N., Gisvold J.J. (1993). Invasive lobular carcinoma of the breast: mammographic findings and extent of disease at diagnosis in 184 patients. *AJR American journal of roentgenology* 161**:** 957-960.

Kreienberg R., Albert U.S., Follmann M., Kopp I., Kühn T., Wöckel A., Zemmler T. (2012). *Interdisziplinäre S3-Leitlinie für die Diagnostik, Therapie und Nachsorge des Mammakarzinoms : Langversion 3.0, Aktualisierung 2012*. Leitlinienprogramm Onkologie der AWMF, Deutschen Krebsgesellschaft e.V. und Deutschen Krebshilfe e.V.

Kretzschmar M. (2000). Transforming growth factor-beta and breast cancer: Transforming growth factor-beta/SMAD signaling defects and cancer. *Breast Cancer Res* 2**:** 107-115.

Kursa M.B., Rudnicki W.R. (2010). Feature Selection with the Boruta Package. *Journal of Statistical Software* 36**:** 1-13.

Kusnezow W., Hoheisel J.D. (2003). Solid supports for microarray immunoassays. *Journal of molecular recognition : JMR* 16**:** 165-176.

Kwak E.L., Bang Y.J., Camidge D.R., Shaw A.T., Solomon B., Maki R.G., Ou S.H., Dezube B.J., Janne P.A., Costa D.B., Varella-Garcia M., Kim W.H., Lynch T.J., Fidias P., Stubbs H., Engelman J.A., Sequist L.V., Tan W., Gandhi L., Mino-Kenudson M., Wei G.C., Shreeve S.M., Ratain M.J., Settleman J., Christensen J.G., Haber D.A., Wilner K., Salgia R., Shapiro G.I., Clark J.W., Iafrate A.J. (2010). Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med* 363**:** 1693-1703.

Lam S., Boyle P., Healey G.F., Maddison P., Peek L., Murray A., Chapman C.J., Allen J., Wood W.C., Sewell H.F., Robertson J.F. (2011). EarlyCDT-Lung: an immunobiomarker test as an aid to early detection of lung cancer. *Cancer Prev Res (Phila)* 4**:** 1126-1134.

Lash G.E., Scaife P.J., Innes B.A., Otun H.A., Robson S.C., Searle R.F., Bulmer J.N. (2006). Comparison of three multiplex cytokine analysis systems: Luminex, SearchLight and FAST Quant. *J Immunol Methods* 309**:** 205-208.

LeJeune S., Leek R., Horak E., Plowman G., Greenall M., Harris A.L. (1993). Amphiregulin, epidermal growth factor receptor, and estrogen receptor expression in human primary breast cancer. *Cancer Res* 53**:** 3597-3602.

Leong A.S. (2004). Quantitation in immunohistology: fact or fiction? A discussion of variables that influence results. *Appl Immunohistochem Mol Morphol* 12**:** 1-7.

Liedtke C., Hatzis C., Symmans W.F., Desmedt C., Haibe-Kains B., Valero V., Kuerer H., Hortobagyi G.N., Piccart-Gebhart M., Sotiriou C., Pusztai L. (2009). Genomic grade index is associated with response to chemotherapy in patients with breast cancer. *J Clin Oncol* 27**:** 3185-3191.

Lilja H., Ulmert D., Vickers A.J. (2008). Prostate-specific antigen and prostate cancer: prediction, detection and monitoring. *Nat Rev Cancer* 8**:** 268-278.

Lin N.U., Winer E.P. (2008). Optimizing endocrine therapy for estrogen receptor-positive breast cancer: treating the right patients for the right length of time. *J Clin Oncol* 26**:** 1919-1921.

Lin X., Liang M., Feng X.H. (2000). Smurf2 is a ubiquitin E3 ligase mediating proteasome-dependent degradation of Smad2 in transforming growth factor-beta signaling. *J Biol Chem* 275**:** 36818-36822.

Linderholm B., Tavelin B., Grankvist K., Henriksson R. (1998). Vascular endothelial growth factor is of high prognostic value in node-negative breast carcinoma. *J Clin Oncol* 16**:** 3121-3128.

Loebke C., Sueltmann H., Schmidt C., Henjes F., Wiemann S., Poustka A., Korf U. (2007). Infrared-based protein detection arrays for quantitative proteomics. *Proteomics* 7**:** 558-564.

Lofblom J., Frejd F.Y., Stahl S. (2011). Non-immunoglobulin based protein scaffolds. *Current opinion in biotechnology* 22**:** 843-848.

Lombardi D., Lacombe M.L., Paggi M.G. (2000). nm23: unraveling its biological function in cell differentiation. *J Cell Physiol* 182**:** 144-149.

Longacre T.A., Ennis M., Quenneville L.A., Bane A.L., Bleiweiss I.J., Carter B.A., Catelano E., Hendrickson M.R., Hibshoosh H., Layfield L.J., Memeo L., Wu H., O'Malley F P. (2006). Interobserver agreement and reproducibility in classification of invasive breast carcinoma: an NCI breast cancer family registry study. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 19**:** 195-207.

Ma X.J., Salunga R., Tuggle J.T., Gaudet J., Enright E., McQuary P., Payette T., Pistone M., Stecker K., Zhang B.M., Zhou Y.X., Varnholt H., Smith B., Gadd M., Chatfield E., Kessler J., Baer T.M., Erlander M.G., Sgroi D.C. (2003). Gene expression profiles of human breast cancer progression. *Proc Natl Acad Sci U S A* 100**:** 5974-5979.

MacBeath G., Schreiber S.L. (2000). Printing proteins as microarrays for high-throughput function determination. *Science* 289**:** 1760-1763.

Mannsperger H.A. (2011). Analyse von Signalproteinen in Zelllinien und Tumorgewebe mit Reverse Phase Proteinarrays (Dissertation), Universität Heidelberg, Heidelberg.

Mannsperger H.A., Gade S., Henjes F., Beissbarth T., Korf U. (2010). RPPanalyzer: Analysis of reverse-phase protein array data. *Bioinformatics* 26**:** 2202-2203.

McIntyre E., Blackburn E., Brown P.J., Johnson C.G., Gullick W.J. (2009). The complete family of epidermal growth factor receptors and their ligands are co-ordinately expressed in breast cancer. *Breast Cancer Res Treat*.

Mengel M., von Wasielewski R., Wiese B., Rudiger T., Muller-Hermelink H.K., Kreipe H. (2002). Inter-laboratory and inter-observer reproducibility of immunohistochemical assessment of the Ki-67 labelling index in a large multi-centre trial. *J Pathol* 198**:** 292-299.

Metzger-Filho O., Tutt A., de Azambuja E., Saini K.S., Viale G., Loi S., Bradbury I., Bliss J.M., Azim H.A., Jr., Ellis P., Di Leo A., Baselga J., Sotiriou C., Piccart-Gebhart M. (2012). Dissecting the heterogeneity of triple-negative breast cancer. *J Clin Oncol* 30**:** 1879-1887.

Metzger Filho O., Ignatiadis M., Sotiriou C. (2011). Genomic Grade Index: An important tool for assessing breast cancer tumor grade and prognosis. *Crit Rev Oncol Hematol* 77**:** 20-29.

