PÄIVIKKI KAURANIEMI

Molecular Consequences of Gene Amplification in Breast Cancer

■

ACADEMIC DISSERTATION

To be presented, with the permission of
the Faculty of Medicine of the University of Tampere,
for public discussion in the lecture room of Finn-Medi 5,
Biokatu 12, Tampere, on March 5th, 2004, at 12 o'clock.

Acta Universitatis Tamperensis 990

ACADEMIC DISSERTATION

University of Tampere, Institute of Medical Technology
Tampere University Hospital, Department of Clinical Chemistry
Finland

Supervised by
Docent Anne Kallioniemi
University of Tampere

Reviewed by
Docent Päivi Peltomäki
University of Helsinki
Professor Åke Borg
University of Lund

Distribution

University of Tampere
Bookshop TAJU
P.O. Box 617
33014 University of Tampere
Finland

Tel. +358 3 215 6055
Fax +358 3 215 7685
taju@uta.fi
http://granum.uta.fi

Cover design by
Juha Siro

Printed dissertation
Acta Universitatis Tamperensis 990
ISBN 951-44-5899-0
ISSN 1455-1616

Electronic dissertation
Acta Electronica Universitatis Tamperensis 322
ISBN 951-44-5900-8
ISSN 1456-954X
http://acta.uta.fi

Tampereen yliopistopaino Oy Juvenes Print
Tampere 2004

Geenimonistuman vaikutukset rintasyövässä

Tässä tutkimuksessa selvitettiin kromosomialueella 17q12 sijaitsevan ERBB2-monistuma-alueen geenien kopiolukuja ja ilmentymistasoja cDNA –mikrosirutekniikan avulla. Tätä tarkoitusta varten koostimme 217 kloonia sisältävän cDNA -mikrosirun, jota käytettiin sekä geenien kopiolukujen että ilmentymistasojen määrittämiseen seitsemässä rintasyöpäsolulinjassa. Vertaamalla geenikopiolukuja geenien ilmentymistasoihin löysimme pienen joukon geenejä, joiden ilmentymistaso oli johdonmukaisesti kasvanut geenimonistuman seurauksena.

Seuraavaksi 17q12 monistuma-alueen rakenne määritettiin käyttämällä fluoresenssi in situ hybridisaatio (FISH) –tekniikkaa ja koettimina kookkaan insertin omaavia genomisia BAC klooneja. Tämän kartoituksen avulla pystyimme määrittämään ns. minimaalisen monistuma-alueen, joka rajoittui yhden BAC-kloonin alueelle sekä rintasyöpäsolulinjoissa että primaari rintakasvaimissa. Tämä 280 kiloemäksen suuruinen minimaalinen monistuma-alue sisältää kymmenen transkriptiä, joista kahdeksan vastaa tunnettuja geenejä ja kaksi vielä tuntemattomia hypoteettisiä proteiineja koodaavia geenejä. Geenimonistuman vaikutuksia geenien ilmentymistasoihin primaareissa rintakasvaimissa tutkittiin kvantitatiivisella RT-PCR menetelmällä. Ilmentymisanalyysin perusteella tunnistimme kuusi geeniä (ERBB2, GRB7, PNMT, MLN64, MGC9753, MGC14832), joilla ilmentymistasojen kasvu oli tilastollisesti yhteydessä geenimonistumaan.

Herceptiini –lääkekäsittelyn vaikutuksia geenien ilmentymistasoihin tarkasteltiin ERBB2-monistuneissa ja ei-monistuneissa rintasyöpäsolulinjoissa genominlaajuista cDNA mikrosiruanalyysiä käyttäen. Herceptiini-käsittely aikaansai annosriippuvaisen solujen kasvun hidastumisen ERBB2-monistuneissa soluissa, kun taas ei-monistuneissa soluissa vaste jäi minimaaliseksi. Monivaiheisen geenien valinta-algoritmin avulla pystyimme tunnistamaan 439 geeniä, joiden ilmenemismuutokset Herceptin-käsittelyn jälkeen parhaiten erottivat ERBB2-monistuneet ja ei-monistuneet solulinjat toisistaan. Näiden geenien erottelukyky varmistettiin vielä hierarkisella klusteroinnilla ja self-organizing map analyysillä. ERBB2-monistuneissa solulinjoissa Herceptiini-käsittely aikaansai useiden RNA:n prosessointiin ja DNA:n korjaukseen osallistuvien geenien ilmentymisen kasvun, kun taas soluadheesioon

osallistuvien ja joidenkin tunnettujen onkogeenien kuten c-FOS ja c-KIT geenien, ilmentyminen laski lääkkeen vaikutuksesta.

Geenimomonistuman genominlaajuisia vaikutuksia geenien ilmentymistasoihin tutkittiin 14 rintasyöpäsolulinjassa käyttämällä 13,824 kloonista koostuvaa cDNA-mikrosirua. Korkean resoluution omaavan kopiolukuanalyysin avulla näistä solulinjoista tunnistettiin yhteensä 24 itsenäistä amplikonia, joiden koko vaihteli 0.2-12 megaemäsparin välillä. Tulokset osoittivat, että sekä matala- että korkea-asteisella geenien kopioluvun kasvulla oli merkittävä vaikutus geenien ilmentymistasoihin kaikkialla genomissa. Kaikkein korkeimmin monistuneista geeneistä 44 %:lla myös ilmentymistaso oli noussut, kun taas korkeimmin ilmentyneistä geeneistä 10.5 % oli myös monistuneita. Satunnaisen permutaatioanalyysin avulla identifioimme 270 geeniä, joiden ilmentyminen oli systemaattisesti kytkeytynyt geenimonistuman kanssa. Näiden joukossa oli useita aiemmin tunnistettuja monistuman kohdegeenejä, mutta myös useita geenejä, joiden toiminnan ei ole aiemmin osoitettu liittyvän rintasyöpään.

CONTENTS

LIST OF ORIGINAL COMMUNICATIONS

This thesis is based on the following communications, referred in the text by their Roman numerals:

I. Kauraniemi P, Bärlund M, Monni O and Kallioniemi A (2001): New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. Cancer Res 61:8235-8240.

II. Kauraniemi P, Kuukasjärvi T, Sauter G and Kallioniemi A (2003): Amplification of a 280 kb core region at the ERBB2 locus leads to activation of two hypothetical proteins in breast cancer. Am J Pathol 163:1979-1984.

III. Kauraniemi P, Hautaniemi S, Autio R, Astola J, Monni O, Elkahloun A and Kallioniemi A (2004): Effects of Herceptin treatment on global gene expression patterns in HER2 amplified and non-amplified breast cancer cell lines. Oncogene 23:1010-1013.

IV. Hyman E*, Kauraniemi P*, Hautaniemi S, Wolf M, Mousses S, Rozenblum E, Ringner M, Sauter G, Monni O, Elkahloun A, Kallioniemi OP and Kallioniemi A (2002): Impact of DNA amplification on gene expression patterns in breast cancer. Cancer Res 62:6240-6245.

* These authors contributed equally

ABBREVIATIONS

ATM	ataxia telangiectasia mutated
AR	amphiregulin
BAC	bacterial artificial chromosome
BFB	breakage-fusion-bridge
BRCA1	breast cancer susceptibility gene 1
BRCA2	breast cancer susceptibility gene 2
BTC	betacellulin
cDNA	complementary DNA
CGH	comparative genomic hybridisation
DM	double minute
EGFR	epidermal growth factor receptor
EPR	epiregulin
ERBB2	avian erythroblastic leukaemia viral oncogene homolog 2
EST	expressed sequence tag
FISH	fluoresecence *in situ* hybridization
FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog
GRB7	growth factor receptor-bound protein 7
HB-EGF	heparin binding epidermal growth factor
HER	human epidermal growth factor receptor like
HMEC	human mammary epithelial cells
HSR	homogenously staining region
IHC	immunohistochemistry
KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
MGC14832	hypothetical protein MGC14832
MGC9753	hypothetical protein MGC9753
MLN64	metastatic lymph node 64
NEUROD2	neurogenic differentiation 2
NRG	neuregulin
P1	P1 artificial chromosome
p53	protein 53
PAC	P1 derived artificial chromosome
PCR	polymerase chain reaction
PNMT	phenylethanolamine N-methyltransferase
PPP1R1B	protein phosphatase 1, regulatory (inhibitor) subunit 1B
RB	retinoblastoma
TCAP	titin-cap
TGFα	transforming growth factor-α
ZNFN1A3	zinc finger protein, subfamily 1A, 3 (Aiolos)

ABSTRACT

cDNA microarray technology was applied for a systematic survey of gene copy numbers and gene expression levels at the ERBB2 amplicon at 17q12. For this purpose a chromosome 17-specific cDNA microarray containing 217 clones from this region was constructed and used for parallel analysis of gene copy numbers and gene expression levels in seven breast cancer cell lines. Direct comparison of data from these two array-based analyses allowed identification of a small set of transcripts showing a consistent pattern of increased copy number and overexpression in the 17q12 amplified cell lines.

Detailed characterization of the structure and extent of the 17q12 amplicon was performed using large insert size BAC clones and interphase fluorescence *in situ* hybridization (FISH). The amplicon mapping revealed a minimal common region of amplification that was restricted to a single BAC clone both in breast cancer cell lines and in primary breast tumors. This 280 kb minimal region contains ten transcripts, eight representing known genes and two encoding uncharacterised hypothetical proteins. The molecular consequences of amplification in primary breast tumors were determined using real time RT-PCR. Expression analysis identified six genes (ERBB2, GRB7, PNMT, MLN64, MGC9753, MGC14832) showing statistically highly significant correlation between gene copy number and expression levels.

Global effects of Herceptin treatment on gene expression levels in ERBB2 amplified and non-amplified cell lines were explored using cDNA microarrays. In growth inhibition assays, Herceptin treatment induced a dose-dependent growth reduction in ERBB2 amplified breast cancer cell lines whereas non-amplified cell lines showed only a minimal effect. Using a step-wise gene selection algorithm a set of 439 genes that best separated the amplified and non-amplified cell lines from each other was identified. The discriminatory power of these genes was validated by both hierarchical clustering and self-organizing map analyses. In the ERBB2 amplified cell lines, Herceptin treatment induced the expression of several genes involved in RNA processing and DNA repair, while expression of cell adhesion mediators and known oncogenes such as c-FOS and c-KIT was downregulated.

To explore the molecular consequences of gene amplification on a global level, a genome-wide gene copy number and expression analysis was performed in 14 breast cancer cell lines

using a cDNA microarray containing 13,824 clones. High-resolution CGH microarray analysis identified 24 independent amplicons ranging in size from 0.2 to 12 Mb. Both low- and high-level copy number increases had a considerable effect on gene expression levels throughout the genome. Of the most highly amplified genes, 44 % showed overexpression whereas 10.5 % of the highly overexpressed genes were amplified. Statistical analysis with random permutation tests revealed a set of 270 genes whose expression levels were systematically attributable to gene amplification status. These included many previously described amplification target genes as well as several genes not previously shown to have a role in breast cancer.

INTRODUCTION

The most common mechanism of oncogene activation in breast cancer and in other solid tumors is gene amplification. Cytogenetically recognized characteristics of gene amplification include double minute chromosomes (DM) and homogenously staining regions (HSR). Double minutes appear as paired, small, spherical, extrachromosomal structures lacking centromeres. Homogenously staining regions are chromosomal segments that lack the typical banding pattern seen in Giemsa staining (reviewed by Schwab, 1999). The size and complexity of the amplified region typically varies from one chromosomal location and tumor type to another. Usually the amplified segment is much larger than the transcription unit of a particular gene and may range from less than 100 kb to several megabases (reviewed by Savelyeva and Schwab, 2001). Recently several small amplicons have been identified in various cancers (Collins et al., 1998; Lin et al., 2000; Huang et al., 2002; Lin et al. 2002). However, these findings probably reflect improved methodologies and the availability of highly accurate and detailed mapping information through the Human Genome Project. Traditionally it has been thought that only a single gene is selected for in a given amplicon. However, recent studies of e.g. 11q13, 17q23 and 20q13 amplicons have shown that amplification and altered expression of multiple genes included in an amplicon probably contribute to tumor progeression (Hui et al., 1997; Bekri et al., 1997; Wu et al., 2001; Bärlund et al., 2000; Anzick et al., 1997; Sen et al., 1997, Collins et al., 1998). The purpose of this study was to evaluate the molecular consequences of gene amplification in breast cancer across the entire genome with special emphasis on the ERBB2 amplicon.

REVIEW OF THE LITERATURE

1. Breast cancer

For the last two decades breast cancer has been the most common malignancy in Finland and other industrial countries and it has been estimated to affect every tenth woman at some point of her life (Pukkala et al., 1997). In Finland, 3,658 new cases of breast cancer were diagnosed in 2001. Preliminary predictions of new breast cancer cases show a steady increase, 3,746 cases being predicted for 2003 (Finnish Cancer Registry. Cancer Statistics at www.cancerregistry.fi, last updated on 13 Aug 2003).

Breast cancer arises from the epithelial cells lining the terminal duct lobular unit and can be classified as either invasive or non-invasive (in situ) carcinoma. In non-invasive breast carcinoma cells remain within the basement membrane whereas in invasive type cancer cells invade the basement membrane of the ducts and lobules. Invasive breast cancer is divided into two main categories, ductal carcinoma consisting of 80% and lobular carcinoma consisting of 10-15 % of cases (reviewed by Sainsbury et al., 2000). Some tumors show a distinct pattern of growth and cellular morphology and are classified based on these specific features, e.g. as tubular, mucinous, medullary, papillary, and cribriform.

Most breast cancers (90-95 %) are sporadic. Several risk factors for breast cancer have been identified in epidemiological studies (reviewed by Fentiman, 2001). These include old age, exogenous estrogens, early age of menarche, late age at first birth and menopause, nulliparity, obesity, residence in western countries, and exposure to radiation. However, the main risk factor is a positive family history of breast cancer. Hereditary breast cancer is determined by a finding of breast cancer in two or more first-degree relatives, early age of onset, bilaterality, and presence of other epithelial cancers (e.g. ovarian cancer) in addition to breast cancer in the family (reviewed by McPherson et al., 2000). Hereditary breast cancer accounts for 5-10 % of breast cancer cases and predisposition is mostly due to germ-line mutations in the two breast cancer susceptibility genes, BRCA1 at 17q11-q21 (Miki et al., 1994) and BRCA2 at 13q12-q13 (Wooster et al., 1995).

2. Basic aspects of cancer genetics

Cancer evolves from a single somatic cell acquiring a series of genetic and epigenetic changes conferring a growth advantage for the cell and leading to progressive conversion of normal cells into cancer cells (reviewed by Ponder, 2001). The strongest evidence for the multistep nature of cancer has been provided by studies of colorectal cancer, where accumulation of specific mutations have been shown to correspond to defined histopathological stages in the progression of the disease (reviewed by Fearon and Vogelstein, 1990). Hanahan and Weinberg (2000) have suggested that the human cancer cell needs to acquire six essential genetic alterations in the cell physiology, including 1) self-sufficiency in growth signals, 2) insensitivity to anti-growth signals, 3) evasion of programmed cell death (apotosis), 4) gain of unlimited growth potential, 5) sustained angiogenesis, and 6) acquired potential for tissue invasion and metastasis, in order to achieve malignant growth.

2.1 Oncogenes and tumor suppressor genes

Genetic changes in the progression of cancer typically affect two different types of genes, tumor suppressor genes and oncogenes. Cell fusion experiments have indicated that the transformed phenotype can often be corrected *in vitro* by fusion with a normal cell, providing evidence that tumorigenesis involves recessive loss of function mutations in certain genes Harris et al., 1969). In normal cells, these so-called tumor suppressor genes control important cellular processes, such as cell cycle (e.g. p53, RB), proliferation, DNA repair (e.g. BRCA1, BRCA2), and apoptosis (e.g. ATM, p53) (reviewed by Teh et al., 1999). Tumor suppressor genes are inactivated by loss-of-function mutations during tumorigenesis. According to the Knudson “two-hit” hypothesis, two hits inactivating both alleles of a tumor suppressor are required for tumor formation (Knudson, 1971; reviewed by Knudson, 2001). These inactivating hits may be acquired through point mutation, deletion, methylation, or gene conversion. However, it has recently been suggested that in some circumstances loss of a single tumor suppressor gene allele can be sufficient to lead to tumor progression. This mechanism is called functional haploinsufficiency (reviewed by Balmain et al., 2003).

In contrast to tumor suppressor genes, oncogenes are activated during tumor development. Proto-oncogenes, the normal cellular counterparts of the activated oncogenes, include secreted growth factors (e.g. PDGF), cell surface receptors (e.g. ERBB family), components

of the intracellular signaling pathways (e.g. RAS), nuclear transcription factors (e.g. MYC), and regulators of cell cycle (e.g. CDKs, cyclins). Activation of oncogenes involves a gain of function mutation, which may be quantitative (increased production of unaltered gene product) or qualitative (production of altered gene product). Several mechanisms lead to oncogene activation and include gene amplification (e.g. ERBB2), point mutation (e.g. RAS), and chromosomal rearrangement that can either result in formation of novel fusion proteins (e.g. ABL-BCR) or transposition of a growth promoting gene to an active chromatin domain (e.g. MYC) (reviewed by Todd and Wong, 1999).

2.2 Gene amplification in cancer

Gene amplification has been shown to occur in response to exposure to cytotoxic drugs and during tumorigenesis. In both cases gene amplification leads to enhanced expression of the gene product. The mechanisms of gene amplification are still largely unknown, although several different models have been suggested. These include e.g. re-replication, unequal exchange, episome excision, and breakage-fusion-bridge cycle models (Wahl 1989; Windle and Wahl, 1992; Stark, 1993). In mammalian cells gene amplification by breakage-fusion-bridge (BFB) cycle is best documented, although factors driving this mechanism are not clear. Initial event in the BFB cycle is double strand breakage occurring within chromosome common fragile sites (Coquelle et al., 1997; Kuo et al., 1998; Coquelle et al., 2002; Hellman et al., 2002). Chromosomal breakage is followed by fusion of the broken species and formation of the dicentric chromosome. Dicentric chromosome then forms a bridge between the opposite poles of the dividing cell. In cell division a break of this chromosome leads to unequal distribution of the genetic material. In successive cell generations, recurrent cycles of BFB result futher accumulation of amplicon copies (Hellman et al., 2002). Gene amplification has been shown to occur commonly in several different malignancies including breast, prostate, lung, ovarian, gastric, pancreatic and colon cancers (Knuutila et al., 1998). First amplified oncogene shown to have clinical significance was MYCN, due to its association with poor prognosis of the patient in neuroblastoma (reviewed by Schwab, 1999).

2.3 Epigenetic changes in cancer

Epigenetic changes, including DNA methylation and modifications of histones, function as an additional mechanism of gene regulation contributing to malignant transformation of the cell (reviewed by Plass, 2002). Epigenetic modifications of the DNA are hereditary, but do not alter the actual DNA sequence. In DNA methylation, a methyl group is added to the 5´ carbon of the cytosine base, typically found as long CG-rich segments concentrated in CpG islands and repeat sequences present in satellite and centromere regions. CpG islands are located in 5´ promoter regions of housekeeping genes and some tissue specifically expressed genes. Most of the CpG islands are unmethylated in normal cells, although other GC-rich sequences (e.g. repetitive sequences) are highly methylated in normal cells. Global hypomethylation has been reported to occur in several human malignancies and it has been suggested to cause genomic instability (reviewed by Jones and Baylin, 2002). Hypermethylation of the promoter CpG islands on the other hand serves as a mechanism to silence tumor suppressor gene activity in cancer cells (reviewed by Plass, 2002; Jones and Baylin, 2002; Garinis et al., 2002; Nephew and Huang, 2003). In addition to DNA methylation, histone modifications have a role in the regulation of gene expression (Rice and Allis, 2001). The acetylation status of histones alters chromatin structure, which in turn regulates gene expression. Deacetylated histones lead to condensation of the chromatin limiting the access of transcription factors in promoter regions of the genes whereas acetylation of the histones relaxes the chromatin structure allowing transcription to occur. The acetylation status of the histones is regulated by two families of proteins, histone acetyltransferases (HAT) and histone deacetylases (HDAC) (reviewed by Marks et al., 2001; Lehrmann et al., 2002).