Meyerson M., Harlow E. (1994). Identification of G1 kinase activity for cdk6, a novel cyclin D partner. *Mol Cell Biol* 14**:** 2077-2086.

Meyuhas O. (2008). Physiological roles of ribosomal protein S6: one of its kind. *Int Rev Cell Mol Biol* 268**:** 1-37.

Migliaccio A., Di Domenico M., Castoria G., de Falco A., Bontempo P., Nola E., Auricchio F. (1996). Tyrosine kinase/p21ras/MAP-kinase pathway activation by estradiol-receptor complex in MCF-7 cells. *EMBO J* 15**:** 1292-1300.

Miles D.W., de Haas S.L., Dirix L.Y., Romieu G., Chan A., Pivot X., Tomczak P., Provencher L., Cortes J., Delmar P.R., Scherer S.J. (2013). Biomarker results from the AVADO phase 3 trial of first-line bevacizumab plus docetaxel for HER2-negative metastatic breast cancer. *Br J Cancer* 108**:** 1052-1060.

Molina M.A., Codony-Servat J., Albanell J., Rojo F., Arribas J., Baselga J. (2001). Trastuzumab (herceptin), a humanized anti-Her2 receptor monoclonal antibody, inhibits basal and activated Her2 ectodomain cleavage in breast cancer cells. *Cancer Res* 61**:** 4744-4749.

Mueller C., Edmiston K.H., Carpenter C., Gaffney E., Ryan C., Ward R., White S., Memeo L., Colarossi C., Petricoin E.F., 3rd, Liotta L.A., Espina V. (2011). One-step preservation of phosphoproteins and tissue morphology at room temperature for diagnostic and research specimens. *PLoS One* 6**:** e23780.

Mueller K.L., Madden J.M., Zoratti G.L., Kuperwasser C., List K., Boerner J.L. (2012). Fibroblast-secreted hepatocyte growth factor mediates epidermal growth factor receptor tyrosine kinase inhibitor resistance in triple-negative breast cancers through paracrine activation of Met. *Breast Cancer Res* 14**:** R104.

Nielsen U.B., Geierstanger B.H. (2004). Multiplexed sandwich assays in microarray format. *J Immunol Methods* 290**:** 107-120.

Nieto Y., Nawaz F., Jones R.B., Shpall E.J., Nawaz S. (2007). Prognostic significance of overexpression and phosphorylation of epidermal growth factor receptor (EGFR) and the presence of truncated EGFRvIII in locoregionally advanced breast cancer. *J Clin Oncol* 25**:** 4405-4413.

Nilsson S., Makela S., Treuter E., Tujague M., Thomsen J., Andersson G., Enmark E., Pettersson K., Warner M., Gustafsson J.A. (2001). Mechanisms of estrogen action. *Physiological reviews* 81**:** 1535-1565.

O'Shaughnessy J., Osborne C., Pippen J.E., Yoffe M., Patt D., Rocha C., Koo I.C., Sherman B.M., Bradley C. (2011). Iniparib plus chemotherapy in metastatic triple-negative breast cancer. *N Engl J Med* 364**:** 205-214.

Otero A.S. (2000). NM23/nucleoside diphosphate kinase and signal transduction. *J Bioenerg Biomembr* 32**:** 269-275.

Paik S., Shak S., Tang G., Kim C., Baker J., Cronin M., Baehner F.L., Walker M.G., Watson D., Park T., Hiller W., Fisher E.R., Wickerham D.L., Bryant J., Wolmark N. (2004). A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 351**:** 2817-2826.

Parker J.S., Mullins M., Cheang M.C., Leung S., Voduc D., Vickery T., Davies S., Fauron C., He X., Hu Z., Quackenbush J.F., Stijleman I.J., Palazzo J., Marron J.S., Nobel A.B., Mardis E., Nielsen T.O., Ellis M.J., Perou C.M., Bernard P.S. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 27**:** 1160-1167.

Paweletz C.P., Charboneau L., Bichsel V.E., Simone N.L., Chen T., Gillespie J.W., Emmert-Buck M.R., Roth M.J., Petricoin I.E., Liotta L.A. (2001). Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. *Oncogene* 20**:** 1981-1989.

Pawlak M., Schick E., Bopp M.A., Schneider M.J., Oroszlan P., Ehrat M. (2002). Zeptosens' protein microarrays: a novel high performance microarray platform for low abundance protein analysis. *Proteomics* 2**:** 383-393.

Peralta Soler A., Knudsen K.A., Salazar H., Han A.C., Keshgegian A.A. (1999). P-cadherin expression in breast carcinoma indicates poor survival. *Cancer* 86**:** 1263-1272.

Perez E.A., Moreno-Aspitia A., Aubrey Thompson E., Andorfer C.A. (2010). Adjuvant therapy of triple negative breast cancer. *Breast Cancer Res Treat* 120**:** 285-291.

Perou C.M., Sorlie T., Eisen M.B., van de Rijn M., Jeffrey S.S., Rees C.A., Pollack J.R., Ross D.T., Johnsen H., Akslen L.A., Fluge O., Pergamenschikov A., Williams C., Zhu S.X., Lonning P.E., Borresen-Dale A.L., Brown P.O., Botstein D. (2000). Molecular portraits of human breast tumours. *Nature* 406**:** 747-752.

Perrot-Applanat M., Di Benedetto M. (2012). Autocrine functions of VEGF in breast tumor cells: adhesion, survival, migration and invasion. *Cell adhesion & migration* 6**:** 547-553.

Pestalozzi B.C., Zahrieh D., Mallon E., Gusterson B.A., Price K.N., Gelber R.D., Holmberg S.B., Lindtner J., Snyder R., Thurlimann B., Murray E., Viale G., Castiglione-Gertsch M., Coates A.S., Goldhirsch A. (2008). Distinct clinical and prognostic features of infiltrating lobular carcinoma of the breast: combined results of 15 International Breast Cancer Study Group clinical trials. *J Clin Oncol* 26**:** 3006-3014.

Polyak K., Kalluri R. (2010). The role of the microenvironment in mammary gland development and cancer. *Cold Spring Harb Perspect Biol* 2**:** a003244.

Press M.F., Bernstein L., Thomas P.A., Meisner L.F., Zhou J.Y., Ma Y., Hung G., Robinson R.A., Harris C., El-Naggar A., Slamon D.J., Phillips R.N., Ross J.S., Wolman S.R., Flom K.J. (1997). HER-2/neu gene amplification characterized by fluorescence in situ hybridization: poor prognosis in node-negative breast carcinomas. *J Clin Oncol* 15**:** 2894-2904.

Raghav K.P., Wang W., Liu S., Chavez-MacGregor M., Meng X., Hortobagyi G.N., Mills G.B., Meric-Bernstam F., Blumenschein G.R., Jr., Gonzalez-Angulo A.M. (2012). cMET and phospho-cMET protein levels in breast cancers and survival outcomes. *Clin Cancer Res* 18**:** 2269-2277.

Remmele W., Stegner H.E. (1987). [Recommendation for uniform definition of an immunoreactive score (IRS) for immunohistochemical estrogen receptor detection (ER-ICA) in breast cancer tissue]. *Pathologe* 8**:** 138-140.

Revillion F., Lhotellier V., Hornez L., Bonneterre J., Peyrat J.P. (2008). ErbB/HER ligands in human breast cancer, and relationships with their receptors, the bio-pathological features and prognosis. *Ann Oncol* 19**:** 73-80.

Ringner M., Fredlund E., Hakkinen J., Borg A., Staaf J. (2011). GOBO: gene expression-based outcome for breast cancer online. *PLoS One* 6**:** e17911.

Robert Koch-Institut, Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (2012). *Krebs in Deutschland 2007/2008*, vol. 8.

Robinson D.R., Wu Y.M., Lin S.F. (2000). The protein tyrosine kinase family of the human genome. *Oncogene* 19**:** 5548-5557.

Robinson W.H., DiGennaro C., Hueber W., Haab B.B., Kamachi M., Dean E.J., Fournel S., Fong D., Genovese M.C., de Vegvar H.E., Skriner K., Hirschberg D.L., Morris R.I., Muller S., Pruijn G.J., van Venrooij W.J., Smolen J.S., Brown P.O., Steinman L., Utz P.J. (2002). Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nat Med* 8**:** 295-301.

Rustin G.J., Marples M., Nelstrop A.E., Mahmoudi M., Meyer T. (2001). Use of CA-125 to define progression of ovarian cancer in patients with persistently elevated levels. *J Clin Oncol* 19**:** 4054-4057.

Sainsbury J.R., Farndon J.R., Needham G.K., Malcolm A.J., Harris A.L. (1987). Epidermal-growth-factor receptor status as predictor of early recurrence of and death from breast cancer. *Lancet* 1**:** 1398-1402.

Saphner T., Tormey D.C., Gray R. (1996). Annual hazard rates of recurrence for breast cancer after primary therapy. *J Clin Oncol* 14**:** 2738-2746.

Sarrio D., Moreno-Bueno G., Hardisson D., Sanchez-Estevez C., Guo M., Herman J.G., Gamallo C., Esteller M., Palacios J. (2003). Epigenetic and genetic alterations of APC and CDH1 genes in lobular breast cancer: relationships with abnormal E-cadherin and catenin expression and microsatellite instability. *Int J Cancer* 106**:** 208-215.

Schroder C., Jacob A., Tonack S., Radon T.P., Sill M., Zucknick M., Ruffer S., Costello E., Neoptolemos J.P., Crnogorac-Jurcevic T., Bauer A., Fellenberg K., Hoheisel J.D. (2010). Dual-color proteomic profiling of complex samples with a microarray of 810 cancer-related antibodies. *Mol Cell Proteomics* 9**:** 1271-1280.

Schuler M., Awada A., Harter P., Canon J.L., Possinger K., Schmidt M., De Greve J., Neven P., Dirix L., Jonat W., Beckmann M.W., Schutte J., Fasching P.A., Gottschalk N., Besse-Hammer T., Fleischer F., Wind S., Uttenreuther-Fischer M., Piccart M., Harbeck N. (2012). A phase II trial to assess efficacy and safety of afatinib in extensively pretreated patients with HER2-negative metastatic breast cancer. *Breast Cancer Res Treat* 134**:** 1149-1159.

Schwenk J.M., Gry M., Rimini R., Uhlen M., Nilsson P. (2008). Antibody suspension bead arrays within serum proteomics. *J Proteome Res* 7**:** 3168-3179.

Simpson E.R., Davis S.R. (2001). Minireview: aromatase and the regulation of estrogen biosynthesis--some new perspectives. *Endocrinology* 142**:** 4589-4594.

Slamon D.J., Clark G.M., Wong S.G., Levin W.J., Ullrich A., McGuire W.L. (1987). Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 235**:** 177-182.

Slamon D.J., Leyland-Jones B., Shak S., Fuchs H., Paton V., Bajamonde A., Fleming T., Eiermann W., Wolter J., Pegram M., Baselga J., Norton L. (2001). Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 344**:** 783-792.

Sloan E.K., Stanley K.L., Anderson R.L. (2004). Caveolin-1 inhibits breast cancer growth and metastasis. *Oncogene* 23**:** 7893-7897.

Sonntag J., Mannsperger H., Jocker A., Korf U. (2011). Microspot immunoassay-based analysis of plasma protein profiles for biomarker discovery strategies. *Methods Mol Biol* 785**:** 237-245.

Sorlie T., Perou C.M., Tibshirani R., Aas T., Geisler S., Johnsen H., Hastie T., Eisen M.B., van de Rijn M., Jeffrey S.S., Thorsen T., Quist H., Matese J.C., Brown P.O., Botstein D., Lonning P.E., Borresen-Dale A.L. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 98**:** 10869-10874.

Sorlie T., Tibshirani R., Parker J., Hastie T., Marron J.S., Nobel A., Deng S., Johnsen H., Pesich R., Geisler S., Demeter J., Perou C.M., Lonning P.E., Brown P.O., Borresen-Dale A.L., Botstein D. (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 100**:** 8418-8423.

Sotgia F., Martinez-Outschoorn U.E., Howell A., Pestell R.G., Pavlides S., Lisanti M.P. (2012). Caveolin-1 and cancer metabolism in the tumor microenvironment: markers, models, and mechanisms. *Annu Rev Pathol* 7**:** 423-467.

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

Sotiriou C., Wirapati P., Loi S., Harris A., Fox S., Smeds J., Nordgren H., Farmer P., Praz V., Haibe-Kains B., Desmedt C., Larsimont D., Cardoso F., Peterse H., Nuyten D., Buyse M., Van de Vijver M.J., Bergh J., Piccart M., Delorenzi M. (2006). Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 98**:** 262-272.

Sousa B., Paredes J., Milanezi F., Lopes N., Martins D., Dufloth R., Vieira D., Albergaria A., Veronese L., Carneiro V., Carvalho S., Costa J.L., Zeferino L., Schmitt F. (2010). P-cadherin, vimentin and CK14 for identification of basal-like phenotype in breast carcinomas: an immunohistochemical study. *Histol Histopathol* 25**:** 963-974.

Sparano J.A., Paik S. (2008). Development of the 21-gene assay and its application in clinical practice and clinical trials. *J Clin Oncol* 26**:** 721-728.

Spurrier B., Ramalingam S., Nishizuka S. (2008). Reverse-phase protein lysate microarrays for cell signaling analysis. *Nat Protoc* 3**:** 1796-1808.

Steeg P.S., Bevilacqua G., Kopper L., Thorgeirsson U.P., Talmadge J.E., Liotta L.A., Sobel M.E. (1988). Evidence for a novel gene associated with low tumor metastatic potential. *J Natl Cancer Inst* 80**:** 200-204.