3. ERBB2 oncogene in breast cancer

3.1 ERBB family of receptors

Amplification and subsequent overexpression of the ERBB2 oncogene is one of the most clinically important genetic aberrations in breast cancer (reviewed by Ross and Fletcher, 1999). The ERBB2 (also known as neu and HER2) gene is located at chromosome 17q12 and encodes a 185 kDa transmembrane glycoprotein that belongs to a family of epidermal growth

factor receptors. This family contains four homologous receptors: the epidermal growth factor receptor EGFR (ERBB1/HER1), ERBB2 (HER2/neu), ERBB3 (HER3), and ERBB4 (HER4) (reviewed by Stern, 2000). All these receptors are composed of a cysteine-rich extracellular ligand binding domain, a lipophilic transmembrane domain, and an intracellular tyrosine kinase domain with a carboxy terminal autophosphorylation segment (reviewed by Rubin and Yarden, 2001). Ligand binding to the extracellular domain induces receptor dimerization and formation of either homo- (e.g. HER1-HER1) or heterodimers (e.g. HER1-HER2) (reviewed by Järvinen and Liu, 2002). Receptor dimerization leads to activation of intrinsic protein tyrosine kinase activity and tyrosine autophosphorylation, which triggers a complex signaling cascade resulting in signal transmission across the cell membrane and intracellular space to nucleus, where activation of the target genes occurs (reviewed by Järvinen and Liu, 2002; Rubin and Yarden, 2001).

Several ligands binding to ERBB receptors have been identified (Figure 1). These include EGF-like ligands activating the EGFR receptor (e.g. EGF, TGFα, amphiregulin, heparin-binding EGF-like growth factor, betacellulin, and epiregulin) and neuregulins binding to ERBB3 and ERBB4 receptors (reviewed by Pinkas-Kramarski et al., 1997; Riese and Stern, 1998; Hynes et al., 2001). Interestingly, no ligand specific for ERBB2 has so far been identified. ERBB ligands are bivalent molecules containing two binding sites, high-affinity narrow-specificity site (N-terminal) and low-affinity broad-specificity site (C-terminal), for ERBB receptors (Tzahar et al., 1997). Although ERBB2 receptor does not function as a high affinity binding site for any of the ligands, it has been shown to be a preferable interaction partner for all other ERBB receptors through the low-affinity binding site of the ligand (Tzahar et al., 1996; Graus-Porta et al., 1997). Heterodimers containing ERBB2 have been shown to contain particularly high signaling potency compared to hetero- and homodimers without ERBB2. This high signaling potency of the ERBB2 containing heterodimers is explained by several different factors (reviewed by Harari and Yarden, 2000; Klapper et al., 2000). First of all, ERBB2 heterodimers are characterized by reduced rate of ligand dissociation from the high affinity receptor (Karunagaran et al., 1996; Alroy and Yarden, 1997; Worthylake et al., 1999). ERBB2 also induces lateral signaling by recruiting and activating other ERBB receptors and has a slower rate of receptor internalisation compared with other ERRB receptors (Graus-Porta et al., 1997; Baulida et al., 1996; Pinkas-Kramarski et al., 1996; Brennan et al., 2000). As a result, ERBB2 containing heterodimers are long-lived leading to prolonged signaling and enhanced activation of downstream signaling pathways.

The variety of homo- and heterodimer combinations of ERBB family members provides great signal diversity because different signaling pathways are activated depending on the dimerisation partners, ligands and signaling proteins involved (Lemmon and Schlessinger, 1994; Beerli and Hynes, 1996; Alroy and Yarden, 1997; Olayioye et al., 1998; Hynes et al., 2001). ERBB signaling networks can be envisaged to consist of three different layers. Signal transduction begins from the input layer, which contains the ligands and their receptors. In the signal processing layer, the identity of the ligand, the heterodimer composition, and the nature of intracellular receptor associated signaling proteins determine the type of signal transmitted to the nucleus. The output layer includes various cellular effects of ERBB signals that involve the regulation of cell growth, division, differentiation, adhesion, motility, and apoptosis (reviewed by Alroy and Yarden, 1997; Yarden and Sliwkowski, 2001). The main downstream signalling pathways activated by ERBB receptors include the mitogen-activated protein kinase (MAPK), phosphatidylinositol 3-kinase (PI 3-K), the Janus kinase, and the phospholipase C gamma (PLC-γ) pathways (reviewed by Olayioye et al., 2000; Prenzel et al., 2001; Yarden and Sliwkowski, 2001).

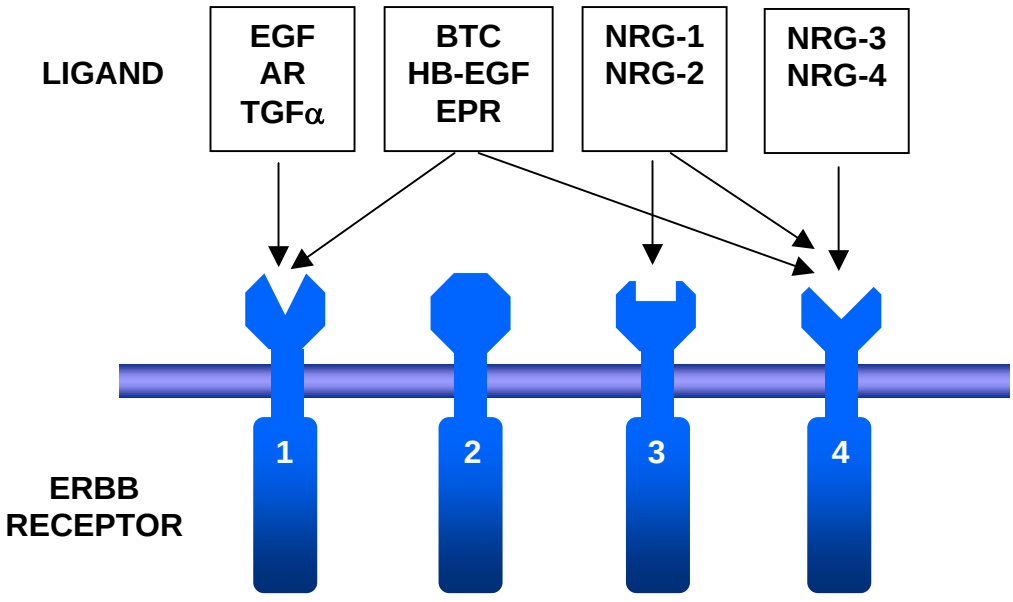

Figure 1. Binding specificities of the peptide growth factors to ERBB family receptors. Epidermal growth factor (EGF), amphiregulin (AR), transforming growth factor-α, betacellulin (BTC), heparin-binding EGF (HB-EGF), epiregulin (EPR), and neuregulin (NRG).

3.2 Clinical relevance of the ERBB2 oncogene in breast cancer

3.2.1 ERBB2 as a prognostic and predictive factor

The ERBB2 oncogene has been shown to be amplified and overexpressed in 10-34% of breast cancers (reviewed by Ross and Fletcher, 1999). The association between ERBB2 amplification and poor prognosis of the patients was first demonstrated in the study by Slamon and co-workers in 1987, where they showed that amplification of ERBB2 is a significant predictor of both overall survival and time to relapse in patients with breast cancer (Slamon et al., 1987). Since then several studies assessing the relationship between ERBB2 abnormalities and breast cancer outcome have been published. A recent meta-analysis summarizing 81 published studies with 27,161 patients revealed that in 73 (90 %) studies either ERBB2 gene amplification or protein overexpression correlated with poor outcome of the patients in either univariate or multivariate analysis (Ross et al., 2003). Only eight of the studies did not show a correlation between ERBB2 status and patient outcome (van de Vijver et al., 1988; Heintz et al, 1990, Clark and McGuire, 1991; Jacquemier et al., 1994; Rosen et al., 1995; Hieken et al., 1996; Reed et al., 2000; Pawlowski et al., 2000).

In addition to the prognostic significance of ERBB2, several studies have evaluated its role in predicting response to therapy. Most commonly amplification and overexpression of ERBB2 has been associated with resistance to hormone therapy (reviewed by Dowsett, 2001; Muss, 2001; Schmid et al., 2002; Nunes and Harris, 2002). Several studies have shown an association between ERBB2 status and resistance to tamoxifen treatment (Giai et al., 1994; Carlomagno et al., 1996; Newby et al., 1997; Burke et al., 1998; Sjogren et al., 1998; Ferrero-Pous et al., 2000; Jukkola et al., 2001; Pinto et al., 2001). Furthermore, it has been shown that ERBB2-positive patients treated with tamoxifen had worse outcome than untreated patients (De Placido et al., 2003). However, contradictory results indicating no association between positive ERBB2 status and tamoxifen resistance have also been obtained (Elledge et al., 1998; Berry et al., 2000; Knoop et al., 2001; Love et al., 2003).

The relationship of ERBB2 status and adjuvant chemotherapy has also been assessed. The earliest studies concentrated on regimens containing CMF (cyclofosfamide, methotrexate, 5-fluorouracil). The majority of these studies showed that ERBB2-positive tumors are resistant to CMF therapy (Gusterson et al., 1992; Giai et al., 1994; Stål et al., 1995; Berns et al., 1995;

Cooke et al., 2001), although a few studies failed to find such association (Miles et al., 1999; Menard et al., 2001; Cooke et al., 2001). In addition, Muss and co-workers showed that ERBB2-positive patients respond better to doxorubicin treatment compared to ERBB2-negative patients (Muss et al., 1994). This finding has subsequently been confirmed in several studies (Thor et al., 1998; Paik et al., 1998; Petit et al., 2001; Harris et al., 2001; Kim et al., 2002). Again, other studies indicated no relationship between ERBB2 status and doxorubicin chemosensitivity (Niskanen et al., 1997; Clahsen et al., 1998; Rozan et al., 1998; Hamilton et al., 2000).

The contradictory results obtained in the studies reviewed above may be partly due to the significant variation in methods used to assess ERBB2 status. Detection of ERBB2 overexpression with immunohistochemistry (IHC) on formalin-fixed paraffin-embedded specimens especially has provided inconsistent results, whereas studies using gene-based detection methods, such as Southern blotting, fluorescence *in situ* hybridization (FISH) and chromogenic in situ hybridization, have more uniformly established the association between ERBB2 and poor outcome (reviewed by Ross and Fletcher, 1999). In conclusion, the overwhelming majority of the studies indicate that ERBB2 status does indeed correlate with patient outcome. However, although strong trends suggesting an association between positive ERBB2 status and resistance to tamoxifen therapy and sensitivity to anthracyclines have been demonstrated, more studies are still needed to confirm these associations.

3.2.2 ERBB2 as a target for therapy

The high prevalence of ERBB2 amplification and overexpression in breast cancer, association of ERBB2 positive status with poor clinical outcome, and cell membrane localization of the ERBB2 receptor enabling extracellular manipulation, have made ERBB2 an ideal target for cancer therapy. One approach to downregulate the ERBB2 signaling pathway was the development of a recombinant humanized monoclonal antibody, Herceptin, targeted against the extracellular domain of ERBB2 (Carter et al., 1992). Several clinical trials have evaluated the safety and efficacy of Herceptin as a single agent therapy in ERBB2-positive metastatic breast cancer showing overall response rates ranging from 11.6 % to 26 % (Baselga et al., 1996; Baselga et al., 1999; Cobleigh et al., 1999; Baselga, 2001; Vogel et al., 2001; Vogel et al., 2002). Herceptin has also been demonstrated to potentiate the antitumor activity of other agents when used in combination with paclitaxel, doxorubicin, docetaxel, cisplatin, and

vinorelbine (Baselga et al., 1998; Pegram et al., 1998; Pegram and Slamon 1999; Meden et al., 2001; Jahanzeb et al., 2002; Burstein et al., 2003; Klos et al., 2003; Montemurro et al., 2003). Recently it has been shown that Herceptin also increases the clinical benefit of first-line standard chemotherapy in metastatic breast cancer overexpressing ERBB2 (Eiermann, 2001; Slamon et al., 2001). To potentiate the effect of the anti-ERBB2 monoclonal antibodies, they can be combined with immunotoxins, radionuclides and prodrugs (revieved by Yarden and Sliwkowski, 2001).

In addition to antibodies targeting the ERBB2 receptor, several other approaches aiming at decreasing the signaling through the ERBB2 pathway have been envisioned (reviewed by Prenzel et al., 2001). These include low molecular weight tyrosine kinase inhibitors (e.g. genistein, emodin, curcumin) targeted against the tyrosine kinase activity of the ERBB2 receptor, transcriptional repression by adenovirus type 5 E1A, antisense oligonucleotides, antisense RNAs, and dominant negative mutants. Inhibition of ERBB2 receptor transfer from endoplasmic reticulum to plasma membrane by single chain antibodies, induction of natural immune response by ERBB2 DNA vaccines, and downregulation of ERBB2 protein and mRNA expression by retinoid acids have also been postulated (reviewed by Yu and Hung, 2000; Yarden and Sliwkowski, 2001).

4. Principles of DNA microarray technology

Implementation of the DNA microarray technology has revolutionized molecular genetics, allowing researchers to study thousands of genes simultaneously in a single experiment. At the same time, the Human Genome Project has gathered increasing amounts of information about human genes, making it possible to construct more complete microarrays containing in principle the entire human genome. As the technology becomes more accessible for all researchers, microarray analysis is finding applications in diverse areas of biology. In cancer research, microarray based expression and copy number profiling can be used e.g. in the identification of amplification target genes, classification of various tumor types, definition of subclasses within a specific cancer type, and prediction of patient outcome. There are various platforms to use for gene expression (oligonucleotide, cDNA) and gene copy number (oligonucleotide, cDNA, genomic clones) analysis, which will be discussed in more detail below (Figure 2).

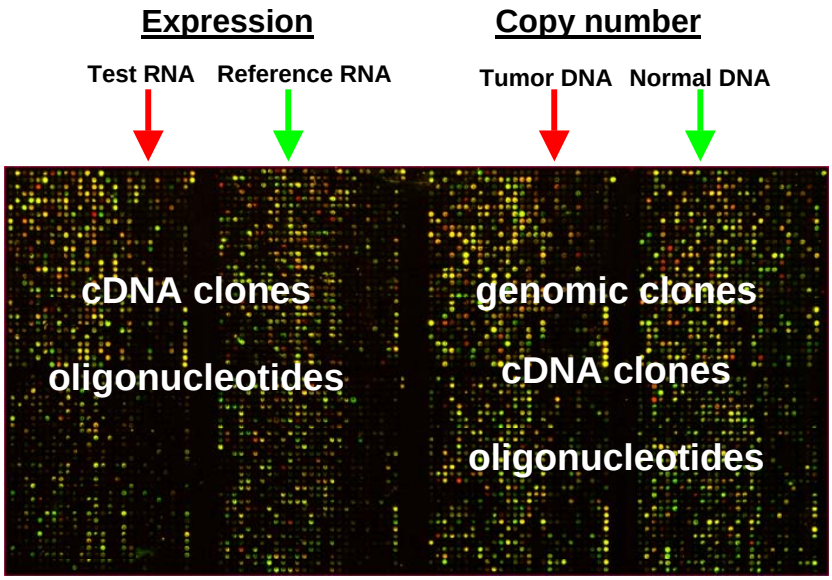

Figure 2. A summary of the different types of array technologies.

4.1 Expression analysis

Monitoring of the RNA expression levels can be done using either cDNA clone or oligonucleotide microarrays. cDNA microarrays are produced by spotting PCR amplified inserts of cDNA clones (approximately 0.6-2.4 kb in size) representing specific genes onto a solid support, typically on glass slides (Schena et al., 1995). DNA targets printed on the array can be selected from various databases, such as GenBank, dbEST, and UniGene, or may represent a random collection of transcribed sequences derived e.g. from a specific cDNA library. Printing of the PCR products onto glass slides is normally carried out by a robot using either contact (fountain pen like device) or non-contact (piezo or ink-jet device) printing mode (reviewed by Duggan et al., 1999). Total RNA or mRNA from the test and reference cells to be examined are reverse transcribed into cDNA and labeled with fluorochoromes. The differentially labeled test and reference samples are then hybridized together onto a microarray slide. The relative expression level of each gene on the microarray is based on a direct comparison between test and reference fluorescence intensities (Schena et al., 1995; DeRisi et al., 1996; reviewed by Duggan et al., 1999).

Another approach to defining expression levels is to use oligonucleotide microarrays (Lipshutz et al., 1995; Lockhart et al., 1996). There are two different platforms for high-density oligonucleotide array manufacture. The first method involves a combination of photolithography and *in situ* synthesis of oligonucleotides (Fodor et al., 1991; Pease et al.,

1994; Lipshutz et al., 1999). Instead of hybridizing two samples labeled with different fluorochromes, for this format only a single sample is hybridized at a time onto the microarray. In addition to the photolithographic method, oligonucleotide microarrays can also be manufactured by spotting of the oligos onto glass slides. For this type of array, differentially labeled test and reference samples are hybridized together as in cDNA arrays. Oligonucleotides synthesized using the photolithographic method are usually 25-mers, whereas spotted oligonucleotides are somewhat longer, 50-70-mers (Kane et al., 2000; Hughes et al., 2001; Holloway et al., 2002; Relogio et al., 2002; Barczak et al., 2003).

The oligonucleotide microarrays provide greater specificity compared to cDNA arrays in gene expression analysis. Oligonucleotide arrays are capable of distinguishing single-nucleotide polymorhism, splice variants, and closely related sequences (Hughes et al, 2001; Holloway et al., 2002). In addition, one can use multiple oligonucleotides hybridizing to different regions of the same RNA, which markedly improves the signal-to-noise ratios and accuracy of quantification and also reduces the rate of false positives. The use of mismatch oligonucleotides containing a single base difference in the middle of the oligo allows direct subtraction of background and cross hybridization signals, further increasing the specificity of the oligonucleotide array analysis (reviewed by Lipshutz et al., 1999). However, because the probe length is much shorter than with cDNAs, oligonucleotides must be carefully designed, since the base composition may greatly affect their performance. Probe length also affects the sensitivity of the array, cDNA clones providing better signal intensity due to longer length of the probes. The main advantage of cDNA microarrays is that cDNA clones are easy and cheap to manufacture, bringing them within the reach of a wider spectrum of researchers (reviewed by Holloway et al., 2002).