Stephens P.J., Tarpey P.S., Davies H., Van Loo P., Greenman C., Wedge D.C., Nik-Zainal S., Martin S., Varela I., Bignell G.R., Yates L.R., Papaemmanuil E., Beare D., Butler A., Cheverton A., Gamble J., Hinton J., Jia M., Jayakumar A., Jones D., Latimer C., Lau K.W., McLaren S., McBride D.J., Menzies A., Mudie L., Raine K., Rad R., Chapman M.S., Teague J., Easton D., Langerod A., Lee M.T., Shen C.Y., Tee B.T., Huimin B.W., Broeks A., Vargas A.C., Turashvili G., Martens J., Fatima A., Miron P., Chin S.F., Thomas G., Boyault S., Mariani O., Lakhani S.R., van de Vijver M., van 't Veer L., Foekens J., Desmedt C., Sotiriou C., Tutt A., Caldas C., Reis-Filho J.S., Aparicio S.A., Salomon A.V., Borresen-Dale A.L., Richardson A.L., Campbell P.J., Futreal P.A., Stratton M.R. (2012). The landscape of cancer genes and mutational processes in breast cancer. *Nature* 486**:** 400-404.

Sweep F.C.G.J., Thomas C.M.G., Schmitt M. (2006). Analytical aspects of biomarker immunoassays in cancer research. In: Gasparini G, Hayes DF (eds). *Biomarkers in breast cancer: Molecular diagnostics for predicting and monitoring therapeutic effect*. Humana Press. pp 17-30.

Takadate T., Onogawa T., Fujii K., Motoi F., Mikami S., Fukuda T., Kihara M., Suzuki T., Takemura T., Minowa T., Hanagata N., Kinoshita K., Morikawa T., Shirasaki K., Rikiyama T., Katayose Y., Egawa S., Nishimura T., Unno M. (2012). Nm23/Nucleoside Diphosphate Kinase-A as a Potent Prognostic marker in Invasive Pancreatic Ductal Carcinoma Identified by Proteomic Analysis of Laser Micro-dissected Formalin-Fixed Paraffin-Embedded Tissue. *Clin Proteomics* 9**:** 8.

Takasaki Y., Deng J.S., Tan E.M. (1981). A nuclear antigen associated with cell proliferation and blast transformation. *The Journal of experimental medicine* 154**:** 1899-1909.

Tibshirani R., Hastie T., Narasimhan B., Chu G. (2002). Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A* 99**:** 6567-6572.

Toedter G., Hayden K., Wagner C., Brodmerkel C. (2008). Simultaneous detection of eight analytes in human serum by two commercially available platforms for multiplex cytokine analysis. *Clin Vaccine Immunol* 15**:** 42-48.

Troncale S., Barbet A., Coulibaly L., Henry E., He B., Barillot E., Dubois T., Hupe P., de Koning L. (2012). NormaCurve: a SuperCurve-based method that simultaneously quantifies and normalizes reverse phase protein array data. *PLoS One* 7**:** e38686.

Tubiana-Hulin M., Stevens D., Lasry S., Guinebretiere J.M., Bouita L., Cohen-Solal C., Cherel P., Rouesse J. (2006). Response to neoadjuvant chemotherapy in lobular and ductal breast carcinomas: a retrospective study on 860 patients from one institution. *Ann Oncol* 17**:** 1228-1233.

Turner N., Pearson A., Sharpe R., Lambros M., Geyer F., Lopez-Garcia M.A., Natrajan R., Marchio C., Iorns E., Mackay A., Gillett C., Grigoriadis A., Tutt A., Reis-Filho J.S., Ashworth A. (2010). FGFR1 amplification drives endocrine therapy resistance and is a therapeutic target in breast cancer. *Cancer Res* 70**:** 2085-2094.

Uhlmann S., Mannsperger H., Zhang J.D., Horvat E.A., Schmidt C., Kublbeck M., Henjes F., Ward A., Tschulena U., Zweig K., Korf U., Wiemann S., Sahin O. (2012). Global microRNA level regulation of EGFR-driven cell-cycle protein network in breast cancer. *Mol Syst Biol* 8**:** 570.

van der Hage J.A., van de Velde C.J., Julien J.P., Tubiana-Hulin M., Vandervelden C., Duchateau L. (2001). Preoperative chemotherapy in primary operable breast cancer: results from the European Organization for Research and Treatment of Cancer trial 10902. *J Clin Oncol* 19**:** 4224-4237.

Van Keymeulen A., Rocha A.S., Ousset M., Beck B., Bouvencourt G., Rock J., Sharma N., Dekoninck S., Blanpain C. (2011). Distinct stem cells contribute to mammary gland development and maintenance. *Nature* 479**:** 189-193.

Varga Z., Diebold J., Dommann-Scherrer C., Frick H., Kaup D., Noske A., Obermann E., Ohlschlegel C., Padberg B., Rakozy C., Sancho Oliver S., Schobinger-Clement S., Schreiber-Facklam H., Singer G., Tapia C., Wagner U., Mastropasqua M.G., Viale G., Lehr H.A. (2012). How reliable is Ki-67 immunohistochemistry in grade 2 breast carcinomas? A QA study of the Swiss Working Group of Breast- and Gynecopathologists. *PLoS One* 7**:** e37379.

Verma S., Miles D., Gianni L., Krop I.E., Welslau M., Baselga J., Pegram M., Oh D.Y., Dieras V., Guardino E., Fang L., Lu M.W., Olsen S., Blackwell K. (2012). Trastuzumab emtansine for HER2-positive advanced breast cancer. *N Engl J Med* 367**:** 1783-1791.

Vetter G., Saumet A., Moes M., Vallar L., Le Bechec A., Laurini C., Sabbah M., Arar K., Theillet C., Lecellier C.H., Friederich E. (2010). miR-661 expression in SNAI1-induced epithelial to mesenchymal transition contributes to breast cancer cell invasion by targeting Nectin-1 and StarD10 messengers. *Oncogene* 29**:** 4436-4448.

Viale G., Giobbie-Hurder A., Regan M.M., Coates A.S., Mastropasqua M.G., Dell'Orto P., Maiorano E., MacGrogan G., Braye S.G., Ohlschlegel C., Neven P., Orosz Z., Olszewski W.P., Knox F., Thurlimann B., Price K.N., Castiglione-Gertsch M., Gelber R.D., Gusterson B.A., Goldhirsch A. (2008). Prognostic and predictive value of centrally reviewed Ki-67 labeling index in postmenopausal women with endocrine-responsive breast cancer: results from Breast International Group Trial 1-98 comparing adjuvant tamoxifen with letrozole. *J Clin Oncol* 26**:** 5569-5575.

Visvader J.E. (2009). Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. *Genes & development* 23**:** 2563-2577.

Walker S.R., Nelson E.A., Zou L., Chaudhury M., Signoretti S., Richardson A., Frank D.A. (2009). Reciprocal effects of STAT5 and STAT3 in breast cancer. *Molecular cancer research : MCR* 7**:** 966-976.

Wallwiener M., Hartkopf A.D., Baccelli I., Riethdorf S., Schott S., Pantel K., Marme F., Sohn C., Trumpp A., Rack B., Aktas B., Solomayer E.F., Muller V., Janni W., Schneeweiss A., Fehm T.N. (2013). The prognostic impact of circulating tumor cells in subtypes of metastatic breast cancer. *Breast Cancer Res Treat* 137**:** 503-510.

Wang T., Niu G., Kortylewski M., Burdelya L., Shain K., Zhang S., Bhattacharya R., Gabrilovich D., Heller R., Coppola D., Dalton W., Jove R., Pardoll D., Yu H. (2004). Regulation of the innate and adaptive immune responses by Stat-3 signaling in tumor cells. *Nat Med* 10**:** 48-54.

Watson C.J., Miller W.R. (1995). Elevated levels of members of the STAT family of transcription factors in breast carcinoma nuclear extracts. *Br J Cancer* 71**:** 840-844.

Webb N.J., Bottomley M.J., Watson C.J., Brenchley P.E. (1998). Vascular endothelial growth factor (VEGF) is released from platelets during blood clotting: implications for measurement of circulating VEGF levels in clinical disease. *Clin Sci (Lond)* 94**:** 395-404.

Weigelt B., Geyer F.C., Natrajan R., Lopez-Garcia M.A., Ahmad A.S., Savage K., Kreike B., Reis-Filho J.S. (2010). The molecular underpinning of lobular histological growth pattern: a genome-wide transcriptomic analysis of invasive lobular carcinomas and grade- and molecular subtype-matched invasive ductal carcinomas of no special type. *J Pathol* 220**:** 45-57.

Weigelt B., Reis-Filho J.S. (2009). Histological and molecular types of breast cancer: is there a unifying taxonomy? *Nat Rev Clin Oncol* 6**:** 718-730.

White A.M., Daly D.S., Varnum S.M., Anderson K.K., Bollinger N., Zangar R.C. (2006). ProMAT: protein microarray analysis tool. *Bioinformatics* 22**:** 1278-1279.

Willmarth N.E., Baillo A., Dziubinski M.L., Wilson K., Riese D.J., 2nd, Ethier S.P. (2009). Altered EGFR localization and degradation in human breast cancer cells with an amphiregulin/EGFR autocrine loop. *Cell Signal* 21**:** 212-219.

Wilson T.R., Lee D.Y., Berry L., Shames D.S., Settleman J. (2011). Neuregulin-1-mediated autocrine signaling underlies sensitivity to HER2 kinase inhibitors in a subset of human cancers. *Cancer Cell* 20**:** 158-172.

Wingren C., Ingvarsson J., Dexlin L., Szul D., Borrebaeck C.A. (2007). Design of recombinant antibody microarrays for complex proteome analysis: choice of sample labeling-tag and solid support. *Proteomics* 7**:** 3055-3065.

Wirapati P., Sotiriou C., Kunkel S., Farmer P., Pradervand S., Haibe-Kains B., Desmedt C., Ignatiadis M., Sengstag T., Schutz F., Goldstein D.R., Piccart M., Delorenzi M. (2008). Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res* 10**:** R65.

Witkiewicz A.K., Dasgupta A., Sotgia F., Mercier I., Pestell R.G., Sabel M., Kleer C.G., Brody J.R., Lisanti M.P. (2009). An absence of stromal caveolin-1 expression predicts early tumor recurrence and poor clinical outcome in human breast cancers. *Am J Pathol* 174**:** 2023-2034.

Wolff A.C., Hammond M.E., Schwartz J.N., Hagerty K.L., Allred D.C., Cote R.J., Dowsett M., Fitzgibbons P.L., Hanna W.M., Langer A., McShane L.M., Paik S., Pegram M.D., Perez E.A., Press M.F., Rhodes A., Sturgeon C., Taube S.E., Tubbs R., Vance G.H., van de Vijver M., Wheeler T.M., Hayes D.F. (2007). American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *J Clin Oncol* 25**:** 118-145.

Wulfkuhle J.D., Aquino J.A., Calvert V.S., Fishman D.A., Coukos G., Liotta L.A., Petricoin E.F., 3rd (2003). Signal pathway profiling of ovarian cancer from human tissue specimens using reverse-phase protein microarrays. *Proteomics* 3**:** 2085-2090.

Wulfkuhle J.D., Berg D., Wolff C., Langer R., Tran K., Illi J., Espina V., Pierobon M., Deng J., DeMichele A., Walch A., Bronger H., Becker I., Waldhor C., Hofler H., Esserman L., Liotta L.A., Becker K.F., Petricoin E.F., 3rd (2012). Molecular analysis of HER2 signaling in human breast cancer by functional protein pathway activation mapping. *Clin Cancer Res* 18**:** 6426-6435.

Wulfkuhle J.D., Speer R., Pierobon M., Laird J., Espina V., Deng J., Mammano E., Yang S.X., Swain S.M., Nitti D., Esserman L.J., Belluco C., Liotta L.A., Petricoin E.F., 3rd (2008). Multiplexed cell signaling analysis of human breast cancer applications for personalized therapy. *J Proteome Res* 7**:** 1508-1517.

Xia W., Mullin R.J., Keith B.R., Liu L.H., Ma H., Rusnak D.W., Owens G., Alligood K.J., Spector N.L. (2002). Anti-tumor activity of GW572016: a dual tyrosine kinase inhibitor blocks EGF activation of EGFR/erbB2 and downstream Erk1/2 and AKT pathways. *Oncogene* 21**:** 6255-6263.

Yakes F.M., Chinratanalab W., Ritter C.A., King W., Seelig S., Arteaga C.L. (2002). Herceptin-induced inhibition of phosphatidylinositol-3 kinase and Akt Is required for antibody-mediated effects on p27, cyclin D1, and antitumor action. *Cancer Res* 62**:** 4132-4141.

Yamamoto Y., Kohara K., Tabara Y., Miki T. (2001). Association between carotid arterial remodeling and plasma concentration of circulating hepatocyte growth factor. *J Hypertens* 19**:** 1975-1979.

Yarden Y., Sliwkowski M.X. (2001). Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol* 2**:** 127-137.

Zardawi S.J., O'Toole S.A., Sutherland R.L., Musgrove E.A. (2009). Dysregulation of Hedgehog, Wnt and Notch signalling pathways in breast cancer. *Histol Histopathol* 24**:** 385-398.

Zhang H.H., Ahn J., Lin X., Park C. (2006). Gene selection using support vector machines with non-convex penalty. *Bioinformatics* 22**:** 88-95.