4.2 Copy number analysis

DNA microarrays have also been adapted for detection and mapping of copy number changes by CGH (Solinas-Toldo et al., 1997, Pinkel et al., 1998; Pollack et al., 1999; Lucito et al., 2003). In this approach, total DNA from tumor and normal genomes is isolated, labeled with different fluorochoromes, and hybridized simultaneously to the array. The results are presented as ratios between tumor and normal reference intensities. The earliest studies applying CGH microarray technique arrayed large-insert size genomic clones, such as cosmids, P1, PAC and BAC clones, on glass slides (Solinas-Toldo et al., 1997; Pinkel et al.,

1998, Albertson et al., 2000). In 1999, application of CGH onto cDNA microarrays was developed by Pollack and coworkers (1999) and recently the first microarray-based CGH study utilizing oligonucleotide microarrays was published (Lucito et al., 2003).

Microarray-based CGH using either genomic, cDNA or oligonucleotide arrays has proved to be suitable for the detection of both increased and decreased copy numbers. The advantages of the genomic array include high quantitativity and sensitivity as a result of hybridization signal integration over long BAC probes (Lucito et al., 2000; Albertson and Pinkel, 2003). The resolution of the CGH microarray is dependent on the number of clones present on the array and the distribution and spacing of the clones throughout the genome. In the case of the genomic array the limiting factor of the resolution is the length of the BAC clones (Pinkel et al., 1998). This is the reason why cDNA and oligonucleotide microarrays provide much better resolution compared to genomic arrays. In cDNA arrays the theoretical limit is a single gene. However, in practice a signal is usually averaged over several adjacent clones in order to increase signal-to-noise ratio (Pollack et al., 1999; Albertson and Pinkel, 2003; Lucito et al., 2003). The best resolution today is obtained by oligonucleotide microarrays, where multiple oligonucleotide probes can be organized for a single gene. A recently published study applying oligonucleotide microarray in copy number analysis achieved an average resolution of 30 kb throughout the whole genome (Lucito et al., 2003).

One of the main disadvantages of genomic arrays is the propagation and printing of BAC clones, which may be quite difficult because of low yields of BAC DNA and the viscous nature of the high molecular weight DNA solutions (Lucito et al., 2000; Albertson and Pinkel, 2003; Albertson, 2003). To overcome these problems, several PCR-based methods have been applied to provide amplification of representations of BAC DNAs (Snijders et al., 2001; Telenius et al., 1992; Fiegler et al., 2003). In CGH on oligonucleotide microarrays, PCR amplification step is performed to create representations of the genome to be analyzed. It is possible that this amplification step might introduce a quantitative bias, if the quality of the sample and reference DNA differs. By contrast to genomic clones, cDNA clones are easy to grow, amplify and print on glass slides. In addition, because cDNA clones are reverse transcribed from mRNA sequences, they are focused on expressed sequences and contain only a few repeat sequences compared to genomic clones, where masking of the repetitive sequences can be sometimes problematic (Pollack et al., 1999; Pollack et al., 2002). Disadvantages of the cDNA array platform include the unability to detect alterations in non-

coding genomic regions and to distinguish genes from pseudogenes. However, the major advantage of using cDNA clones or oligonucleotides as hybridization targets is that the exact same microarray can be used for both gene copy number and expression analysis, providing rapid correlation between gene copy number aberrations and changes occurring in gene expression levels (reviewed by Albertson and Pinkel, 2003).

5. Microarray applications in cancer research

5.1 Expression profiling

DNA microarray based expression analyses have been widely used to classify tumors. Most commonly molecular classification studies have been applied to a single cancer type but multiclass classification has also been performed (Ramaswamy et al., 2001; Su et al., 2001; Yeang et al., 2001; Sheddan et al., 2003). Ramaswamy and colleagues (2001) determined the expression patterns in 218 tumor samples representing 14 common tumor types using oligonucleotide microarrays containing 16,063 genes. Multiclass classification using a supervised learning method accurately assigned 78 % of the tumors according to the tissue of origin. Poorly differentiated tumors could not be correctly classified, demonstrating that they are molecularly distinct from their well differentiated counterparts (Ramaswamy et al., 2001). In another study Su et al. (2001) identified a set of 110 classifier genes whose expression patterns accurately predicted the tissue of origin for 90 % of tumors, including 9 out of 12 metastatic tumors.

Expression profiling has also facilitated classification of tumor types sharing the same kind of histology. In the study by Khan and co-workers (2001), artificial neural networks (ANN) were used to classify small round blue cell tumors (SRBCTs) that are otherwise difficult to distinguish due to similar histology into four separate diagnostic categories consisting of neuroblastoma, rhabdomyosarcoma, non-Hodgkin lymphoma, and the Ewing family of tumors. Expression analysis of 63 training samples by cDNA microarray containing 6,567 genes identified a set of 96 classifier genes. A subsequent set of 25 blinded samples was correctly classified using these 96 classifier genes demonstrating the usefulness of this method for accurate class prediction (Khan et al., 2001).

Most commonly expression profiling has been used to define subclasses within a specific tumor type (Golub et al., 1999; Alizadeh et al., 2000; Bittner et al., 2000; Perou et al., 2000; Sorlie et al., 2001; Dhanasekaran et al., 2001; Hedenfalk et al., 2001; Kim et al., 2002; Hedenfalk et al., 2003; Takahashi et al., 2003; Frederiksen et al., 2003; Tay et al., 2003). Cancer classification was first applied to human acute leukemia using an oligonucleotide microarray containing 6,817 genes. Using an initial collection of known samples, gene sets that were differentially expressed in acute myelogenous leukemia (AML) and acute lymphocytic leukemia (ALL) were identified and used to classify unknown leukemia samples into the correct categories based solely on the gene expression profiles (Golub et al., 1999). Expression based class discovery has also been used in solid tumors. For example, analysis of gene expression profiles in cutaneous malignant melanomas was able to define two putative subsets of tumors (Bittner et al., 2000).

Molecular classification has also been performed in breast cancer (Perou et al., 2000; Sorlie et al., 2001; Hedenfalk et al., 2001; Hedenfalk et al., 2003). In a study analyzing 65 breast adenocarcinomas from 42 individuals by cDNA microarray containing 8,102 genes it was possible to classify these tumors into three distinct subtypes, ERBB2 positive, estrogen receptor positive luminal-like, and basal-like, by hierarchical cluster analysis (Perou et al., 2000). In a subsequent study in a larger set of 78 breast tumors using identical cDNA microarray the same three subgroups of tumors were demonstrated (Sorlie et al., 2001). As a novel finding, the estrogen receptor positive luminal-like group could be further divided into at least two subgroups with distinctive expression profiles. In addition, survival analysis indicated that the basal-like and the ERBB2 positive subtypes are associated with the shortest survival times and that the two luminal-like subtypes represent clinically distinct groups with different outcomes (Sorlie et al., 2001). In breast cancer, expression profiling has also been used to accurately classify hereditary tumors from BRCA1 and BRCA2 mutation carriers (Hedenfalk et al., 2001). In addition, novel subclasses were discovered among the non-BRCA1/2 breast cancer patients, most likely reflecting the heterogeneous genetic origin of these tumors (Hedenfalk et al., 2003).

Gene expression profiling has recently been used to predict disease outcome in breast cancer patients (Sgroi et al., 1999; Martin et al., 2000; West et al., 2001; Van´t Veer et al., 2002; Van de Vijver et al., 2002; Ahr et al., 2002; Huang et al., 2003). For example, differential display and cDNA arrays were used in the identification of four clusters of genes associated with

three major clinical parameters of breast tumors (ER status, tumor stage and tumor size). Expression patterns of these four different clusters were used to group breast tumors into distinct therapeutic categories (Martin et al., 2000). In another study Van´t Veer and co-workers (2002) identified a set of 70 genes being able to accurately classify breast cancer patients into poor prognosis and good prognosis groups based on expression profiles. The set of 70 classifier genes was futher validated in a study with 295 patients having either lymph-node-negative or lymph-node-positive breast cancer. Interestingly, the prognosis of the patients did not seem to be dependent on lymph node status, because lymph-node negative and positive patients were distributed almost equally between the two prognostic categories (van de Vijver et al., 2002). Recently Huang and co-workers (2003) were able to classify breast tumors in defined subgroups associated with lymph-node metastases and recurrency based on gene expression patterns.

Expression studies have also challenged the prevailing model of metastasis suggesting that capacity to metastasize arises late in tumorogenesis and only from rare cells of the primary tumor (Fidler and Kripke, 1977). In addition to the studies reviewed above, where gene expression profiling made it possible to distinguish tumor classes with different prognosis and metastatic potential, other studies supporting a variant model of cancer metastasis have been published (Ramaswamy et al., 2003; Schmidt-Kittler et al., 2003; Woelfle et al., 2003; Kang et al., 2003). Ramaswamy and co-workers (2003) addressed this issue directly by comparing gene expression profiles of adenocarcinoma metastases of different tumor types to those of unmatched primary adenocarcinomas. A specific expression signature capable of distinguishing primary tumors from metastases was identified. However, a subset of primary tumors carried an expression signature similar to that of metastatic tumors, suggesting that the metastatic potential is present in the bulk of primary tumors (Ramaswamy et al., 2003). Similarly, the study by Schmidt-Kittler and colleagues (2003) suggests that tumor cells disseminate from the primary tumor at a much earlier state of tumor progression than previously expected. Their findings also indicate that the disseminated cell population is not the most advanced clone within the primary tumor, as expected by the prevailing model of metastasis (Schmidt-Kittler, 2003).

5.2 High resolution copy number analysis

Microarray based CGH has been most frequently used for the analysis of copy number increases previously described by chromosomal CGH and for the rapid identification of putative amplification target genes. Utilization of CGH microarrays has enabled detailed structural chracterization of amplicons and definition of specific amplicon boundaries. For example, amplification affecting the q-arm of chromosome 20 has been studied in detail using genomic arrays containing large insert size BAC clones (Pinkel et al., 1998; Albertson et al., 2000). Such an analysis provided detailed information of the 20q amplicon structure and revealed two independent amplification peaks. In addition, the CYP24 gene was identified as a novel candidate target gene for the 20q amplicon in breast cancer (Albertson et al., 2000). Similar detailed characterization of specific previously discovered amplicons has been performed in breast and gastric cancers as well as testicular germ cell tumors by using region-specific cDNA microarrays (Monni et al., 2001; Varis et al., 2002; Zafarana et al., 2003).

Clark and co-workers (2002) used an alternative approach to identify amplification target genes. They prepared a cDNA library from a breast cancer cell line containing 17q12, 17q23, and 20q13 amplifications and constructed a cDNA microarray by printing randomly selected clones from this library. Parallel gene copy number and expression analysis was performed to identify putative amplification target genes that were overexpressed (Clark et al., 2002). The main advantages of this technique are that it is not limited to a pre-selected set of genes and that genes overexpressed in the sample are highly represented in the array. However, at the same time the analysis is limited to genes expressed in the original sample.

Recently several studies utilizing CGH microarray in genome-wide copy number analysis have been published, both using genomic arrays (Snijders et al., 2001; Cai et al., 2002; Massion et al., 2002; Veltman et al., 2003) and cDNA microarrays (Pollack et al., 1999; Pollack et al., 2002; Clark et al., 2003; Squire et al., 2003). The first genome-wide BAC array contained 2460 BAC and P1 clones throughout the genome providing an average resolution of approximately 1.4 Mb (Snijders et al., 2001). A similar resolution was also obtained in the analysis of genome-wide copy number changes in bladder tumors. In addition to previously identified alterations, many new aberrations including high-level amplifications and homozygous deletions involving small genomic region were identified (Veltman et al., 2003). The first genome-wide copy number profiling utilizing cDNA microarrays was performed by

Pollack and co-workers in 1999. The results obtained confirmed previous data on copy number changes in higher resolution and parallel expression analysis using an identical cDNA microarray enabled identification of candidate amplification target genes (Pollack et al., 1999). Recently Clark et al., (2003) demonstrated that CGH microarrays could also be applied for deletion detection. In this study, CGH on cDNA microarrays was used for genome-wide screening of genetic loss in prostate cancer. Copy number analysis in prostate cancer cell lines allowed specific mapping of homozygous deletions and led to the identification of two novel regions of complete loss at 17q21.31 and 10q23.1 (Clark et al., 2003).

As in expression profiling, CGH microarray data have also been used for disease classification (Wilhelm et al., 2002; Fritz et al., 2002; Weiss et al., 2003). For example, array-based CGH was used to evaluate copy number changes in different types of renal cell carcinomas (Wilhelm et al., 2002). Copy number information obtained from 24 BAC clones was used in hierarchical clustering and resulted in the classification of 33 of the 34 malignant tumors into correct histological subtype. As would be expected, normal kidney samples and benign neoplasms with no copy number changes were grouped together (Wilhelm et al., 2002). Similarly, Fritz and co-workers (2002) utilized CGH microarray based gene copy number analysis to classify dedifferentiated and pleomorphic liposarcomas. In this study, gene expression profiling was not able to provide as accurate a classification as that obtained by gene copy number analysis (Fritz et al., 2002). However, this result may be partly due to the relatively low number of genes analyzed in the expression survey. Recently, Weiss et al. (2003) applied CGH microarrays in the classification of gastric adenocarcinomas. Hierarchical cluster analysis revealed three different groups of gastric cancers showing different patterns of genomic aberrations. In addition, these three groups correlated with lymph node status and survival, thus indicating the possible clinical utility of copy number based tumor classifications (Weiss et al., 2003).

AIMS OF THE STUDY

1) To define the structure and extent of the ERBB2 amplicon at 17q12 in breast cancer cell lines and primary breast tumors.

2) To identify amplified and overexpressed genes at the ERBB2 amplicon in breast cancer.

3) To evaluate the effects of Herceptin treatment on gene expression patterns in ERBB2 amplified and non-amplified cell lines.

4) To perform genome-wide mapping of DNA amplifications in breast cancer using high-resolution CGH on cDNA microarray.

5) To evaluate the global impact of gene amplification on gene expression levels in breast cancer.

MATERIALS AND METHODS

1. Cell lines (I, III, IV)

A total of 23 established breast cancer cell lines were used in this study. Breast cancer cell lines BT-474, BT-20, MDA-361, MDA-436, MDA-453, MDA-468, HCC202, HCC1419, HCC1428, HCC1954, UACC-732, UACC-812, UACC-893, UACC-3133, ZR-75-1, ZR-75-30, Hs578t, MCF7, SKBR-3, T-47D and HBL-100 were obtained from American Type Culture Collection (Manassas, VA) and MX-1 and EFM-192A from the German Collection of Micro-organisms and Cell Cultures (Braunschweig, Germany). Normal human mammary epithelial cells, HMECs, were obtained from Clonetics (Walkersville, MD). All cells were grown under recommended culture conditions.

2. Clinical tumor samples (II, IV)

A tissue microarray containing 612 formalin-fixed, paraffin embedded primary breast tumors from the years 1985-1995 was used. Tumor collection was obtained from the Institute of Pathology, University of Basel, Switzerland. All samples were anonymous, archival tissue specimens with associated clinicopathological information. The use of these tissue specimens in retrospective analyses was approved by the Ethics Committee of the University of Basel and by the NIH Institutional Review Board.

For Study II, 36 freshly frozen primary breast tumor specimens were obtained from the Tampere University Hospital (Tampere, Finland). The use of these samples for research purposes was approved by the Ethics Committee of the Pirkanmaa Hospital District. Tumors were pre-selected according to the histochemical status of the ERBB2 protein and the 17q12 amplification status was verified by FISH. The amplified (n=15) and non-amplified (n=21) tumor groups were also matched regarding standard clinicopathological features. According to a modified Bloom and Richardson classification, 8 of the tumors were histological grade II and 28 were grade III. There were 18 tumors with less than 2cm in diameter and 16 more than 2 cm. Fifteen tumors were axillary lymph node positive and 21 node negative. Routine

immunohistochemical analysis of the hormone receptor status showed that 26 tumors were estrogen receptor positive and 20 progesterone receptor positive.

3. Growth inhibition assay (III)

The response of breast cancer cells to the Herceptin treatment was determined by using the AlamarBlue™ Assay (Alamar Biosciences, Sacramento, CA). The recombinant DNA-derived humanized monoclonal antibody Herceptin (Trastuzumab) was obtained from Genentech, Inc. (South San Francisco, CA). Exponentially growing cells were harvested and plated on 96 well plates at defined density, ranging from $2.5 \times 10^4 - 6.5 \times 10^4$ cells/ml depending on cell line. After 48 hours of culture in recommended medium, Herceptin was added at concentrations of 1, 10, 100, and 500 µg/ml. Equal amounts of fresh medium lacking Herceptin were added to the control wells of each cell line. Plates were incubated for 72 hours, after which the Herceptin containing medium was removed and fresh medium was added for an additional 18 hours. To determine the number of proliferating cells, AlamarBlue was added to the wells in an amount equal to 10 % of the total culture medium volume. The AlamarBlue reagent contains an incorporated fluorometric/colorimetric oxidation-reduction (REDOX) growth indicator. Metabolic activity of the growing cells results in a chemical reduction of the REDOX indicator, which both fluoresces and changes color in response. After 4-10 hours of incubation, absorbance was measured at 540 nm and 620 nm wavelengths with a spectrophotometer (ELISA reader). The percent difference in growth inhibition between Herceptin-treated and untreated control cells were calculated according to manufacturer´s instructions (Alamar Biosciences). All experiments were performed in triplicate.

4. cDNA microarrays (I, II, IV)

4.1 Composition of cDNA microarrays

For Study I, a chromosome 17-specific cDNA microarray was constructed (Figure 3) using the information available in GeneMap´99 (Monni et al., 2001). The array contained a total of 636 cDNA clones, including 217 from 17q12 (interval D17S933-D17S930; 293-325 cR) and

156 from 17q23 (interval D17S791-D17S795; 333-435 cR). Genome-wide cDNA microarrays containing a total of 14,380 and 13,824 clones were used in Studies III and IV respectively. The majority of the arrayed clones represented known genes, with only 939 and 244 corresponding to uncharacterized ESTs. The preparation and printing of cDNA clones on glass slides were performed as described (De Risi et al., 1996; Shalon et al., 1996; Mousses et al., 2000).

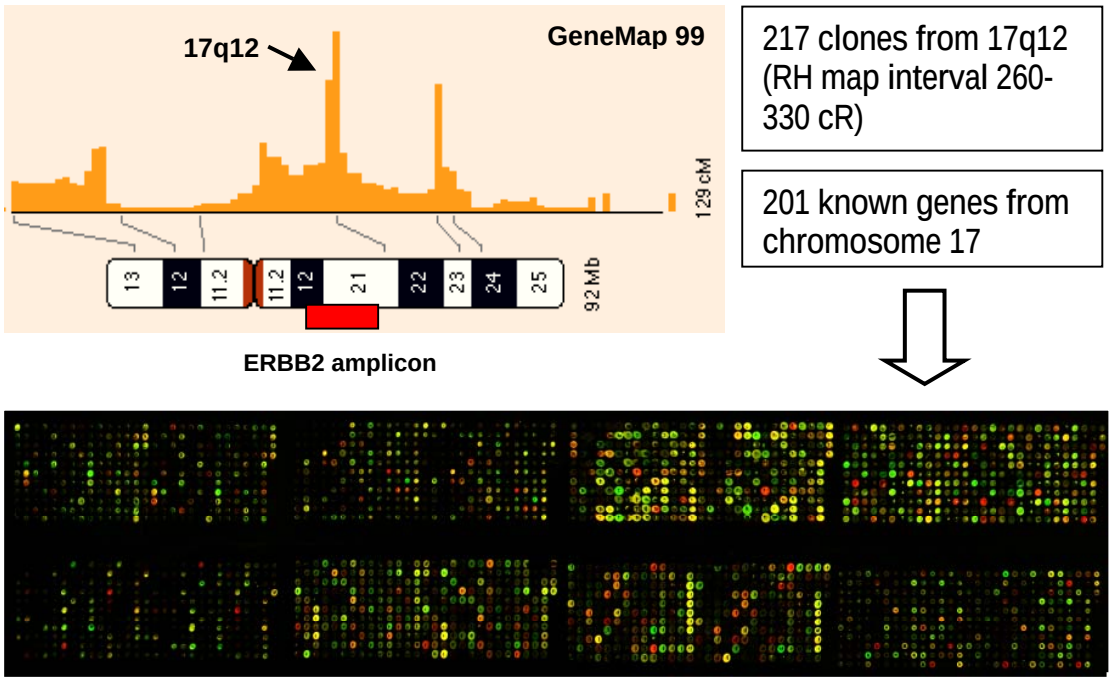

Figure 3. Construction of the chromosome-17 specific cDNA microarray used in the comprehensive copy number and expression profiling of 17q12 region (I).

4.2 Gene copy number and expression analyses

cDNA microarray technology was used for the analysis of gene copy numbers (I, IV) and gene expression levels (I, III, IV) in breast cancer cell lines (Table 1). Genomic DNA from breast cancer cell lines was extracted using Qiagen Blood and Cell Culture DNA Maxi Kit (Qiagen Inc., Valencia, CA) and normal placental (I) or human white blood cell DNA (IV) was used as a reference. CGH on cDNA microarrays was done as previously described (Pollack et al., 1999; Monni et al., 2001). Briefly, genomic DNAs were digested for 14-18 h with *AluI* and *RsaI* (Life Technologies, Inc., Rockville, MD) and purified by phenol/chloroform extraction. Six µg of digested cell line DNAs were labeled with Cy3-

dUTP and 6-8 μg of normal reference DNA with Cy5-dUTP using Bioprime Labeling Kit (Life Technologies, Inc., Rockville, MD). Hybridization was done as described (Monni et al., 2001), and posthybridization washes according to the protocol by (Mousses et al., 2000).

Table 1. Summary of microarray studies.

Study	Array (no. clones)	Reference	Test
I	chromosome-17 specific (636)	8 μg, placental DNA, Cy5	6 μg, genomic DNA, Cy3
I	chromosome-17 specific (636)	16 μg, MDA-436 mRNA, Cy5	4 μg, mRNA or 80 ug total RNA, Cy3
III	genome-wide (14380)	3 μg, mRNA (untreated), Cy3	3 μg, mRNA (treated), Cy5
IV	genome-wide (13824)	6 μg, white blood cell DNA, Cy5	6 μg, genomic DNA, Cy3
IV	genome-wide (13824)	40 μg pool of total RNA (Stratagene), Cy3	3.5 μg mRNA, Cy5

For the expression analyses, total RNA or mRNA was extracted from breast cancer cell lines using the Trizol reagent or FastTrack 2.0 Kit (Invitrogen, Carlsbad, CA). In Study I, the breast cancer cell line MDA-436 with no copy number increase at 17q12 and, in Study IV, a commercially available reference (Universal Human Reference RNA; Stratagene, La Jolla, CA) were used as a standard reference in all experiments. In Study III, breast cancer cell cultured in 10 μg/ml Herceptin and control cells grown in drug-free medium were harvested after 24, 48 and 72 hours of incubation. mRNA isolated from drug-treated test cells was hybridized against mRNA isolated from untreated control cells of the same cell line. All RNAs were labeled using oligo(dT)-primed polymerisation by SuperScript II reverse transcriptase (Life Technologies, Inc., Rockville, MD). Amounts and fluorescent labels used in the labeling reaction were as follows: in Study I, 4 μg of test mRNA or 80 μg of total RNA was labelled with Cy3-dUTP (Amersham Pharmacia Biotech, Piscataway, NJ) and 16 μg of reference mRNA with Cy5-dUTP (Amersham Pharmacia Biotech); in Study III, 3 μg of test mRNA was labeled with Cy5-dUTP and 3 μg reference mRNA with Cy3-dUTP; and in Study IV, 3.5 μg test mRNA was labeled with Cy5-dUTP and 40 μg of reference total RNA with Cy3-dUTP. Hybridization of the labeled cDNAs on microarray slides and posthybridization washes were performed as described (Mousses et al., 2000; Monni et al., 2001).

4.3 Data analysis

In all cDNA and CGH microarray analyses, a laser confocal scanner (Agilent Technologies, Palo Alto, CA) was used to measure the fluorescence intensities at the target locations. The test and reference fluorescent images were scanned separately and the data was analysed using DEARRAY software (Chen et al., 1997). After background subtraction, copy number and expression level of each gene was determined by comparing the average intensity of the clone in the test sample to the average intensity in the reference. For the copy number analysis, ratios were normalized based on the distribution of the ratios of all targets on the array (I, IV) and for the expression analysis based on the 88 housekeeping genes (I, IV). Low quality measurements, referring to expression data with test and reference intensity <100 fluorescent units and/or spot size <50 units (III, IV) and copy number data with mean reference intensity <100 fluorescent units (IV), were excluded from the analysis and were treated as missing values. In each study, specific statistical analyses were used as outlined below.

4.3.1 Study I

Copy number levels of chromosome 17 specific genes in breast cancer cell lines were evaluated against normal placental DNA as a reference and expression levels against MDA-436 cells, with no copy number increase in the 17q12 region. Clones showing a copy number ratio ≥ 1.5 and an expression ratio >3 in breast cancer cell lines relative to the reference, without such increase in HMECs, were considered to be amplified and overexpressed.

4.3.2 Study III

Data derived from 11,311 clones after quality filtering was normalized using Local Weighted Scatter Plot Smoother (LOWESS) analysis (Cleveland, 1979; Yang et al., 2001) for each print-tip group. Hierarchical clustering of the 11,311 genes at 24, 48, and 72 h time points was performed using the average linkage method with uncentered correlation as the distance function (Eisen et al., 1998). A collection of statistical analyses was then used to identify genes whose expression changed (either increased or decreased) as a function of time during the drug treatment. First, equalization transformation (Bolstad et al., 2003) was used to

convert ratios across the three time-series experiments (24, 48 and 72h incubation of the drug) of each cell line to a similar range. Next, linear regression analysis was used to convert the time-series observations of each sample into a single number *T* (Hautaniemi et al., 2003a), where the sign of *T* illustrates the direction of change in expression level (a positive sign is obtained when the expression level of the gene increases and negative when expression level decreases) and the value of *T* reflects the magnitude of expression change. As a result of this *T* value computation, each gene in each sample is characterized by a single quantity that illustrates whether expression levels increase, decrease or remain the same during the time-series of the drug treatment. These *T* values were then utilized in a stepwise gene selection algorithm to identify genes that would best separate the ERBB2 amplified cell lines from those with no amplification. For this purpose, BT-474, SKBR-3 and ZR-75-30 were assigned as amplified and MDA-436 and MCF-7 cell lines as non-amplified. MDA-453 was excluded because it demonstrates only low-level amplification and moderate expression of ERBB2 and thus did not fit into either group. Genes showing a difference greater than one in median expression level between the amplified and non-amplified groups were included for the analysis. In stepwise gene selection, weight (W) for each of the selected 2990 genes was computed using Fisher´s linear discriminant. New genes were added to the stepwise gene selection program until the weight reached a value of three. To verify that the result was statistically significant a randomization test was run for the 2990 genes. Clustering of the discriminatory genes was performed using the hierarchical clustering with average linkage method and Euclidean distance as the distance function and the SOM algorithm (Kohonen, 2001; Hautaniemi et al., 2003b) using the toolbox made by Vesanto et al., (2000).

4.3.3 Study IV

A statistical approach was used to determine the effect of gene copy number on gene expression patterns throughout the genome. First, CGH and cDNA data was log-transformed and normalized by median centering of the values in each cell line. cDNA ratios for each gene were also median centered across all 14 cell lines. Next, the distribution of fluorescence ratios was used to define cutpoints for increased (ratio >1.43; representing the upper 5% of the CGH ratios across all experiments) and decreased (ratio <1.43; representing the lower 5%) copy number. The CGH data for each gene was then divided in two groups representing gene copy number ratio values in amplified and non-amplified breast cancer cell lines. The correlation

between gene amplification and gene expression was assessed by calculating a weight for each gene as follows:

$$w_g = \frac{m_{g1} - m_{g0}}{\sigma_{g1} + \sigma_{g0}}$$

where m_{g1}, σ_{g1} and m_{g0}, σ_{g0} denote the means and standard deviations for the expression levels for amplified and non-amplified cell lines respectively. The statistical significance of each weight was tested by changing the order of amplified and non-amplified cell line groups randomly (10,000 times) and then comparing the re-calculated weight with the original weight. The probability that a gene had equal or greater weight with the random order of cell lines was denoted by alpha (α). Low alpha values are obtained when the association between amplification and gene expression is strong.

4.3.4 Genomic localization of cDNA clones and amplicon mapping (IV)

Each cDNA clone on the genome-wide microarray was assigned to a Unigene cluster using the Unigene Build 141 (http://research.nhgri.nih.gov/microarray/downloadable_cdna.html) information. The August 2001 freeze of the University of California Santa Cruz´s Golden Path database was used to create a database containing genomic sequence alignment information for mRNA sequences. The chromosome and exact base pair positions for each cDNA clone were then obtained by relating these two data sets. Amplicons were defined based on the following criteria: a CGH copy number ratio >2.0 in at least two adjacent clones in two or more cell lines or a CGH ratio >2.0 in at least three adjacent clones in a single cell line. In addition, the amplicon start and end positions were extended to include neighboring non-amplified clones (ratio <1.5). The amplicon size determination was partially dependent on local clone density.

5. Fluorescence in situ hybridization (I, II, IV)

5.1 Interphase FISH on breast cancer cell lines (I)

Several gene-specific BAC/PAC clones were used as probes in this study (Table 2). Clones were identified by performing sequence similarity searches against the nr and htgs databases using the blastn program and the specificity of the probes was confirmed by PCR with gene-

specific primers. Probe for the human EGFR gene and centromeric probes for chromosomes 7 and 17 were commercially available (Vysis inc., Downers Grove, IL). Gene-specific probes were labeled with SpectrumOrange-dUTP (Vysis) using random priming and SpectrumGreen-labeled centoromere probes were used as a reference (Vysis). Metaphase and interphase cell preparations from the breast cancer cell lines and normal blood lymphocytes were prepared as described (Kallioniemi et al., 1992). Control hybridizations to normal metaphase chromosomes were done to verify that the labeled probes recognized a single copy target. Dual-color FISH to breast cancer cell lines and normal lymphocytes was performed as described earlier (Bärlund et al., 2000). Hybridization signals were evaluated using either a Zeiss or an Olympus fluorescence microscope equipped with a dual band-pass fluorescence filter (Chromatechnology, Brattleboro, VT) and a 63x oil-immersion objective. Approximately 50 non-overlapping nuclei with intact morphology based on 4,6-diamidino-2-phenylindole (DAPI) counterstaining were scored to determine the number of hybridization signals for each test and reference probe in breast cancer cell lines.

Table 2. Genomic clones used in this study

GenBank accession number	Clone name
AC010761	RP11-386F9
AC006449	CTB-58E17
AC021102	RP11-129G15
AC005288	CTB-131K11
AC009283	RP11-390P24
AC040933	CTD-2019C10
AC019095	RP11-386H17
AC068669	RP11-749I16
AC015851	RP5-1112G21
AC018629	RP11-5809
AC004585	RP5-1028K7
AC009789	RP11-361K8

5.2 FISH on tumor tissue microarray (II, IV)

FISH to tissue microarray was performed as described (Andersen et al., 2001) using SpectrumOrange-dUTP labeled gene-specific BAC probes and SpectrumGreen labeled chromosome 17 centromere probe as a reference. The formalin-fixed tissue microarray sections on microscope slides were deparaffinized and treated according to the Paraffin Pretreatment Reagent Kit protocol (Vysis): the slides were fixed in 4% paraformaldehyde for

10 min, denaturated at 94°C for 5 min in Tth-buffer (10 mM Tris-HCl, pH 8.9, 0.1 M KCl, 1.5 mM MgCl, 50 µg/ml BSA, 0.05% Tween 20® (v/v)), treated with Proteinase K (10µg/ml in PBS) at 37°C for 10 min, dehydrated, and air-dried. After an overnight hybridization, the slides were washed in 0.4xSSC/0.3%NP-40 at 72°C for 1 min, at RT 2XSSC for 2 min, and then counterstained with 4', 6-diamidino-2-phenylindole in antifade solution. Hybridization signals were evaluated using a Zeiss fluorescence microscope (Carl Zeiss, Inc., Thornwood, NY). The number of cells counted varied from one tumor sample to another, with a minimum of 50 cells analyzed for each case. Strict criteria were applied to define amplification to ensure the accuracy of scoring. Specimens containing a 3-fold (Study II) or a 2-fold (Study IV) or higher increase in the number of test probe signals, as compared with the chromosome 17 centromere signals in at least 10% of the tumor cells were considered to be amplified.

6. Northern hybridization (I)

Total RNA from breast cancer cell lines was extracted using TRIzol reagent (Life Technologies, Inc., Rockville, MD). Fifteen µg of RNAs were subjected to electrophoresis and transferred onto a Nytran membrane (Schleicher & Schuell, Keene, NH). PCR products or sequence-verified cDNA clone inserts corresponding to ERBB2, GRB7, MLN64, and EST 485823 were labeled with [α-^{32}P]-dCTP using random priming (*redi*prime II, Amersham Pharmacia, UK) and GAPDH probe (Clontech, Palo Alto, CA) was used as a reference to confirm equal loading of the samples. The blot was first prehybridized for 1h at 42ºC in NorthernMax Prehybridization/Hybridization Buffer (Ambion., Inc., Austin, TX) together with Herring Sperm DNA (Sigma Chemical Co., St. Louis, MO) and then hybridized overnight in the prehybridization solution at 45ºC-48ºC. The blot was washed several times with 2X SSC-0.05% SDS at RT and then twice in 0.1X SSC-01% SDS at 55ºC. Hybridized probe was detected by using a PhosphorImager (Molecular Dynamics, Sunnyvale, CA).

7. Reverse transcriptase-polymerase chain reaction (II, IV)

7.1 RT-PCR (IV)

RT-PCR was used to verify expression levels of HOXB7 gene in breast cancer cell lines. Reverse transcription and PCR amplification of HOXB7 and GAPDH reference gene were performed using 10 ng mRNA as a template (Access RT-PCR System, Promega Corporation, Madison, WI). First strand cDNA synthesis was done at 48°C for 45 min and PCR amplification as follows: initial denaturation for 2 min at 94°C, denaturation at 94°C for 30s, annealing at 58°C for 45s, and extension at 70°C for 1 min, 27 and 20 cycles for *HOXB7* and *GAPDH*, respectively. Primers for HOXB7 were: 5'-GAGCAGAGGGACTCGGACTT-3' and 5'-GCGTCAGGTAGCGATTGTAG-3' and for GAPDH: 5'-TCCTGGAAGATGG TGATGGGAT-3' and 5'- TGAAGGTCGGAGTCAACGGATT-3'. The *HOXB7* expression levels were determined relative to *GAPDH*.

7.2 Quantitative RT-PCR (II)

Gene expression analyses in freshly frozen tumor tissue samples were performed using quantitative RT-PCR. A representative area from each primary breast sample was selected based on hematoxylin-eosin staining of tissue sections. Total RNA was isolated from 2 mm core biopsies using Qiagen RNeasy MiniKit (Qiagen Inc., Valencia, CA). In order to avoid genomic DNA contamination, samples were treated with RNase Free Dnase I (Epicentre, Madison, WI) at 37ºC for 30 min, followed by inactivation of the enzyme at 65ºC for 15 min. First-strand cDNA synthesis was performed using Superscript II reverse transcriptase and random hexamer primers (Invitrogen, Carlsbad, CA). Molecular Beacon Probe Sets for eight known genes (ERBB2, GRB7, MLN64, PNMT, NEUROD2, ZNFN1A3, TCAP, PPP1R1B), two hypothetical proteins (MGC14832, MGC9753), and the housekeeping gene TPB as well as specific double stranded DNA standards for each gene were obtained from Gorilla Genomics, Inc. (Alameda, CA).

Quantitative real-time PCR analyses were performed using LightCycler equipment (Roche, Mannheim, Germany; Wittwer et al., 1997). The total reaction volume of 20 µl consisted of 2 µl of 10x PCR Buffer, 0.4 µl 50x Probe Mix (Gorilla Genomics, Inc.), 0.4 µl TITANIUM™ Taq DNA polymerase (Clontech Laboratories, Inc., Alameda, CA), and 1 µl of cDNA sample or DNA standard, adjusted to final volume with sterile H_2O. The PCR program consisted of initial denaturation at 95ºC followed by 45 cycles of denaturation at 95ºC for 10s, annealing at 55ºC for 10s, and elongation at 72ºC for 6s. Quantitative analysis was performed using the LightCycler software according to manufacturer´s instructions. First the proportional baseline fluorescence adjustment was performed and then the fit-point method was used to determine the crossing point value, representing the cycle number where the fluorescence level for each sample was the same. The crossing point value is dependent on the starting amount of the target mRNA in the sample, small target amounts taking more cycles to reach the specific fluorescence level. A dilution series of the gene-specific DNA standard was used as a template in each PCR run to prepare a standard curve by plotting the crossing point of each standard against the logarithmic value of its concentration. The concentrations of the unknown samples were then determined by setting their crossing points to the standard curve. The expression levels of the genes studied were normalized against the housekeeping gene TBP.

8. Statistical methods (II, IV)

In Study II, association of 17q12 amplification with clinocopathological parameters of the primary breast tumors was determined using paired Student´s t-Test analysis. Comparison of the median gene expression levels between the 17q12 amplified and non-amplified tumor groups was performed using the nonparametric Mann-Whitney test. All p-values were two-tailed. In Study IV, the survival analysis for the HOXB7 gene amplification was performed using the Kaplan-Meier method and the log-rank test.

9. Genomic databases (I-IV)

During the course of this study the following sources of information were most often used:

National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/)
Unigene Resources (http://www.ncbi.nlm.nih.gov/Unigene)
Unigene Builds (http://research.nhgri.nih.gov/microarray/downloadable_cdna.html)
MapView (http://www.ncbi.nlm.nih.gov/mapview/)
GeneMap (http://www.ncbi.nlm.nih.gov/genemap)
GeneOntology Consortium (http://www.geneontology.org/)
Source (http://source.stanford.edu/cgi-bin/sourceSearch)
University of California Santa Cruz Genome Bioinformatics (http://www.genome.ucsc.edu)

RESULTS

1. Characterization of the ERBB2 amplicon in breast cancer (I, II)

1.1 Copy number and expression analysis of the 17q12 amplicon by chromosome 17 specific cDNA microarray (I)

In order to identify amplified and overexpressed genes at the 17q12 region, comprehensive expression and copy number analysis across the 17q12 amplicon was performed using a custom made chromosome 17-specific cDNA microarray containing 636 cDNA clones. Seven breast cancer cell lines (BT-474, HCC202, MDA-362, MDA-453, SKBR-3, UACC-812, UACC-893) known to have gain or amplification at 17q12 by chromosomal CGH (Forozan et al., 2000; Larramendy et al., 2000) were selected for expression analysis. In addition four cell lines (BT-474, SKBR-3, UACC-812, UACC-893) were applied for CGH microarray analysis in order to obtain copy number information of all the clones on the array. Focused examination of the expression levels at 17q12 region identified 20 transcripts, including 11 known genes and 9 uncharacterized ESTs that were overexpressed in three or more of the cell lines studied. Comparison between the expression and copy number data in four cell lines indicated that 12 of these 20 transcripts (CDC6, ERBB2, GRB7, MLN51, MLN64, ZNF144, as well as ESTs 48582, 244062, 418240, 767775, 1623927, and 810305) were consistently overexpressed when amplified.