Appendix

Table 11: Summary of target proteins (n = 128) and corresponding gene name information.

target protein	gene name	full gene name	phosphorylation site	project[1]
acetyl-CoA carboxylase	ACACA	acetyl-CoA carboxylase alpha		A
AKT1	AKT1	v-akt murine thymoma viral oncogene homolog 1		A, B
AKT1/2	AKT1/AKT2	v-akt murine thymoma viral oncogene homolog 1/2	S473	A, B
AKT1/2	AKT1/AKT2	v-akt murine thymoma viral oncogene homolog 1/2	T308	A, B
AKT2	AKT2	v-akt murine thymoma viral oncogene homolog 2		A, B
APC	APC	adenomatous polyposis coli		A, B
ATM	ATM	ataxia telangiectasia mutated		A, B
ATR	ATR	ataxia telangiectasia and Rad3 related		A, B
BAX	BAX	BCL2-associated X protein		A, B
βCatenin	CTNNB1	catenin (cadherin-associated protein), beta 1		A, B
βCatenin	CTNNB1	catenin (cadherin-associated protein), beta 1	S33/S37/T41	A, B
BCL-2	BCL2	B-cell CLL/lymphoma 2		A, B
BCL-XL	BCL2L1	BCL2-like 1		A, B
BRCA2	BRCA2	breast cancer 2, early onset		A, B
caveolin-1	CAV1	caveolin 1, caveolae protein		A
CBL	CBL	Cas-Br-M (murine) ecotropic retroviral transforming sequence		A
CDK1	CDK1	cyclin-dependent kinase 1		A, B
CDK2	CDK2	cyclin-dependent kinase 2		A, B
CDK6	CDK6	cyclin-dependent kinase 6		A, B
cJUN	JUN	jun proto-oncogene		A
Claudin 1	CLDN1	claudin 1		A, B
Claudin 3	CLDN3	claudin 3		A, B

[1]Used for biomarker discovery project (A); Used for pathway activation profiling project (B)

target protein	gene name	full gene name	phosphorylation site	project[1]
COL4A3BP	COL4A3BP	collagen, type IV, alpha 3 (Goodpasture antigen) binding protein		A
cRAF	RAF1	v-raf-1 murine leukemia viral oncogene homolog 1	S259	A, B
CREB	CREB1	cAMP responsive element binding protein 1		A
cyclin B1	CCNB1	cyclin B1		A, B
cyclin D1	CCND1	cyclin D1		A, B
cytokeratin 18	KRT18	keratin 18		A
cytokeratin 8	KRT8	keratin 8	S23	A
E-cadherin	CDH1	cadherin 1, type 1, E-cadherin (epithelial)		A, B
EGFR	EGFR	epidermal growth factor receptor		A
EpCAM	EPCAM	epithelial cell adhesion molecule		A, B
HER2	HER2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)		A
HER2	HER2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)	Y1112	A
HER2	HER2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)	Y1248	A
ERBB3	ERBB3	v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)		A
ERBB4	ERBB	v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)		A
ERBB4	ERBB4	v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)	Y1162	A
ERK1	MAPK3	mitogen-activated protein kinase 3		A, B
ERK1/2	MAPK3/MAPK1	mitogen-activated protein kinase 3/1	T202/Y204/T185/Y187	A, B

[1]Used for biomarker discovery project (A); Used for pathway activation profiling project (B)

target protein	gene name	full gene name	phosphorylation site	project[1]
ERα	ESR1	estrogen receptor 1		A
FAK	PTK2	PTK2 protein tyrosine kinase 2	S843	A, B
FIH	HIF1AN	hypoxia inducible factor 1, alpha subunit inhibitor		A, B
FoxO3a	FOXO3	forkhead box O3	S318/S321	A
GATA3	GATA3	GATA binding protein 3		A
GRB2	GRB2	growth factor receptor-bound protein 2		A
GSK3A	GSK3A	glycogen synthase kinase 3 alpha		A, B
GSK3A	GSK3A	glycogen synthase kinase 3 alpha	S21	A, B
GSK3A/B	GSK3A/B	glycogen synthase kinase 3 alpha/beta	Y279/Y216	A, B
GSK3B	GSK3B	glycogen synthase kinase 3 beta		A, B
GSK3B	GSK3B	glycogen synthase kinase 3 beta	S9	A, B
Integrin B1	ITGB1	integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)		A, B
Integrin B3	ITGB3	integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61)		A, B
Ki-67	MKI67	antigen identified by monoclonal antibody Ki-67		A
LAMB1	LAMB1	LAMB1 laminin, beta 1		A
MCL-1	MCL1	myeloid cell leukemia sequence 1		A, B
MEK	MAP2K1	mitogen-activated protein kinase kinase 1		A, B
MEK	MAP2K1	mitogen-activated protein kinase kinase 1	S217/S221	A, B
MET	MET	met proto-oncogene (hepatocyte growth factor receptor)		A
metadherin	MTDH	metadherin		A
MNK1	MKNK1	MAP kinase interacting serine/threonine kinase 1		A, B
mTOR	mTOR	mechanistic target of rapamycin (serine/threonine kinase)		A, B
mTOR	mTOR	mechanistic target of rapamycin (serine/threonine kinase)	S2448	A, B

[1]Used for biomarker discovery project (A); Used for pathway activation profiling project (B)

target protein	gene name	full gene name	phosphorylation site	project[1]
N-cadherin	CDH2	cadherin 2, type 1, N-cadherin		A, B
NDKA	NME1	non-metastatic cells 1, protein (NM23A)		A
NFκB	RELA	v-rel reticuloendotheliosis viral oncogene homolog A (avian)		A, B
NFκB	RELA	v-rel reticuloendotheliosis viral oncogene homolog A (avian)	S536	A, B
NOTCH2	NOTCH2	notch 2		A, B
NOTCH3	NOTCH3	notch 3		A, B
p27	CDKN1B	cyclin-dependent kinase inhibitor 1B (p27, Kip1)		A, B
p38	MAPK14	mitogen-activated protein kinase 14		A, B
p38	MAPK14	mitogen-activated protein kinase 14	T180/Y182	A, B
p53	TP53	tumor protein p53		A, B
p53	TP53	tumor protein p53	S15	A, B
p70S6K	RPS6KB1	ribosomal protein S6 kinase, 70kDa, polypeptide 1		A, B
p70S6K	RPS6KB1	ribosomal protein S6 kinase, 70kDa, polypeptide 1	T389	A, B
p70S6K	RPS6KB1	ribosomal protein S6 kinase, 70kDa, polypeptide 1	T421/S424	A, B
p90RSK	RPS6KA1	ribosomal protein S6 kinase, 90kDa, polypeptide 1	S380	A, B
PAK1	PAK1	p21 protein (Cdc42/Rac)-activated kinase 1		A, B
PAK2	PAK2	p21 protein (Cdc42/Rac)-activated kinase 2		A, B
PARP	PARP1	poly (ADP-ribose) polymerase 1		A, B
P-cadherin	CDH3	cadherin 3, type 1, P-cadherin (placental)		A, B
PCNA	PCNA	proliferating cell nuclear antigen		A
PDI	P4HB	prolyl 4-hydroxylase, beta polypeptide		A
PDK1	PDPK1	3-phosphoinositide dependent protein kinase-1		A, B

[1]Used for biomarker discovery project (A); Used for pathway activation profiling project (B)

target protein	gene name	full gene name	phosphorylation site	project[1]
PDK1	PDPK1	3-phosphoinositide dependent protein kinase-1	S241	A, B
PI3K_p110	PIK3CA	phosphoinositide-3-kinase, catalytic, alpha polypeptide		A, B
PI3K_p85	PIK3R1	phosphoinositide-3-kinase, regulatory subunit 1 (alpha)		A, B
PKA	PRKACA	protein kinase, cAMP-dependent, catalytic, alpha		A
PKCα	PRKCA	protein kinase C, alpha		A, B
PKCα	PRKCA	protein kinase C, alpha	S657/Y658	A, B
PLCγ	PLCG1	phospholipase C, gamma 1		A
PR	PGR	progesterone receptor		A
PRAS40	AKT1S1	AKT1 substrate 1 (proline-rich)		A, B
PRAS40	AKT1S1	AKT1 substrate 1 (proline-rich)	T246	A, B
PTEN	PTEN	phosphatase and tensin homolog		A, B
PTEN	PTEN	phosphatase and tensin homolog	T366/S370	A, B
pyruvate dehydrogenase	PDHA1	pyruvate dehydrogenase (lipoamide) alpha 1		A
RB	RB1	retinoblastoma 1		A, B
RB	RB1	retinoblastoma 1	S807/S811	A, B
RKIP	PEBP1	phosphatidylethanolamine binding protein 1		A, B
ROCK1	ROCK1	Rho-associated, coiled-coil containing protein kinase 1		A, B
ROCK2	ROCK2	Rho-associated, coiled-coil containing protein kinase 2		A, B
RPS6	RPS6	ribosomal protein S6		A, B
RPS6	RPS6	ribosomal protein S6	S235/S236	A, B
RPS6	RPS6	ribosomal protein S6	S240/S244	A, B
RSK	RPS6KA1	ribosomal protein S6 kinase, 90kDa, polypeptide 1		A, B
SDHA	SDHA	succinate dehydrogenase complex, subunit A, flavoprotein (Fp)		A

[1]Used for biomarker discovery project (A); Used for pathway activation profiling project (B)

target protein	gene name	full gene name	phosphorylation site	project[1]
SHP1	PTPN6	protein tyrosine phosphatase, non-receptor type 6		A
SHP2	PTPN11	protein tyrosine phosphatase, non-receptor type 11		A
SMAD2	SMAD2	SMAD family member 2		A, B
SMAD7	SMAD7	SMAD family member 7		A, B
SMURF2	SMURF2	SMAD specific E3 ubiquitin protein ligase 2		A, B
SRC	SRC	v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian)		A, B
SRC	SRC	v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian)	Y416	A, B
STARD10	STARD10	StAR-related lipid transfer (START) domain containing 10		A
STAT1	STAT1	signal transducer and activator of transcription 1		A, B
STAT1	STAT1	signal transducer and activator of transcription 1	Y701	A, B
STAT3	STAT3	signal transducer and activator of transcription 3 (acute-phase response factor)		A, B
STAT3	STAT3	signal transducer and activator of transcription 3 (acute-phase response factor)	Y705	A, B
STAT5	STAT5B	signal transducer and activator of transcription 5B	Y694/Y699	A, B
TIE2	TEK	TEK tyrosine kinase, endothelial		A
TOP2A	TOP2A	topoisomerase (DNA) II alpha 170kDa		A
TSC1	TSC1	tuberous sclerosis 1		A, B
TSC2	TSC2	tuberous sclerosis 2		A, B
TSC2	TSC2	tuberous sclerosis 2	T1462	A, B
VEGFR2	KDR	KDR kinase insert domain receptor (a type III receptor tyrosine kinase)		A
vimentin	VIM	vimentin		A

[1]Used for biomarker discovery project (A); Used for pathway activation profiling project (B)

Table 12: Patient characteristics of the subgroup (n = 36) used for biomarker identification. Median age was 62 years (range 39-85 years).

characteristic	number of patients	%
pT category (UICC 2009)		
pT1	19	53
pT2	15	42
pT3	0	0
pT4	2	6
lymph node status[1]		
positive	9	25
negative	26	72
histologic grade		
1	14	39
2	0	0
3	22	61
ERα status[2]		
positive	36	100
negative	0	0
PR status[2]		
positive	34	94
negative	2	6
HER2 status[3]		
positive	1	3
negative	35	97

[1]Number does not add up to 36 due to a case with missing data; [2]positive: IRS 3-12, negative: IRS 0-2 (Remmele and Stegner 1987); [3]positive: IHC 3+ or average HER2 gene copy number > 6, negative: IHC 0/1+ or average HER2 gene copy number < 4 (Wolff et al. 2007).

Table 13: Patient characteristics of the ERα-positive subgroup (n = 109). Median age was 64 years (range 31-86 years).

characteristic	number of patients	%
pT category (UICC 2009)		
pT1	44	40
pT2	55	50
pT3	6	6
pT4	4	4
lymph node status[1]		
positive	38	35
negative	70	64
histologic grade		
1	14	13
2	73	67
3	22	20
ERα status[2]		
positive	109	100
negative	0	0
PR status[2]		
positive	102	94
negative	7	6
HER2 status[3]		
positive	4	4
negative	105	96

[1]Number does not add up to 109 due to a case with missing data; [2]positive: IRS 3-12, negative: IRS 0-2 (Remmele and Stegner 1987); [3]positive: IHC 3+ or average HER2 gene copy number > 6, negative: IHC 0/1+ or average HER2 gene copy number < 4 (Wolff et al. 2007).

Table 14: List of target proteins differentially expressed between the RPPA_orange and the RPPA_purple group. The cut-off to define differential expression was chosen as p-value < 0.001 (Wilcoxon rank sum test, not adjusted for multiple testing).

target protein	p-value	higher expressed in group
pAKT (S473)	1.19^{e-9}	orange
pAKT (T308)	1.47^{e-5}	orange
ATR	1.37^{e-10}	orange
pbCatenin (S33/S37/T41)	3.83^{e-7}	orange
BCL-XL	2.66^{e-6}	orange
CDK1	1.39^{e-13}	orange
CDK6	1.76^{e-6}	orange
Claudin 1	1.17^{e-4}	orange
pcRAF (S259)	2.29^{e-5}	purple
Cyclin D1	1.59^{e-6}	orange
pFAK (S843)	9.23^{e-11}	orange
FIH	1.90^{e-7}	orange
GSK3A	7.74^{e-6}	purple
pGSK3A (S21)	8.49^{e-10}	orange
GSK3B	9.27^{e-5}	purple
pGSK3B (S9)	9.07^{e-9}	orange
Integrin B1	2.29^{e-9}	orange
MNK1	1.17^{e-4}	orange
mTOR	4.18^{e-8}	purple
N-cadherin	2.29^{e-12}	orange
NOTCH2	$4.88e^{-5}$	orange
NOTCH3	2.99^{e-12}	orange
p70S6K	3.22^{e-5}	purple
pp70S6K (T389)	4.04^{e-11}	orange
pp70S6K (T421/S424)	3.51^{e-4}	purple
PAK1	2.09^{e-5}	purple
PAK2	8.57^{e-5}	orange
pPDK1 (S241)	2.41^{e-7}	purple
PI3K_p85	1.32^{e-11}	orange
pPRAS40 (T246)	1.12^{e-5}	orange

target protein	p-value	higher expressed in group
RB	1.34^{e-5}	orange
RKIP	6.23^{e-5}	purple
ROCK1	1.19^{e-9}	orange
RPS6	1.00^{e-4}	purple
SMAD2	3.66^{e-8}	purple
SMAD7	1.19^{e-9}	orange
SMURF2	1.13^{e-10}	orange
SRC	9.35^{e-7}	orange
STAT1	1.04^{e-6}	orange
pSTAT1 (Y701)	6.19^{e-8}	orange
STAT3	3.22^{e-5}	purple
pSTAT3 (Y705)	1.04^{e-6}	orange
pSTAT5 (Y694/Y699)	8.49^{e-10}	orange
pTSC2 (T1462)	2.96^{e-5}	orange