1.2 Defining the minimal common region of 17q12 amplification (I, II)

To determine the structure of the ERBB2 amplicon and to obtain information on the physical order and location of the overexpressed and amplified clones identified by cDNA microarray, a detailed copy number analysis across a 6 Mb region at 17q12 was performed in 16 breast cancer cell lines and 330 primary breast tumors. Interphase FISH analyses using large-insert-size BAC clones revealed threefold or higher amplification with at least one of the probes in 11 (75 %) cell lines and 53 (16.1 %) primary breast tumors. Both the level of amplification and the size of the amplicon varied depending on the cell line or primary tumor examined. The highest amplification amplitudes (> 50 copies/cell) were seen in the BT-474, HCC1419,

HCC1954, and UACC-893 cell lines. The size of the amplicon was largest in EFM-192 and UACC-812 and smallest in MDA-361 and MDA-453 cell lines. The majority of amplified primary breast tumors (86.8 %) showed co-amplification of at least three adjacent probes, indicating that the size of the amplicon in this region is large and the pattern of amplification is contiguous. The most frequently amplified clone both in the breast cancer cell lines and in the primary breast tumors was BAC 2019C10, which defined the minimal common region of the 17q12 amplification in breast cancer.

1.3 Expression analysis of genes located at the minimal common region of 17q12 amplification (I, II)

Next the expression levels of the transcripts located in the minimal common region of 17q12 amplification were determined. Based on the information obtained from human genome databases, the minimal region defined by BAC 2019C10 spans approximately 280 kb of genomic DNA and contains ten transcripts. Eight of these represent known genes (ERBB2, GRB7, MLN64, PPP1R1B, PNMT, NEUROD2, TCAP, ZNFN1A3) whereas two encode uncharacterized hypothetical proteins (MGC9753, MGC14832). Four of the minimal region genes (ERBB2, GRB7, MLN64, MGC9753=EST 48582) had been included in the chromosome 17-specific cDNA microarray analysis and were highly expressed. Northern hybridization performed in 14 breast cancer cell lines confirmed that expression of all four transcripts was consistently elevated in 17q12 amplified cell lines (BT-474, SKBR3, UACC-812, UACC-893, EFM-192A, HCC1419, HCC1954, ZR-75-30, MDA-361) compared with (MCF7, HBL-100, MDA-436, MX-1) non-amplified cell lines and normal mammary epithelial cells (HMEC). A discrepancy was observed only with MDA-453 cell line, which showed increased expression by cDNA microarray but not by Northern analysis.

Expression analysis of the ten minimal region genes in primary breast cancer was carried out in a set of 36 freshly-frozen tumors consisting of 15 amplified and 21 non-amplified cases using quantitative RT-PCR. Expression of the TCAP and NEUROD2 genes was either very weak or absent in all tumor samples. For the remaining eight genes (ERBB2, GRB7, MLN64, PNMT, MGC9753, MGC14832, ZNFN1A3, PPP1R1B) the median value of expression was calculated in the amplified and non-amplified tumor groups and the statistical significance between median expression and gene amplification status was evaluated using nonparametric Mann-Whitney test. Six of the genes (ERBB2, GRB7, PNMT, MLN64, MGC9753,

MGC14832) showed a highly significant association between the amplification status and the expression level. The most significant difference ($P < 0.0001$) was demonstrated for ERBB2 and GRB7 genes, which displayed >7 and >14 times higher median expression level in the amplified compared to the non-amplified tumors. For ZNFN1A3 and PPP1R1B genes no significant difference in the expression patterns between the two tumor groups was observed.

2. Validation of CGH microarray results by FISH (I)

To evaluate the sensitivity and specificity of the CGH microarray technique, a comparison of the copy number ratios obtained by CGH microarray and by FISH was performed in four cell lines (I). For this comparison, CGH microarray copy number ratios and relative FISH copy number ratios (obtained using the corresponding BAC clones) greater than 1.5 were classified as amplified. A total of 71 individual comparisons were made, of which 53 were correctly classified as amplified and 10 as nonamplified with both of the methods, indicating a 89% concordance (Figure 4). Of the remaining 8 comparisons, 5 showed false negative and 3 false positive copy number ratios by CGH microarray. However, all false positive observations were obtained with a single EST clone that showed increased copy number ratios in three of the four cell lines, although FISH copy numbers were normal. In general CGH microarray showed lower copy number ratios than FISH analyses, thus underestimating the true copy number increases.

3. Effects of Herceptin treatment in ERBB2 amplified and non-amplified breast cancer cell lines (III)

In Study III, the effects of Herceptin treatment on gene expression levels in ERBB2 amplified (BT-474, SK-BR-3, ZR-75-30, MDA-453) and non-amplified (MCF7 and MDA-436) breast cancer cell lines were analyzed using a genome-wide cDNA microarray containing 14,380 cDNA clones. Growth suppression assay performed with four different concentrations (1, 10, 100, 500 μg/ml) of Herceptin demonstrated a dose-dependent growth reduction in breast cancer cell lines with ERBB2 amplification, whereas non-amplified cell lines showed only minimal effect. Interestingly, MDA-453 cell line with low-level ERBB2 amplification did not show any response with 1 μg/ml Herceptin similarly to non-amplified cell lines. However,

with higher doses of Herceptin a clear growth reduction was observed, although only a slight difference in response was observed with doses of 100 µg/ml and 500 µg/ml. The difference in mean growth suppression between ERBB2 amplified and non-amplified cell line groups was statistically significant ($p < 0.05$, two-tailed t-test) with Herceptin doses of 10, 100, and 500 µg/ml.

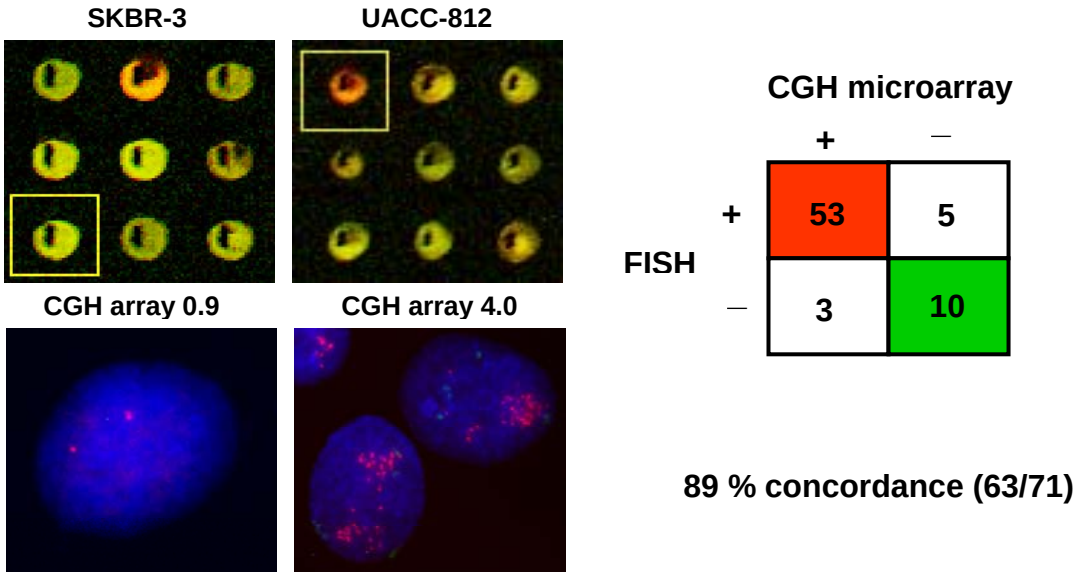

Figure 4. The panels on the left show examples of the CGH microarray hybridizations in the SKBR-3 and UACC-812 cell lines and corresponding FISH analysis. Comparison of copy number ratios obtained by CGH array and FISH are shown on the right. Comparisons correctly classified as amplified with both methods are shaded with red and comparisons correctly classified as non-amplified with green.

Global time-course expression analysis was performed by incubating cells for 24, 48, and 72 hours with 10 µg/ml Herceptin and then comparing the expression levels in the treated cells to those in untreated cells of the same cell line using the cDNA microarray. In order to evaluate the overall expression changes caused by Herceptin treatment hierarchical clustering was performed. In all six cell lines, the three different time-points clustered together indicating that the global expression patterns within each cell line were more similar than those between different samples. Furthermore, the hierarchical clustering dendrogram demonstrated that the 48 and 72 hour time points were more similar to one another than to the 24 hour time point in each cell line.

A combination of statistical methods (Hautaniemi et al., 2003a) was used to identify a specific set of genes whose temporal gene expression patterns differed most between ERBB2 amplified and non-amplified cell lines. This set of 439 discriminatory genes contained 212 down-regulated and 227 upregulated genes after the Herceptin treatment in the amplified vs. non-amplified cell lines. Temporal gene expression changes among the 439 discriminatory genes were analyzed using two different clustering approaches. Both clustering methods clearly separated the up- and down-regulated discriminatory gene sets. Hierarchical clustering showed separate amplified and non-amplified cell line clusters at the opposite ends of the dendrogram and the self-organizing map revealed two major clusters of genes with almost opposite profiles in the amplified and non-amplified cell lines. In addition, both clustering methods demonstrated that in MDA-453 cells with only low-level amplification and modest increase in ERBB2 expression a large fraction of genes showed gene expression changes similar to those seen in the non-amplified cell lines rather than in the amplified cell lines. During the antibody treatment, variation of the ERBB2 gene expression level was less than two-fold in all cell lines, indicating that the drug treatment did not have a significant effect on the expression of the ERBB2 receptor itself.

Based on the publicly available gene ontology information (http://source.stanford.edu/cgi-bin/sourceSearch) functional information was obtained for 59 % of the discriminatory genes. The majority of these genes encode for proteins implicated in key cellular processes like transcription, signal transduction, protein processing, cell metabolism, and transport. Genes encoding proteins involved in various RNA processing reactions, e.g. biogenesis of spliceosomal snRNAs, RNA 3´-end processing, and nuclear pre-mRNA splicing were typically upregulated in ERBB2 amplified cell lines as a result of Herceptin treatment. Herceptin also induced expression of genes involved in various DNA repair pathways, e.g. nucleotide-excision, double-strand break, and DNA mismatch repair. However, genes encoding cell adhesion proteins (e.g. JUP, CTNND2, and CNTN1) or well-characterized oncoproteins (e.g. FOS and KIT) showed decreasing expression levels in the amplified cell lines during the drug treatment.

4. Genome-wide copy number and expression analysis in breast cancer cell lines (IV)

4.1 Global effect of gene copy number on gene expression levels (IV)

To explore the global impact of gene copy number alterations on gene expression levels, parallel genome-wide gene copy number and expression analysis in 14 breast cancer cell lines (BT-20, BT-474, HCC1428, Hs578t, MCF7, MDA-361, MDA-436, MDA-453, MDA-468, SKBR-3, T-47D, UACC812, ZR-75-1, ZR-75-30) was performed using a cDNA microarray containing 13,824 clones. Comparison between gene copy number and expression data demonstrated that gene copy number alterations had a substantial effect on gene expression levels throughout the genome. Of the most highly amplified transcripts (i.e. CGH copy number ratio > 2.5) up to 44 % were also overexpressed, meaning that their ratio belonged to a global upper 7 % of the expression ratios. Among genes displaying normal gene copy numbers, only 6 % were overexpressed. The same trend was also observed with the low-level copy number aberrations, although the effect on gene expression level was less dramatic. On the contrary, 10 % of the transcripts displaying high-level expression (cDNA expression ratio >10) showed increased copy number, indicating that a large fraction of genes are overexpressed in the genome due to other mechanisms than gene amplification.

4.2 Genome-wide characterization of independent amplicons in breast cancer (IV)

High-resolution CGH microarray analysis was used to identify amplified regions in the 14 breast cancer cell lines. Exact basepair locations for 11,994 cDNA clones (86.8%) on the array were obtained from the GoldenPath Aug 2001 freeze (www.genome.ucsc.edu). The average spacing of the clones on the array was approximately 267 kb throughout the genome. Array CGH analysis identified altogether 24 independent amplicons (Figure 5), ranging in size from 200 kb to 12 Mb of DNA. These included all commonly amplified regions e.g. 1q21 (Chr 1: 173.92 – 177.25 Mb), 17q12-q21.2 (Chr 17: 39.79 – 42.80 Mb), 17q22-q23 (Chr 17: 63.81 – 69.70 Mb), and 20q13 (Chr 20: 51.32 – 59.12 Mb) previously identified by chromosomal CGH, thus validating the array CGH approach. In addition to known amplicons,

two novel amplicons at 9p13 (chr9: 38.65 – 39.25 Mb) and 17q21.3 (chr17: 52.47 - 55.80 Mb) were identified.

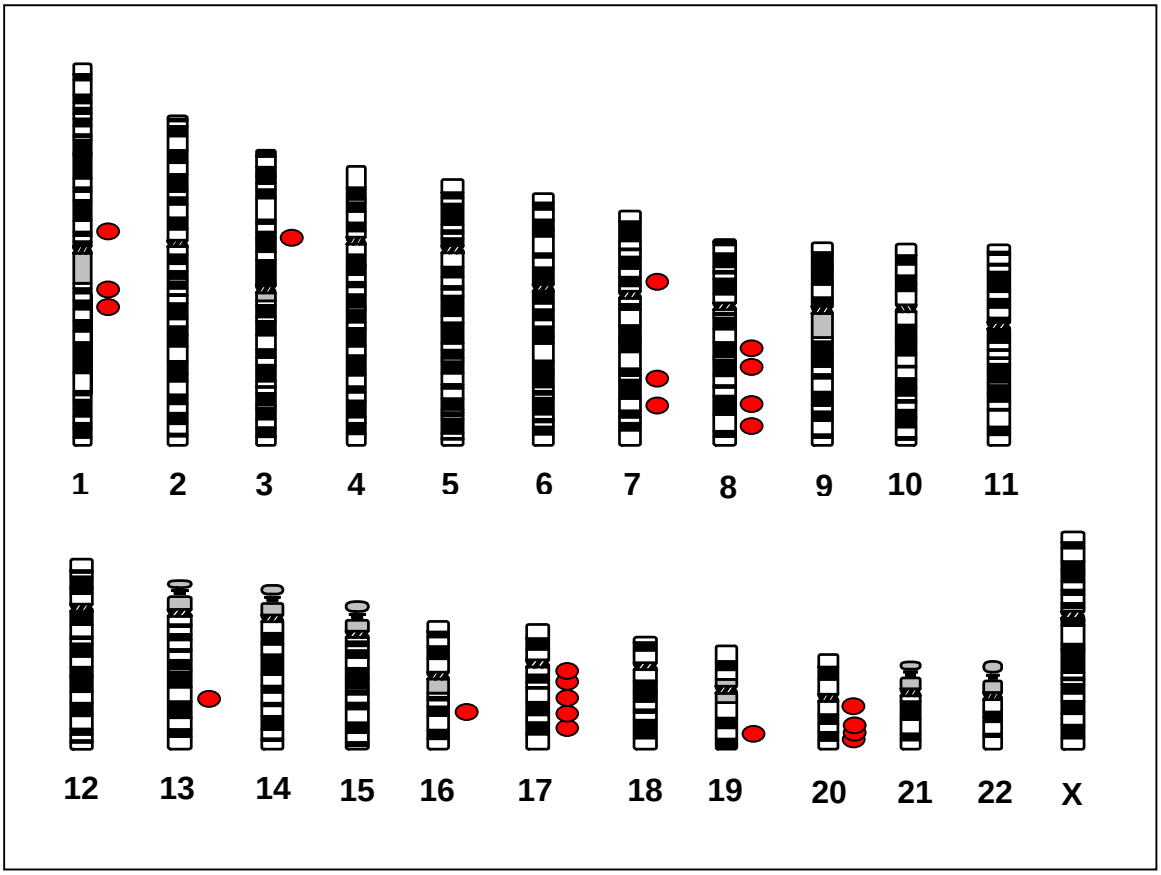

Figure 5. Summary of the chromosomal locations of 24 independent amplicon identified using high-resolution CGH microarray analysis.

4.3 Identification of 270 amplified and highly expressed genes (IV)

To find genes whose expression level were most closely linked to increased copy number, a statistical algorithm was used. Comparison of gene copy number and expression levels across the 14 cell lines revealed a list of 270 genes whose expression was significantly influenced by gene copy number. In addition to the previously defined amplification target genes, e.g. HER2 at 17q12, MYC at 8q24, EGFR at 7p11-p12, and RPS6KB1 at 17q23, this list included numerous novel genes whose activation by gene amplification in breast cancer has not previously been reported. One example of this kind of gene was HOXB7 located in the novel previously undescribed amplicon at 17q21.3. FISH to tissue microarray showed that amplification of the HOXB7 gene was present in 10.2 % of 363 primary breast cancers and

amplification was associated with poor prognosis of the patients (Figure 6). These results validate the presence of the novel amplicon, as well as the performance of the statistical algorithm. In order to evaluate the possible functional roles of the 270 genes, publicly available gene ontology information (http://www.geneontology.org/) was used. Based on the gene ontology data, 91 genes represented hypothetical proteins or genes with no functional annotation, whereas the remaining 179 genes had associated functional information available. The majority of these genes (151/179, 84%) were involved in cellular processes e.g. apoptosis, cell proliferation, signal transduction and transcription, that could be directly linked to cancer.

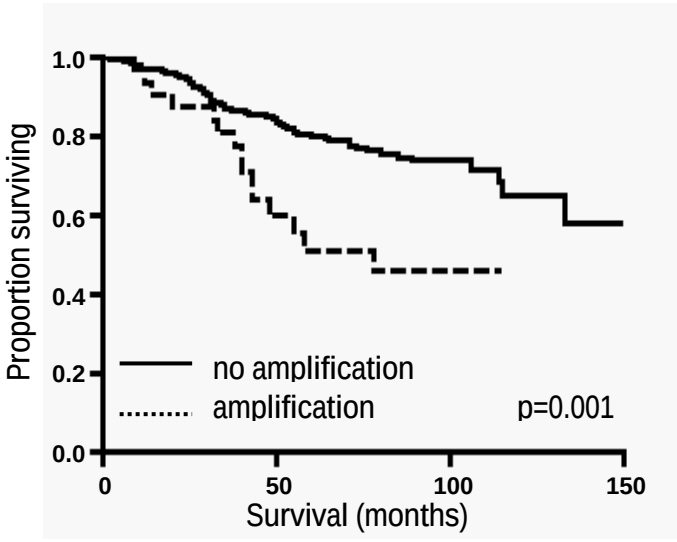

Figure 6. Kaplan-Meier survival analysis of the HOXB7 amplification in 363 primary breast tumors.