Table 15: Patient characteristics of the subgroup (n = 82) used for measurement of growth factor concentrations in tumor samples. Median age was 64 years (range 37-86 years).

characteristic	number of patients	%
pT category (UICC 2009)		
pT1	31	38
pT2	42	51
pT3	5	6
pT4	4	5
lymph node status[1]		
positive	30	37
negative	51	62
histologic grade		
1	8	10
2	53	65
3	21	25
ERα status[2]		
positive	82	100
negative	0	0
PR status[2]		
positive	77	94
negative	5	6
HER2 status[3]		
positive	3	4
negative	79	96
menopause status[1]		
pre	16	20
post	63	77

[1]Number does not add up to 82 due to cases with missing data; [2]positive: IRS 3-12, negative: IRS 0-2 (Remmele and Stegner 1987); [3]positive: IHC 3+ or average HER2 gene copy number > 6, negative: IHC 0/1+ or average HER2 gene copy number < 4 (Wolff et al. 2007).

Table 16: Samples removed from further analysis of growth factor concentrations in plasma samples (n = 30).

patient ID	cause of removal form sample set
ID_60, 67, 88, 160, 166, 175, 177, 202	no surgery
ID_182, 232, 260, 299	recurrent/metastatic breast cancer
ID_212	Angiosarcoma
ID_198	male
ID_247	Melanoma
ID_64, 75, 236	DCIS
ID_180, 205	histology not further specified
ID_12, 33, 77, 119, 240, 252, 266, 282, 318, 328	multicentric/bilateral tumors with different characteristics

Table 17: Patient characteristics of the subgroup (n = 196) used for measurement of growth factor concentrations in plasma samples. Median age was 62 years (range 31-86 years).

characteristic	number of patients	%
pT category (UICC 2009)		
pT1	93	47
pT2	93	47
pT3	5	3
pT4	5	3
lymph node status[1]		
positive	71	36
negative	124	63
histologic grade		
1	22	11
2	123	63
3	51	26
ERα status[2]		
positive	168	86
negative	28	14
PR status[2]		
positive	161	82
negative	35	18
HER2 status[3]		
positive	12	6
negative	184	94
Ki-67 status[1,4]		
positive	95	48
negative	96	49
menopause status[1]		
pre	47	24
post	138	70

[1]Number does not add up to 196 due to cases with missing data; [2]positive: IRS 3-12, negative: IRS 0-2 (Remmele and Stegner 1987); [3]positive: IHC 3+ or average HER2 gene copy number > 6, negative: IHC 0/1+ or average HER2 gene copy number < 4 (Wolff et al. 2007); [4]positive: ≥ 14% nuclear staining of Ki-67, negative < 14% nuclear staining of Ki-67 (Cheang et al. 2009).

Table 18: Patient characteristics of the subgroup (n = 168) used for measurement of growth factor concentrations in plasma samples. Median age was 62 years (range 31-86 years).

characteristic	number of patients	%
pT category (UICC 2009)		
pT1	80	48
pT2	80	48
pT3	4	2
pT4	4	2
lymph node status[1]		
positive	65	39
negative	102	61
histologic grade		
1	22	13
2	117	70
3	29	17
ERα status[2]		
positive	168	100
negative	0	0
PR status[2]		
positive	157	93
negative	11	7
HER2 status[3]		
positive	9	5
negative	159	95
menopause status[1]		
pre	39	23
post	119	71

[1]Number does not add up to 168 due to cases with missing data; [2]positive: IRS 3-12, negative: IRS 0-2 (Remmele and Stegner 1987); [3]positive: IHC 3+ or average HER2 gene copy number > 6, negative: IHC 0/1+ or average HER2 gene copy number < 4 (Wolff et al. 2007).

Acknowledgements

The time as PhD student was an awesome adventure and this dissertation would not have been possible without the contribution of several individuals.

Foremost, I thank Roland Kontermann for accepting me as external PhD student and his valuable feedback during my TAC meetings and the IZI retreats.

Furthermore, I thank Stefan Wiemann for the opportunity to pursue my PhD thesis at the Division of Molecular Genome Analysis in Heidelberg and for his constant support and interest in the project.

I would like to express my gratitude to Ulrike Korf for her supervision and encouragement, but also for her trust to give me the freedom to develop and pursue my own ideas. I admire her “never give up” attitude.

I would like to thank Gerhard Moldenhauer and Andreas Schneeweiss for participation in my TAC meetings and valuable discussions.

I thank the Helmholtz International Graduate School for Cancer Research for offering an excellent advanced training program.

A big thank to Sabrina Schumacher, Corinna Becki, Daniela Heiss, Maike Wosch, and Lisa Schäfer for their constant support and technical assistance in the lab, especially for the “thousands” of Western blots and much more.

For valuable discussions about protein microarrays and beyond, I thank all colleagues of the Quantitative Proteomics Group: Heiko Mannsperger, Christian Schmidt, Frank Götschel, Frauke Henjes, Daniela Berg, and Ramesh Ummanni. Thanks for sharing my enthusiasm about RPPA.

I am grateful to Christian Bender, Tim Beißbarth, Silvia von der Heyde, Manuel Nietert, and Anika Jöcker. Their motivation and bioinformatics support added considerably to this dissertation.

I express my gratitude to all former and current MGA´s for their various forms of support leading to a motivating and enjoyable atmosphere. Especially I would like to mention Sabine Klauck, Ewald Münstermann, and Daniela Fischer.

Furthermore, I want to thank Andreas Schneeweiss, Monika Fischer, Christian Lange, Jörg Heil, Hans-Peter Sinn, Sebastian Aulmann, Steve Wagner, Dorit Arlt as well as Barbara Burwinkel and her team. Without their enormous effort, the “GENOME” sample collection would not have been possible. I also express my sincere gratitude to all patients who agreed to participate in this study.

Many thanks to Heiko, Christian, Rupi, Jan, Frauke, Frank, Mark, Maria, Stephan, Ola, and Kirti, for the great time we spend in and outside the lab, at conferences or retreats, and of course at plenty of unforgettable lunch and coffee breaks with scientific and non-scientific discussions.

Most importantly, I thank my family for their love, support, and encouragement during my entire life, especially during my PhD time in Heidelberg!

Erklärung

Ich erkläre hiermit, dass ich die vorgelegte Dissertation selbst verfasst und mich keinen anderen als den von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe. Des weiteren erkläre ich, dass ich an keiner anderen Stelle ein Prüfungsverfahren beantragt oder die Dissertation in dieser oder einer anderen Form bereits anderweitig als Prüfungsarbeit verwendet habe.

Heidelberg, den 15.03.2013

Johanna Sonntag