DISCUSSION

1. Cell lines versus primary tumors in breast cancer research

Cell lines are widely used as *in vitro* models in cancer research. The use of cell lines offers many advantages. First of all, they are easy to handle and, due to their unlimited replication potential, endless resources of DNA and RNA are guaranteed. In addition, they are easy to store as multiple frozen stocks from which culture can be started at any time, e.g. in case of contamination. Many studies have compared the characteristics of breast cancer cell lines with those of primary breast tumors in order to determine whether the cell lines provide a representative model system (Wistuba et al., 1998; Forozan et al., 1999; Forozan et al., 2000; Larramendy et al., 2000). Good concordance has been observed in terms of different cellular and genetic features (e.g. cell morphology, estrogen and progesterone receptor status, HER2 and p53 status, and allelic loss) (Wistuba et al., 1998). Studies using comparative genomic hybridization have shown that most chromosomal aberrations were the same in cell lines and primary tumors (Forozan et al., 2000; Larramendy et al., 2000). However, cell lines displayed additional novel alterations compared to primary tumors (Forozan et al., 2000). Such changes are most likely due to genotypic and phenotypic drift occurring during the continual culture of the cells. In this thesis, breast cancer cell lines where used in Studies I, III and IV. Amplicon mapping of the 17q12 region in breast cancer cell lines (I) and in primary tumors (II) showed that the amplicon structure was very similar in both cases, confirming the usefulness of cell lines as an *in vitro* model of breast cancer.

2. Two approaches for the identification of amplification target genes

Traditionally identification of amplification target genes has followed conventional positional cloning strategies (Xu and Li, 2000). These strategies are based on the initial identification of a region of interest. For example, information from cytogenetic or CGH analyses is used to select a target chromosomal location. Subsequent efforts are directed at the construction of a complete physical map of the selected region. The physical map typically consists of overlapping large-insert-size genomic clones (e.g. YACs, BACs, PACs, and P1s). The

physical map serves two different purposes. It provides additional tools for further refinement of the region of interest, i.e. the minimal region of amplification, and is the basis for the identification and localization of transcribed sequences that represent putative amplification target genes. The construction of physical maps is tedious and time-consuming. Fortunately, due to the Human Genome Project (Fey, 2002), complete physical and transcript maps are currently electronically available for all human chromosomes, making physical mapping efforts no longer necessary. Despite these advances, the definition of a minimal common region of amplification is still critical and usually means that a large set of tumors must be evaluated. Tissue microarray (TMA) technology (Kononen et al., 1998), which allows simultaneous analysis of hundreds or even thousands of tumor samples, has greatly facilitated such efforts. However, the analysis of the TMA data is still very time-consuming and labor-intensive. Once the minimal common region of amplification has been defined, expression levels of all transcripts within this region have to be evaluated in order to identify putative target genes.

DNA microarray technology provides another approach for the identification of amplification target genes. Utilization of the array based CGH technique (Pollack et al., 1999) enables direct correlation of gene copy number information with gene expression data, because exactly the same cDNA microarray can be used in both experiments. Recently, oligonucleotide microarrays have been applied for copy number analysis (Lucito et al., 2003) thus providing another platform for parallel copy number and expression analysis. In principle, cDNA and oligonucleotide microarrays can be constructed to contain all approximately 30,000 human genes, thus providing genome-wide coverage. When the expression and copy number data is linked to the genomic position information, it is possible to identify all amplified chromosomal regions and genes as well as to directly extract those genes showing increased expression within the amplicons. Thus the combination of two microarray-based technologies enables direct correlation of copy number and expression data on a gene-by-gene basis throughout the genome. This combination is likely to considerably facilitate the identification of putative amplification target genes in cancer.

3. Putative target genes of the 17q12 amplicon (I, II)

Comprehensive characterization of the molecular consequences of the 17q12 amplification in breast cancer was performed using both a cDNA microarray-based strategy and convensional positional cloning. A parallel expression and copy number analysis of the amplified region was performed using a chromosome 17-specific cDNA microarray. The cDNA microarray was constructed based on the information available in the radiation hybrid map (http://www.ncbi.nlm.nih.gov/genemap) and contained all obtainable transcripts from the 17q12 region as well as all known genes from the entire chromosome 17. Combination of the expression and copy number analyses provided a powerful tool for the direct identification of amplified and overexpressed genes in the 17q12 amplicon and revealed a small set of putative amplification target genes. However, it has to be noted that the selection of the clones for the cDNA microarray, which was based on the radiation hybrid map, constitutes a clear limitation to the analysis. First of all, all genes are not present in the radiation hybrid maps and secondly, the order and position of genes is sometimes inaccurate. Thus, the chromosome 17-specific cDNA microarray analysis provided an initial characterization of the extent of the amplified region, but did not present a complete characterization of the genes involved.

In order to obtain more comprehensive information on the extent of the 17q12 amplicon in breast cancer and the actual physical location and order of genes within this region, a physical mapping effort based on the information obtained from the chromosome-17 specific cDNA microarray was initiated. Amplicon mapping using 11 BAC clones located within a ~13 Mb region at 17q12 was performed in 16 breast cancer cell lines using interphase FISH. Based on the results from the cell line analyses, five BAC clones located within a 2 Mb region were selected for further copy number analysis in primary breast tumors. The amplicon mapping implicated a common minimal region of amplification in breast cancer cell lines and primary breast tumors defined by a single BAC clone (2019C10). According to the human genome databases, this clone spans only 280 kb of genomic DNA and contains ten transcripts. Expression analysis of these ten transcripts demonstrated a strong, statistically significant correlation between amplification and expression levels for four known genes (ERBB2, GRB7, MLN64, PNMT) and two genes coding for hypothetical proteins (MGC9753, MGC14832). These results indicated that, like several other amplicons in human tumors (Hui et al., 1997; Bekri et al., 1997; Wu et al., 2001; Bärlund et al., 2000; Anzick et al., 1997; Sen

et al., 1997, Collins et al., 1998), amplification of the 17q12 region in breast cancer leads to simultaneous activation of multiple genes. However, on the contrary to e.g. 11q13 amplification, where several independent amplification target genes e.g cyclinD1, INT2, and EMS1 have been proposed (reviewed by Ormandy et al., 2003), amplification of the 17q12 region has never been reported to occur without amplification of the ERBB2 gene itself.

The GRB7 and MLN64 genes have previously been shown to be co-amplified and overexpressed with ERBB2 (Stein et al., 1994; Tomasetto et al., 1995; Bieche et al., 1996) and based on their function, they have been suggested to have a role in cancer. GRB7 codes for an SH2-domain containing growth factor receptor tyrosine kinase that transmits intracellular signals. Recent studies have suggested that GRB7 is involved in cell invasion and migration and might thus have a role in metastasis (Tanaka et al., 1998; Shen et al., 2002). MLN64 (also known as STARTD3 and CAB1) shares significant homology with the steroidogenic acute regulatory protein (STAR), which plays a key role in steroid hormone biosynthesis (Moog-Lutz et al., 1997). Overexpression of MLN64 could facilitate steroid hormone production in cancer cells, which might affect the growth of hormone-responsive tumors such as breast cancer (Akiyama et al., 1997). PNMT is a terminal enzyme of the catecholamine pathway catalyzing the synthesis of epinephrine (adrenaline) from norepinephrine (Hoehe et al., 1992). PNMT was recently shown to be overexpressed in breast cancer (Dressman et al., 2003), but otherwise its possible role in breast cancer is still unknown.

In addition to known genes, two uncharacterized hypothetical proteins (MGC9753 and MGC14832) were identified as potential target genes of the 17q12 amplification in this study. The function of MGC14832 is currently unknown, because it shows no sequence similarity to any other known genes or proteins. However, MGC9753, also known as CAB2, was recently shown to encode a seven-transmembrane receptor with extracellular N-terminal six-cysteine domain (Katoh and Katoh, 2003). However, the exact location of the MGC9753 in the cell has not been determined. Localization of MGC9753 in plasma membrane would make it an accessible target for various anti-cancer therapies and thus an interesting candidate target gene for breast cancer. Additional studies to evaluate the possible significance of these genes are clearly warranted and could include transfection studies to overexpress the gene, silencing of the gene expression e.g. by siRNA, as well as protein level studies to determine protein expression, subcellular localization and possible interacting protein partners.

Over the years, multiple genes have been reported to be co-amplified with ERBB2 (van de Vijver et al., 1987, Keith et al., 1993; Stein et al., 1994; Tomasetto et al., 1995; Bièche et al., 1996; Zhu et al., 1999). However, comprehensive characterization of the 17q12 amplicon structure and consequences of the amplification on gene expression levels have not been reported before. Most of the previously published co-amplified genes, such as the thyroid hormone receptor α (THRA1, also known as ERBA), retinoid acid receptor α (RARA), topoisomerase II α (TOP2A), peroxisome proliferator-activated receptor binding protein (PPARBP) as well as the MLN50, MLN51, and MLN62 genes, are located outside the minimal common region of amplification defined in Studies I and II (Figure 7). The presence of these multiple co-amplified genes is likely to be explained by the fact that in most cases, as demonstrated in Studies I and II, the amplicon at the 17q12 region is large and continuous.

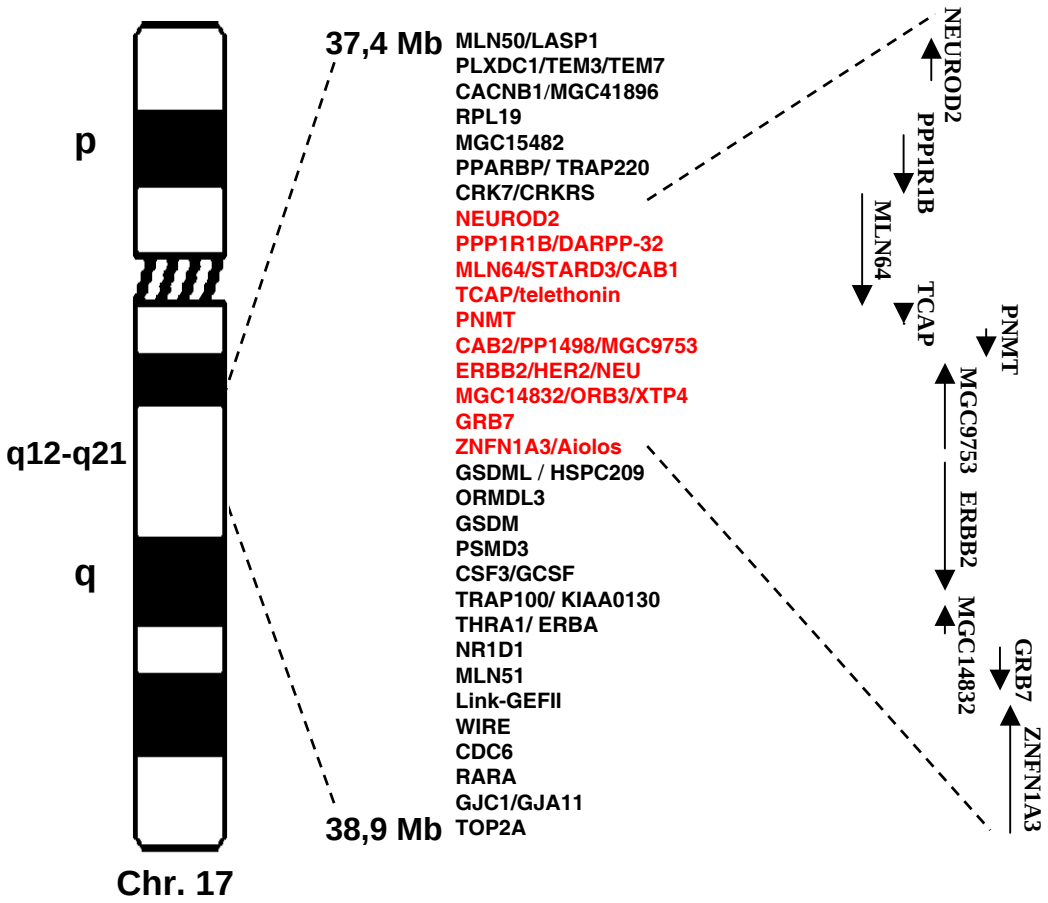

Figure 7. Schematic representation of the location of known genes within a 1.5 Mb region at the ERBB2 amplicon at 17q12-q21 according to NCBI human genome database (http://www.ncbi.nlm.nih.gov/mapview). Genes mapping to the 280 kb minimal region identified in study I and II are shown in red. Orientation of the minimal region transcripts is indicated with arrows.

Other recent studies have also attempted to identify novel amplified genes from the 17q12 region (Clark et al., 2002; Luoh, 2002; Dressman et al., 2003). However, none of these studies have evaluated a minimal region of 17q12 amplification or performed systematic characterization of all the transcripts in this region. Clark and co-workers (2002) utilized a combination of CGH and cDNA microarray analysis, where the clones for the microarray were randomly selected from a cDNA library prepared from BT-474 breast cancer cell line containing the 17q12 amplicon. This approach was used to ensure the presence of novel genes instead of preselected known genes on the array (Clark et al., 2002). With this technique Clark and co-workers were able to identify three (TRAP100, PSMD3, HSPC209) amplified and overexpressed genes from the 17q12 region (Clark et al., 2002). All these genes are located outside the minimal region of amplification defined in Studies I and II, TRAP100 being fartest away (~190 kb) and HSPC209 being nearest (~60 kb).

In a study by Luoh (2002), a >300 kb commonly amplified segment of the 17q12 amplicon containing thirteen genes was identified. Multiplex RT-PCR analysis performed in six breast cancer cell lines showed a near-perfect correlation between amplification and overexpression with eleven (CrkRS, DARPP32, PNMT, TCP, Hs.91668 = MGC9753, CAB1, HER2, GRB7, Aiolos, PSMD3 and TRAP100) genes (Luoh, 2002). These included eight of the ten genes located in the minimal region of amplification defined in Studies I and II. The remaining three genes located outside the minimal region contained TRAP100 and PSMD3 genes also identified by Clark and co-workers and CrkRS gene, which is located in close proximity to minimal region (Luoh, 2002). Identification of the genes with similar expression patterns to ERBB2 in 34 primary breast tumors using cDNA microarray revealed a set of 20 genes from which six genes (RPL19, NR1d1, PSMD3, PNMT, MLN64, GRB7) mapped to the same chromosomal location as ERBB2. Gene copy number analysis with two of these genes (PNMT, MLN64) in 12 tumors confirmed that co-amplification with ERBB2 statistically correlated with increased expression (Dressman et al., 2003), which was in agreement with the results of Study II.

4. Herceptin-induced temporal gene expression changes in breast cancer (III)

Herceptin is a humanized recombinant monoclonal antibody targeted against the extracellular domain of the ERBB2 oncogene. Several studies have indicated that it provides effective treatment in a subset of metastatic breast tumors with ERBB2 overexpression (Baselga et al., 1996, Cobleigh et al., 1999; Slamon et al., 2001; Vogel et al., 2002). However, for an unknown reason not all ERBB2 positive tumors show a favorable response. Thus, it is important to understand effects of Herceptin treatment. To this end, expression analysis by cDNA microarray was performed in ERBB2 amplified and non-amplified cell lines in order to explore the effects of Herceptin treatment on gene expression levels. As expected based on the literature, ERBB2 amplified and non-amplified cell lines showed different responses to Herceptin treatment in growth inhibition assays, ERBB2 amplified cells being more sensitive to the drug. Interestingly, the response seemed to be correlated with the ERBB2 expression level, which was shown in the case of the MDA-453 cell line that contains low-level amplification of ERBB2 leading to only a modest increase in ERBB2 expression. Like ERBB2 non-amplified cell lines, MDA-453 cells did not show any significant growth suppression effect with the lowest dose (1 μg/ml) of Herceptin. With higher doses a clear growth reduction effect was seen, but there did not seem to be any marked difference in the response between doses of 100 μg/ml and 500 μg/ml of Herceptin. These observations might be explained by the fact that the low expression level of ERBB2 leads to the production of fewer copies of the receptor than high-level expression, resulting in earlier saturation of the receptors. These findings might also provide some clues as to why not all ERBB2 positive tumors respond to Herceptin treatment. It may be that Herceptin therapy is useful only for those patients whose tumors display high-level expression of the ERBB2 receptor. Unfortunately all other cell lines, except MDA-453, used in this study contained high-level expression of the ERBB2, thus preventing proper testing of this hypothesis.

Hierarchical clustering of the time-series experiments (24, 48, 72 hours incubation of the drug) indicated that the major effect of Herceptin is obtained between 24 and 48 hours of drug treatment. A combination of statistical methods was used to identify a set of genes that best discriminated the ERBB2 amplified and non-amplified cell lines from each other. This set of 439 genes divided into 212 and 227 genes showing respectively decreasing and increasing expression levels after Herceptin treatment in the amplified vs. non-amplified cell lines. The

discriminatory power of these genes was validated by two clustering methods, hierarchical clustering and self organizing map. Clustering analysis also illustrated the bivalent nature of the MDA-453 cell line, where a large fraction of genes showed expression changes similar to those seen in the non-amplified cell line group.

Publicly available gene ontology information (http://source.stanford.edu/cgi-bin/sourceSearch) was used to obtain functional information of the discriminatory genes. Analysis of the gene ontology data demonstrated that many important cellular processes e.g. transcription, signal transduction, and protein processing were included in the list of functional properties of the discriminatory genes. For example, Herceptin treatment induced the expression of genes encoding proteins involved in RNA processing (e.g. biogenesis of spliceosomal snRNAs; SIP1) and DNA repair (e.g. double-strand break; MRE11) and on the contrary, downregulated genes encoding cell adhesion proteins (e.g. junction plakoglobin, JUP) or well characterized oncoproteins (e.g. FOS and KIT) in ERBB2 amplified cell lines. This study serves as a preliminary effort in the identification of downstream effects of Herceptin treatment. However, further studies must be conducted in order to determine whether the discriminatory genes identified here have any clinical relevance.

The microarray format provides a useful analytical tool to study the effects of various drug treatments (Hoshida et al., 2002; Zhou et al., 2002; Kaneta et al., 2002; Maxwell et al., 2003; Taxman et al., 2003). For example, in the study by Kaneta and co-workers (2002) expression profiling was used to identify 79 genes that were differentially expressed between responders and non-responders of chronic myeloid leukemia patients treated with ST1571. Expression of 15 or 30 of these genes was used to develop a prediction system that could separate responders and non-responders into specific groups (Kaneta et al., 2002). Recently the use of time-series experiments has provided additional information on the temporal processes involved in cell cycle, infection and response to environmental conditions or drug treatment. However, the analysis of the data derived from time-series experiments is difficult due to biological and experimental variation, such as possible differences in sampling rates, timing of the biological processes, and lack of repeats (Bar-Joseph et al., 2003). Thus analytical methods used in static expression experiments cannot be directly applied to time-series experiments, which had led to recent introduction of several algorithms specifically for time-series expression data analysis (Aach and Church, 2001; Filkov et al., 2001; Kasturi et al., 2003; Park et al., 2003; Hautaniemi et al., 2003a).

5. Genome-wide copy number and expression analysis in breast cancer (IV)

In Study IV, high-resolution genome-wide copy number analysis using CGH microarray containing 13,824 clones was performed. Altogether 24 independent amplicons were identified and included all those previously detected to be commonly amplified by chromosomal CGH. In addition, two novel amplicons at 9p13 and 17q21.3 were identified. These have not previously been detected by chromosomal CGH, most likely due to their small size or close proximity to other amplicons. The 9p13 amplicon spans only 600 kb and the 17q21.3 amplicon is located between the 17q12 and 17q22-q23 amplicons. Nowadays, when the Human Genome Project has been completed all approximately 30,000 human genes could be printed on a single microarray slide thus allowing an even more accurate localization of amplicons in cancer.

Parallel gene copy number and expression analysis enabled us to determine globally the impact of gene copy number alterations on gene expression patterns. Our results clearly illustrated that gene amplification had a considerable impact on gene expression levels, both in the case of high- and low-level copy number increases in breast cancer cell lines. However, although gene amplification was shown to be a major determinant of gene expression levels, it did not always lead to overexpression of the gene. Expression analysis also demonstrated that a large proportion of genes are overexpressed due to some other mechanism than gene amplification. Recent studies on acute myeloid leukemia and prostate cancer have shown a similar relationship between gene copy number and gene expression levels in the case of chromosomal aneuploidy, with regions of chromosomal gain showing increased expression and regions of chromosomal loss decreased expression of the genes (Virtaneva et al., 2001; Phillips et al., 2001). The results of Study IV were also in agreement with the study by Pollack and coworkers, showing that changes in DNA copy number had a large direct effect on global gene expression levels both in breast cancer cell lines and in primary breast tumors (Pollack et al., 2002). In contrast, a study analyzing the relationship between gene amplification with gene expression profiles in primary colon tumors and colon cancer cell lines demonstrated that only a very small proportion (3.8 %) of the genes located on the amplified region showed increased expression levels (Platzer et al., 2002). Such contradictory results may be partly due to the different methods used in these studies. In the study by Platzer and co-workers amplified regions were mapped by chromosomal CGH technique, the resolution of which is limited to 10-20 Mb, thus making it difficult to detect small amplicons

and define the exact boundaries of the amplified regions (Platzer et al., 2002). By contrast in Study IV and in the study by Pollack et al. the copy number and expression comparison was made on a gene-by-gene basis using the combination of CGH and cDNA microarray (Pollack et al., 2002).

In addition to the general global analysis of the influence of copy number on gene expression, we carried out a more systematic search in order to identify genes whose expression levels were most affected by amplification status across the 14 cell lines. This search, performed using a statistical analysis with random permutation tests, identified a set of 270 genes (representing 2 % of the all genes printed on the array) whose expression was most clearly linked to gene copy number throughout the whole data set. Among these 270 genes were previously described amplified and overexpressed genes e.g. ERBB2, MYC, EGFR, S6K and AIB3, which confirmed the relevance of the set. In addition to previously known amplification target genes the set included a large number of genes that have not previously been implicated in breast cancer and can therefore provide information of the novel pathogenic mechanisms of breast cancer development. For example, the HOXB7 gene, located in the novel amplicon at 17q21.3, was shown to be amplified in 10.2 % of 363 primary breast cancers and amplification was associated with poor prognosis of the patients. In a recent study, amplification of the EMSY gene at 11q13.5 was shown to be associated with worse survival in node negative breast cancer, providing another example of a novel amplified gene with prognostic signifigance (Hughes-Davies et al., 2003). Functional analysis of the 270 genes provided very interesting results, with 84% of the genes having available functional information being involved in cellular processes (e.g. apoptosis, cell proliferation, signal transduction) that could be directly linked to cancer progression. Although it is not assumed that all of these genes have important and independent roles in tumorigenesis, further characterization of these genes is clearly warranted. This type of comprehensive list characterizing the putative amplification target genes throughout the genome in breast or any other cancer has not previously been published. The possible role of these genes in breast cancer progression and in the development of novel therapeutic drugs remains to be seen.

SUMMARY AND CONCLUSIONS

The main purpose of this study was to identify novel amplified and overexpressed genes in breast cancer and to explore the transcriptional effects of Herceptin treatment on breast cancer cells.

Comprehensive evaluation of the ERBB2 amplicon (at 17q12) in breast cancer cell lines and primary breast tumors revealed a consistent amplicon structure in breast cancer. Amplicon mapping using large-insert-size BAC clones and interphase FISH defined a minimal common region of amplification restricted to a single BAC clone 2019C10. Based on the information on human genome databases, the minimal region of amplification spans approximately 280 kb of genomic DNA and contains ten transcripts. Expression analysis in primary breast tumors using real time RT-PCR showed that six of these genes, four known genes (ERBB2, GRB7, PNMT, MLN64) and two uncharacterized hypothetical proteins (MGC9753, MGC14832), display a statistically highly significant association between gene copy number and expression levels. The possible function of these genes in breast cancer needs to be ascertained.

Growth suppression assays using increasing concentrations of Herceptin demonstrated a dose-dependent growth reduction in breast cancer cell lines with ERBB2 amplification whereas non-amplified cell lines were practically resistant. The effects of Herceptin treatment on gene expression levels in ERBB2 amplified and non-amplified cell lines were evaluated by comparing genome-wide expression levels in Herceptin treated vs. untreated cells. A combination of statistical methods was used to identify a set of 439 genes whose temporal gene expression patterns differed most significantly between the ERBB2 amplified and non-amplified cell lines. Both hierarchical and self-organizing map clustering methods clearly distinguished the amplified and non-amplified cell lines, confirming the validity of the discriminatory gene set. Functional information of these genes showed that the majority of them are involved in key cellular processes like transcription, signal transduction, protein processing, cell metabolism, and transport. Detailed characterization of these genes could provide new insights into the treatment of ERBB2-positive breast cancer.

Genome-wide high-resolution copy number analysis in 14 breast cancer cell lines using CGH on cDNA microarray identified a total of 24 independent amplicons. In addition to more precise localization of previously known amplicons, two novel amplicons at 9p13 and 17q21.3 were identified. Parallel gene copy number and expression analysis was performed to determine the global effects of gene copy number alterations on gene expression levels. The results obtained demonstrate that amplification has a substantial effect on gene expression levels, with 44% of the highly amplified genes showing increased expression level. In addition, statistical comparison of gene copy number and gene expression levels led to the identification of 270 genes whose expression was most tightly correlated with gene amplification. This list included all previously characterized amplification target genes and several genes that have not previously been implicated in breast cancer. Interestingly, gene ontology information showed that 84 % of these genes were involved in cellular processes, such as apoptosis, cell proliferation, signal transduction, and transcription that could be directly linked to cancer. Therefore, this set of genes is likely to have an active role in breast cancer pathogenesis.

ACKNOWLEDGEMENTS

This study was carried out in the Laboratory of Cancer Genetics, Institute of Medical Technology, University of Tampere and Tampere University Hospital, and at the Cancer Genetics Branch, National Human genome Research Institute (NHGRI), National Institutes of Health (NIH), USA, during the years 2000-2003.

My most profound gratitude is due to my supervisor, Docent Anne Kallioniemi, M.D., Ph.D., for providing me with such an interesting research project and for her continous support and encouragement over the years. It has really been a pleasure to work with someone with such wide knowledge and expertise in molecular biology and medicine. I also want to thank Anne for her friendship and sense of humour. I feel Anne is simply the best supervisor a young scientist could dream of. I gratefully acknowledge Professor Olli-Pekka Kallioniemi, M.D., Ph.D., for providing the research facilities at the beginning of the project. Olli-Pekka is also warmly thanked for his enthusiastic and inspirating guidance during my stay at the NIH.

I wish to express my sincere thanks to the Heads of the Department of Clinical Chemistry, Docent Erkki Seppälä, M.D., Ph.D., and the Institute of Medical Technology, Professor Olli Silvennoinen, M.D., Ph.D., for the excellent research facilities.

Professor Päivi Peltomäki, M.D., Ph.D., University of Helsinki and Professor Åke Borg, Ph.D., University of Lund, are warmly thanked for their valuable comments on this thesis.

The work of Ms. Virgiana Mattila, M.A. for carefully reviewing the language of this manuscript is gratefully acknowledged.

I am also gratefull to all my co-authors Maarit Bärlund, M.D., Ph.D., Outi Monni, M.D., Ph.D., Tuula Kuukasjärvi, M.D., Ph.D., Professor Guido Sauter, M.D., Sampsa Hautaniemi, Dtech, Reija Autio, M.Sc., Professor Jaakko Astola, Ph.D., Abdel Elkahloun, Ph.D., Elizabeth Hyman M.D., Maija Wolf, Ph.D., Spyro Mousses, Ph.D., Ester Rozenblum, Ph.D., and Professor Markus Ringnér, Ph.D., for their valuable contribution.

I cordially thank Professor Tapio Visakorpi, M.D., Ph.D., for his support and help over the years. I also wish to express warm thanks to Professor Jorma Isola, M.D., Ph.D., and Docent Minna Tanner, M.D., Ph.D., for introducing me to the fascinating world of science and for their guidance in my early steps as a researcher.

I express my sincere thanks to Ms. Kati Rouhento for her skillful and invaluable help in the laboratory, as well as her friendship over the years. I also want to thank Mrs. Arja Alkula "Raaka-Arska eli Raakkis" for her friendship and excellent assistance in the cell culture.

Warm thanks are due to all the members of the Anne´s lab. Jenita, Emma-Leena and Ritva "Ritu-Pitu" are especially thanked for their encouragement and support during the writing of this manuscript. I want to express special thanks to Maarit for her guidance at the beginning of this project and her unfailing support and friendship over the years. Ritva is also warmly thanked for her tremendous help in several different practical issues.

Thanks from the bottom of my heart to all the dear friends I met during my stay at the NIH. Kaisa and Minna "Allisen Akka" are thanked for being such excellent roommates. I also want to thank Minna for all the fun times we have shared together. Piia, Maija, Sampsa, and Outi are thanked for their friendship both in science and in private life, likewise all my other friends at the NIH, Elizabeth, Spyro, Ester, Natalie, Don, Galen, and Robbie, for the friendly atmosphere in the lab.

I also want to thank all the close friends in the old and present Laboratory of Cancer Genetics. My special thanks are due to my dearest friends, Annika Rökman, M.Sc., for organizing my free-time and being the ultimate party companion, and Kati Porkka "Porkkelis", Ph.D., and Marika Linja, "Linjuri", Ph.D., for their valuable friendship over the years.

All the members of "Projekti 2005" Outi Kauppinen, Hanna-Leena "Hantta" Mäkinen and Kati Malmi are thanked for all the fun we have shared together.

Thanks also to my best friends Noora Kotaja, Ph.D., Maritta Ilmonen, M.Sc., and Piia Karisola, M.Sc., for all the joyful excursions we have made around the world and the wonderful times we have shared together. Daily e-mails have been invaluable lifeline all these years.

I would like to thank my beloved parents, my mother Anja and my father Tapio, for their never-ending love, support and belief in me throughout my life. I also want to thank my dear brother Pasi and her wife Annika, and their wonderful children Sanni and Matias, for all the great times together. Your family has shown me what really matters in life.

Finally, with all my heart, I would like to thank Iivari for sharing all the good and bad times with me and for his continous love and support over the years.

This study was financially supported by the Foundation for Finnish Cancer Institute, the Finnish Cancer Society, the Finnish and Pirkanmaa Cultural Foundations, the Pirkanmaa Cancer Society, the Ida Montin Foundation, the Emil Aaltonen Foundation, the Medical Research Fund of Tampere University Hospital, the Finnish Breast Cancer Group, the Research and Science Foundation of Farmos, the Foundation for the Development of Laboratory Medicine, and the Scientific Foundation of the City of Tampere.

Tampere, January 2004

Päivikki Kauraniemi

REFERENCES

Aach J and Church GM (2001): Aligning gene expression time series with time warping algorithms. Bioinformatics 17:495-508.

Ahr A, Karn T, Solbach C, Seiter T, Strebhardt K, Holtrich U and Kaufmann M (2002): Identification of high risk breast-cancer patients by gene expression profiling. Lancet 359:131-132.

Akiyama N, Sasaki H, Ishizuka T, Kishi T, Sakamoto H, Onda M, Hirai H, Yazaki Y, Sugimura T and Terada M (1997): Isolation of a candidate gene, CAB1, for cholesterol transport to mitochondria from the c-ERBB-2 amplicon by a modified cDNA selection method. Cancer Res 57:3548-3553.

Albertson DG, Ylstra B, Segraves R, Collins C, Dairkee SH, Kowbel D, Kuo WL, Gray JW and Pinkel D (2000): Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. Nat Genet 25:144-146.

Albertson D (2003): Profiling breast cancer by array CGH. Breast Cancer Res Treat 78:289-298.

Albertson DG and Pinkel D (2003): Genomic microarrays in human genetic disease and cancer. Hum Mol Gen 12:145-152.

Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO and Staudt LM (2000): Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403:503-511.

Alroy I and Yarden Y (1997): The ErbB signaling network in embryogenesis and oncogenesis: signal diversification through combinatorial ligand-receptor interactions. FEBS Lett 410:83-86.

Andersen CL, Hostetter G, Grigoryan A, Sauter G and Kallioniemi A (2001): Improved procedure for fluorescence in situ hybridization on tissue microarrays. Cytometry 45:83-86.

Anzick SL, Kononen J, Walker RL, Azorsa DO, Tanner MM, Guan XY, Sauter G, Kallioniemi OP, Trent JM and Meltzer PS (1997): AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. Science 277:965-968.

Balmain A, Gray J and Ponder B (2003): The genetics and genomics of cancer. Nat Genet 33:238-244.

Barczak A, Rodriguez MW, Hanspers K, Koth LL, Tai YC, Bolstad BM, Speed TP and Erle DJ (2003): Spotted long oligonucleotide arrays for human gene expression analysis. Genome Res 13:1775-1785.

Bar-Joseph Z, Gerber G, Simon I, Gifford DK and Jaakkola TS (2003): Comparing the continuous representation of time-series expression profiles to identify differentially expressed genes. Proc Natl Acad Sci U S A 100:10146-10151.

Baselga J, Tripathy D, Mendelsohn J, Baughman S, Benz CC, Dantis L, Sklarin NT, Seidman AD, Hudis CA, Moore J, Rosen PP, Twaddell T, Henderson IC and Norton L (1996): Phase II study of weekly intravenous recombinant humanized anti-p185HER2 monoclonal antibody in patients with HER2/neu-overexpressing metastatic breast cancer. J Clin Oncol 14:737-744.

Baselga J, Norton L, Albanell J, Kim YM and Mendelsohn J (1998): Recombinant humanized anti-HER2 antibody (Herceptin) enhances the antitumor activity of paclitaxel and

doxorubicin against HER2/neu overexpressing human breast cancer xenografts. Cancer Res 58:2825-2831.

Baselga J, Tripathy D, Mendelsohn J, Baughman S, Benz CC, Dantis L, Sklarin NT, Seidman AD, Hudis CA, Moore J, Rosen PP, Twaddell T, Henderson IC and Norton L (1999): Phase II study of weekly intravenous trastuzumab (Herceptin) in patients with HER2/neu-overexpressing metastatic breast cancer. Semin Oncol 26:78-83.

Baselga J (2001): Phase I and II clinical trials of trastuzumab. Ann Oncol 12:S49-55.

Baulida J, Kraus MH, Alimandi M, Di Fiore PP and Carpenter G (1996): All ErbB receptors other than the epidermal growth factor receptor are endocytosis impaired. J Biol Chem 271:5251-5257.

Beerli RR and Hynes NE (1996): Epidermal growth factor-related peptides activate distinct subsets of ErbB receptors and differ in their biological activities. J Biol Chem 271:6071-6076.

Bekri S, Adelaide J, Merscher S, Grosgeorge J, Caroli-Bosc F, Perucca-Lostanlen D, Kelley PM, Pebusque MJ, Theillet C, Birnbaum D and Gaudray P (1997): Detailed map of a region commonly amplified at 11q13-->q14 in human breast carcinoma. Cytogenet Cell Genet 79:125-131.

Berns EM, Foekens JA, van Staveren IL, van Putten WL, de Koning HY, Portengen H and Klijn JG (1995): Oncogene amplification and prognosis in breast cancer: relationship with systemic treatment. Gene 159:11-18.

Berry DA, Muss HB, Thor AD, Dressler L, Liu ET, Broadwater G, Budman DR, Henderson IC, Barcos M, Hayes D and Norton L (2000): HER-2/neu and p53 expression versus tamoxifen resistance in estrogen receptor-positive, node-positive breast cancer. J Clin Oncol 18:3471-3479.

Bieche I, Tomasetto C, Regnier CH, Moog-Lutz C, Rio MC and Lidereau R (1996): Two distinct amplified regions at 17q11-q21 involved in human primary breast cancer. Cancer Res 56:3886-3890.

Bittner M, Meltzer P, Chen Y, Jiang Y, Seftor E, Hendrix M, Radmacher M, Simon R, Yakhini Z, Ben-Dor A, Sampas N, Dougherty E, Wang E, Marincola F, Gooden C, Lueders J, Glatfelter A, Pollock P, Carpten J, Gillanders E, Leja D, Dietrich K, Beaudry C, Berens M, Alberts D and Sondak V (2000): Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature 406:536-540.

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

Brennan PJ, Kumagai T, Berezov A, Murali R, Greene MI and Kumogai T (2000): HER2/neu: mechanisms of dimerization/oligomerization. Oncogene 19:6093-6101.

Burke HB, Hoang A, Iglehart JD and Marks JR (1998): Predicting response to adjuvant and radiation therapy in patients with early stage breast carcinoma. Cancer 82:874-877.

Burstein HJ, Kuter I, Campos SM, Gelman RS, Tribou L, Parker LM, Manola J, Younger J, Matulonis U, Bunnell CA, Partridge AH, Richardson PG, Clarke K, Shulman LN and Winer EP (2001): Clinical activity of trastuzumab and vinorelbine in women with HER2-overexpressing metastatic breast cancer. J Clin Oncol 19:2722-2730.

Bärlund M, Monni O, Kononen J, Cornelison R, Torhorst J, Sauter G, Kallioniemi OP and Kallioniemi A (2000): Multiple genes at 17q23 undergo amplification and overexpression in breast cancer. Cancer Res 60:5340-5344.

Bärlund M, Monni O, Kononen J, Cornelison R, Torhorst J, Sauter G, Kallioniemi OP and Kallioniemi A (2000) Multiple genes at 17q23 undergo amplification and overexpression in breast cancer. Cancer Res 60:5340-5344.

Cai WW, Mao JH, Chow CW, Damani S, Balmain A and Bradley A (2002): Genome-wide detection of chromosomal imbalances in tumors using BAC microarrays. Nat Biotechnol 20:393-396.

Carlomagno C, Perrone F, Gallo C, De Laurentiis M, Lauria R, Morabito A, Pettinato G, Panico L, D'Antonio A, Bianco AR and De Placido S (1996): c-erb B2 overexpression decreases the benefit of adjuvant tamoxifen in early-stage breast cancer without axillary lymph node metastases. J Clin Oncol 14:2702-2708.

Carter P, Presta L, Gorman CM, Ridgway JB, Henner D, Wong WL, Rowland AM, Kotts C, Carver ME and Shepard HM (1992): Humanization of an anti-p185HER2 antibody for human cancer therapy. Proc Natl Acad Sci U S A 89:4285-4289.

Chen Y, Dougherty ER and Bittner ML (1997): Ratio-based decisions and the quantitative analysis of cDNA microarray images. J Biomed Optics 2:364-374.

Clahsen PC, van de Velde CJ, Duval C, Pallud C, Mandard AM, Delobelle-Deroide A, van den Broek L, Sahmoud TM and van de Vijver MJ (1998): p53 protein accumulation and response to adjuvant chemotherapy in premenopausal women with node-negative early breast cancer. J Clin Oncol 16:470-479.

Clark GM and McGuire WL (1991): Follow-up study of HER-2/neu amplification in primary breast cancer. Cancer Res 51:944-948.

Clark J, Edwards S, John M, Flohr P, Gordon T, Maillard K, Giddings I, Brown C, Bagherzadeh A, Campbell C, Shipley J, Wooster R and Cooper CS (2002): Identification of amplified and expressed genes in breast cancer by comparative hybridization onto microarrays of randomly selected cDNA clones. Genes Chromosomes Cancer 34:104-114.

Clark J, Edwards S, Feber A, Flohr P, John M, Giddings I, Crossland S, Stratton MR, Wooster R, Campbell C and Cooper CS (2003): Genome-wide screening for complete genetic loss in prostate cancer by comparative hybridization onto cDNA microarrays. Oncogene 22:1247-1252.

Cleveland W (1979): Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 74:829-836.

Cobleigh MA, Vogel CL, Tripathy D, Robert NJ, Scholl S, Fehrenbacher L, Wolter JM, Paton V, Shak S, Lieberman G and Slamon DJ (1999): Multinational study of the efficacy and safety of humanized anti-HER2 monoclonal antibody in women who have HER2-overexpressing metastatic breast cancer that has progressed after chemotherapy for metastatic disease. J Clin Oncol 17:2639-2648.

Collins C, Rommens JM, Kowbel D, Godfrey T, Tanner M, Hwang SI, Polikoff D, Nonet G, Cochran J, Myambo K, Jay KE, Froula J, Cloutier T, Kuo WL, Yaswen P, Dairkee S, Giovanola J, Hutchinson GB, Isola J, Kallioniemi OP, Palazzolo M, Martin C, Ericsson C, Pinkel D, Albertson D, Li WB and Gray JW (1998): Positional cloning of ZNF217 and NABC1: genes amplified at 20q13.2 and overexpressed in breast carcinoma. Proc Natl Acad Sci U S A 95:8703-8708.

Cooke T, Reeves J, Lanigan A and Stanton P (2001): HER2 as a prognostic and predictive marker for breast cancer. Ann Oncol 12:S23-28.

Coquelle A, Pipiras E, Toledo F, Buttin G, and Debatisse M (1997): Expression of fragile sites triggers intrachromosomal mammalian gene amplification and sets boundaries to early amplicons. Cell 89:215-225.

Coquelle A, Rozier L, Dutrillaux B and Debatisse M (2002): Induction of multiple double-strand breaks within an hsr by meganucleaseI-SceI expression or fragile site activation leads to formation of double minutes and other chromosomal rearrangements. Oncogene 21:7671 – 7679.

De Placido S, De Laurentiis M, Carlomagno C, Gallo C, Perrone F, Pepe S, Ruggiero A, Marinelli A, Pagliarulo C, Panico L, Pettinato G, Petrella G and Bianco AR (2003): Twenty-year results of the Naples GUN randomized trial: predictive factors of adjuvant tamoxifen efficacy in early breast cancer. Clin Cancer Res 9:1039-1046.

DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA and Trent JM (1996): Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat Genet 14:457-460.

Dhanasekaran SM, Barrette TR, Ghosh D, Shah R, Varambally S, Kurachi K, Pienta KJ, Rubin MA and Chinnaiyan AM (2001): Delineation of prognostic biomarkers in prostate cancer. Nature 412:822-826.

Dowsett M (2001): Overexpression of HER-2 as a resistance mechanism to hormonal therapy for breast cancer. Endocr Relat Cancer 8:191-195.

Dressman MA, Baras A, Malinowski R, Alvis LB, Kwon I, Walz TM and Polymeropoulos MH (2003): Gene expression profiling detects gene amplification and differentiates tumor types in breast cancer. Cancer Res 63:2194-2199.

Duggan DJ, Bittner M, Chen Y, Meltzer P and Trent JM (1999): Expression profiling using cDNA microarrays. Nat Genet. 21:10-14.

Eiermann W (2001): Trastuzumab combined with chemotherapy for the treatment of HER2-positive metastatic breast cancer: pivotal trial data. Ann Oncol 12:S57-62.

Eisen M, Spellman P, Brown P and Botstein D (1998): Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95: 14863-14868.

Elledge RM, Green S, Ciocca D, Pugh R, Allred DC, Clark GM, Hill J, Ravdin P, O'Sullivan J, Martino S and Osborne CK (1998): HER-2 expression and response to tamoxifen in estrogen receptor-positive breast cancer: a Southwest Oncology Group Study. Clin Cancer Res 4:7-12.

Fearon ER and Vogelstein B (1990): A genetic model for colorectal tumorigenesis. Cell 61:759-767.

Fentiman IS (2001): Fixed and modifiable risk factors for breast cancer. Int J Clin Pract 55:527-530.

Ferrero-Pous M, Hacene K, Bouchet C, Le Doussal V, Tubiana-Hulin M and Spyratos F (2000): Relationship between c-erbB-2 and other tumor characteristics in breast cancer prognosis. Clin Cancer Res 6:4745-4754.

Fey MF (2002): Impact of the Human Genome Project on the clinical management of sporadic cancers. Lancet Oncol 3:349-356.

Fidler IJ and Kripke ML (1977): Metastasis results from preexisting variant cells within a malignant tumor. Science 197:893-895.

Fiegler H, Carr P, Douglas EJ, Burford DC, Hunt S, Scott CE, Smith J, Vetrie D, Gorman P, Tomlinson IP and Carter NP (2003): DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. Genes Chromosomes Cancer 36:361-374.

Filkov V, Skiena S and Zhi J (2002): Analysis techniques for microarray time-series data. J Comput Biol 9:317-330.

Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT and Solas D (1991): Light-directed, spatially addressable parallel chemical synthesis. Science 251:767-773.

Forozan F, Veldman R, Ammerman CA, Parsa NZ, Kallioniemi A, Kallioniemi OP and Ethier SP (1999): Molecular cytogenetic analysis of 11 new breast cancer cell lines. Br J Cancer 81:1328-1334.

Forozan F, Mahlamaki EH, Monni O, Chen Y, Veldman R, Jiang Y, Gooden GC, Ethier SP, Kallioniemi A and Kallioniemi OP (2000): Comparative genomic hybridization analysis

of 38 breast cancer cell lines: a basis for interpreting complementary DNA microarray data. Cancer Res 60:4519-4525.

Frederiksen CM, Knudsen S, Laurberg S and Orntoft TF (2003): Classification of Dukes' B and C colorectal cancers using expression arrays. J Cancer Res Clin Oncol 129:263-271.

Fritz B, Schubert F, Wrobel G, Schwaenen C, Wessendorf S, Nessling M, Korz C, Rieker RJ, Montgomery K, Kucherlapati R, Mechtersheimer G, Eils R, Joos S and Lichter P (2002): Microarray-based copy number and expression profiling in dedifferentiated and pleomorphic liposarcoma. Cancer Res 62:2993-2998.

Garinis GA, Patrinos GP, Spanakis NE and Menounos PG (2002): DNA hypermethylation: when tumour suppressor genes go silent. Hum Genet 111:115-127.

Giai M, Roagna R, Ponzone R, De Bortoli M, Dati C and Sismondi P (1994): Prognostic and predictive relevance of c-erbB-2 and ras expression in node positive and negative breast cancer. Anticancer Res 14:1441-1450.

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD and Lander ES (1999): Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286:531-537.

Graus-Porta D, Beerli RR, Daly JM and Hynes NE (1997): ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling. EMBO J 16:1647-1655.

Gusterson BA, Gelber RD, Goldhirsch A, Price KN, Save-Soderborgh J, Anbazhagan R, Styles J, Rudenstam CM, Golouh R and Reed R (1992): Prognostic importance of c-erbB-2 expression in breast cancer. International (Ludwig) Breast Cancer Study Group. J Clin Oncol 10:1049-1056.

Hamilton A, Larsimont D, Paridaens R, Drijkoningen M, van de Vijver M, Bruning P, Hanby A, Houston S, Treilleux I, Guastalla JP, Van Vreckem A, Sylvester R and Piccart M (2000): A study of the value of p53, HER2, and Bcl-2 in the prediction of response to doxorubicin and paclitaxel as single agents in metastatic breast cancer: a companion study to EORTC 10923. Clin Breast Cancer 1:233-240.

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

Hautaniemi S, Kauraniemi P, Rämö P, Yli-Harja O, Astola J and Kallioniemi A (2003a): A strategy for identifying class-separating genes in drug-treatment microarray data. Report 1, Institute of Signal Processing, Tampere University of Technology, Finland.

Hautaniemi S, Yli-Harja O, Astola J, Kauraniemi P, Kallioniemi A, Wolf M, Ruiz J, Mousses S and Kallioniemi OP (2003b): Analysis and visualization of gene expression microarray data in human cancer using self-organizing maps. Machine Learning 52:45-66.

Harari D and Yarden Y (2000): Molecular mechanisms underlying ErbB2/HER2 action in breast cancer. Oncogene 19:6102-6114.

Harris H, Miller OJ, Klein G, Worst P and Tachibana T (1969): Suppression of malignancy by cell fusion. Nature 233:363-368.

Harris LN, Yang L, Liotcheva V, Pauli S, Iglehart JD, Colvin OM and Hsieh TS (2001): Induction of topoisomerase II activity after ErbB2 activation is associated with a differential response to breast cancer chemotherapy. Clin Cancer Res 7:1497-504.

Hedenfalk I, Duggan D, Chen Y, Radmacher M, Bittner M, Simon R, Meltzer P, Gusterson B, Esteller M, Kallioniemi OP, Wilfond B, Borg A and Trent J (2001): Gene-expression profiles in hereditary breast cancer. N Engl J Med 344:539-548.

Hedenfalk I, Ringner M, Ben-Dor A, Yakhini Z, Chen Y, Chebil G, Ach R, Loman N, Olsson H, Meltzer P, Borg A and Trent J (2003): Molecular classification of familial non-BRCA1/BRCA2 breast cancer. Proc Natl Acad Sci U S A 100:2532-2537.

Heintz NH, Leslie KO, Rogers LA and Howard PL (1990): Amplification of the c-erb B-2 oncogene and prognosis of breast adenocarcinoma. Arch Pathol Lab Med 114:160-163.

Hellman A, Zlotorynski E, Scherer SW, Cheung J, Vincent JB, Smith DI, Trakhtenbrot L and Kerem B (2002): A role for common fragile site induction in amplification of human oncogenes. Cancer Cell 1:89-97.

Hieken TJ, Mehta RR and Shilkaitis A (1996): Her-2/neu and p53 expression in breast cancer: valid prognostic markers when assessed by direct immunoassay, but not by immunohistochemistry. Proc Am Soc Clin Oncol 15:113a.

Hoehe MR, Plaetke R, Otterud B, Stauffer D, Holik J, Byerley WF, Baetge EE, Gershon ES, Lalouel JM and Leppert M (1992): Genetic linkage of the human gene for phenylethanolamine N-methyltransferase (PNMT), the adrenaline-synthesizing enzyme, to DNA markers on chromosome 17q21-q22. Hum Mol Genet 1:175-178.

Holloway AJ, van Laar RK, Tothill RW and Bowtell DD (2002): Options available--from start to finish--for obtaining data from DNA microarrays II. Nat Genet 32:481-489.

Hoshida Y, Moriyama M, Otsuka M, Kato N, Goto T, Taniguchi H, Shiratori Y, Seki N and Omata M (2002): Identification of genes associated with sensitivity to 5-fluorouracil and cisplatin in hepatoma cells. J Gastroenterol 37:92-95.

Huang E, Cheng SH, Dressman H, Pittman J, Tsou MH, Horng CF, Bild A, Iversen ES, Liao M, Chen CM, West M, Nevins JR and Huang AT (2003): Gene expression predictors of breast cancer outcomes. Lancet 361:1590-1596.

Huang X, Gollin SM, Raja S and Godfrey TE (2002): High-resolution mapping of the 11q13 amplicon and identification of a gene, TAOS1, that is amplified and overexpressed in oral cancer cells. Proc Natl Acad Sci U S A 99:11369-11374.

Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW, Lefkowitz SM, Ziman M, Schelter JM, Meyer MR, Kobayashi S, Davis C, Dai H, He YD, Stephaniants SB, Cavet G, Walker WL, West A, Coffey E, Shoemaker DD, Stoughton R, Blanchard AP, Friend SH and Linsley PS (2001): Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat Biotechnol 19:342-347.

Hughes-Davies L, Huntsman D, Ruas M, Fuks F, Bye J, Chin SF, Milner J, Brown LA, Hsu F, Gilks B, Nielsen T, Schulzer M, Chia S, Ragaz J, Cahn A, Linger L, Ozdag H, Cattaneo E, Jordanova ES, Schuuring E, Yu DS, Venkitaraman A, Ponder B, Doherty A, Aparicio S, Bentley D, Theillet C, Ponting CP, Caldas C and Kouzarides T (2003): EMSY links the BRCA2 pathway to sporadic breast and ovarian cancer. Cell 115:523-535.

Hui R, Campbell DH, Lee CS, McCaul K, Horsfall DJ, Musgrove EA, Daly RJ, Seshadri R and Sutherland RL (1997): EMS1 amplification can occur independently of CCND1 or INT-2 amplification at 11q13 and may identify different phenotypes in primary breast cancer. Oncogene 15:1617-1623.

Hynes NE, Horsch K, Olayioye MA and Badache A (2001): The ErbB receptor tyrosine family as signal integrators. Endocr Relat Cancer 8:151-159.

Jacquemier J, Penault-Llorca F, Viens P, Houvenaeghel G, Hassoun J, Torrente M, Adelaide J and Birnbaum D (1994): Breast cancer response to adjuvant chemotherapy in correlation with erbB2 and p53 expression. Anticancer Res 14:2773-2778.

Jahanzeb M, Mortimer JE, Yunus F, Irwin DH, Speyer J, Koletsky AJ, Klein P, Sabir T and Kronish L (2002): Phase II trial of weekly vinorelbine and trastuzumab as first-line therapy in patients with HER2(+) metastatic breast cancer. Oncologist 7:410-417.

Jones PA and Baylin SB (2002): The fundamental role of epigenetic events in cancer. Nat Rev Genet 3:415-428.

Jukkola A, Bloigu R, Soini Y, Savolainen ER, Holli K and Blanco G (2001): c-erbB-2 positivity is a factor for poor prognosis in breast cancer and poor response to hormonal or chemotherapy treatment in advanced disease. Eur J Cancer 37:347-354.

Järvinen TAH and Liu ET (2002): Novel cancer therapies targeted at the product of the HER-2/neu gene. Applied Genomics and Proteomics 1:3-14.

Kane MD, Jatkoe TA, Stumpf CR, Lu J, Thomas JD and Madore SJ (2000). Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. Nucleic Acids Res 28:4552-4557.

Kaneta Y, Kagami Y, Katagiri T, Tsunoda T, Jin-nai I, Taguchi H, Hirai H, Ohnishi K, Ueda T, Emi N, Tomida A, Tsuruo T, Nakamura Y and Ohno R (2002): Prediction of sensitivity to STI571 among chronic myeloid leukemia patients by genome-wide cDNA microarray analysis. Jpn J Cancer Res 93:849-856.

Kang Y, Siegel PM, Shu W, Drobnjak M, Kakonen SM, Cordón-Cardo C, Guise TA and Massagué J (2003): A multigenic program mediating breast cancer metastasis to bone Cancer Cell 3:537-549.

Karunagaran D, Tzahar E, Beerli RR, Chen X, Graus-Porta D, Ratzkin BJ, Seger R, Hynes NE and Yarden Y (1996): ErbB-2 is a common auxiliary subunit of NDF and EGF receptors: implications for breast cancer. EMBO J 15:254-264.

Kasturi J, Acharya R and Ramanathan M (2003): An information theoretic approach for analyzing temporal patterns of gene expression. Bioinformatics 19:449-458.

Katoh M and Katoh M (2003): MGC9753 gene, located within PPP1R1B-STARD3-ERBB2-GRB7 amplicon on human chromosome 17q12, encodes the seven-transmembrane receptor with extracellular six-cystein domain. Int J Oncol 22:1369-1374.

Keith WN, Douglas F, Wishart GC, McCallum HM, George WD, Kaye SB and Brown R (1993): Co-amplification of erbB2, topoisomerase II alpha and retinoic acid receptor alpha genes in breast cancer and allelic loss at topoisomerase I on chromosome 20. Eur J Cancer 29A:1469-1475.

Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C and Meltzer PS (2001): Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nat Med 7:673-679.

Kim R, Tanabe K, Uchida Y, Osaki A and Toge T (2002): The role of HER-2 oncoprotein in drug-sensitivity in breast cancer. Oncol Rep 9:3-9.

Kim S, Dougherty ER, Shmulevich L, Hess KR, Hamilton SR, Trent JM, Fuller GN and Zhang W. Identification of combination gene sets for glioma classification. Mol Cancer Ther 1:1229-1236.

Klapper LN, Kirschbaum MH, Sela M and Yarden Y (2000): Biochemical and clinical implications of the ErbB/HER signaling network of growth factor receptors. Adv Cancer Res 77:25-79.

Klos KS, Zhou X, Lee S, Zhang L, Yang W, Nagata Y and Yu D (2003): Combined trastuzumab and paclitaxel treatment better inhibits ErbB-2-mediated angiogenesis in breast carcinoma through a more effective inhibition of Akt than either treatment alone. Cancer 98:1377-1385.

Knudson AG Jr (1971): Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci USA 68:820-823.

Knudson AG (2001): Two genetic hits (more or less) to cancer Nat Rev Cancer 1:157-162.

Knuutila S, Bjorkqvist AM, Autio K, Tarkkanen M, Wolf M, Monni O, Szymanska J, Larramendy ML, Tapper J, Pere H, El-Rifai W, Hemmer S, Wasenius VM, Vidgren V

and Zhu Y (1998): DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies. Am J Pathol 152:1107-1123.

Kohonen T. (ed). (2001). *Self-Organizing Maps*. Springer, 3rd edition, Heidelberg, Germany.

Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G and Kallioniemi OP (1998): Tissue microarrays for high-throughput molecular profiling of tumor specimens. Nat Med 4:844-847.

Kuo MT, Sen S, Hittelman WN and Hsu TC (1998): Chromosomal fragile sites and DNA amplification in drug-resistant cells. Biochem Pharmacol 56:7-13.

Larramendy ML, Lushnikova T, Bjorkqvist AM, Wistuba II, Virmani AK, Shivapurkar N, Gazdar AF and Knuutila S (2000): Comparative genomic hybridization reveals complex genetic changes in primary breast cancer tumors and their cell lines. Cancer Genet Cytogenet 19:132-138.

Lehrmann H, Pritchard LL and Harel-Bellan A (2002): Histone acetyltransferases and deacetylases in the control of cell proliferation and differentiation. Adv Cancer Res 86:41-65.

Lemmon MA and Schlessinger J (1994): Regulation of signal transduction and signal diversity by receptor oligomerization. Trends Biochem Sci 19:459-463.

Lin L, Prescott MS, Zhu Z, Singh P, Chun SY, Kuick RD, Hanash SM, Orringer MB, Glover TW and Beer DG (2000): Identification and characterization of a 19q12 amplicon in esophageal adenocarcinomas reveals cyclin E as the best candidate gene for this amplicon. Cancer Res 60:7021-7027.

Lin L, Miller CT, Contreras JI, Prescott MS, Dagenais SL, Wu R, Yee J, Orringer MB, Misek DE, Hanash SM, Glover TW and Beer DG (2002): The hepatocyte nuclear factor 3 alpha gene, HNF3alpha (FOXA1), on chromosome band 14q13 is amplified and overexpressed in esophageal and lung adenocarcinomas. Cancer Res 62:5273-5279.

Lipshutz RJ, Morris D, Chee M, Hubbell E, Kozal MJ, Shah N, Shen N, Yang R and Fodor SP (1995): Using oligonucleotide probe arrays to access genetic diversity. Biotechniques 19:442-447.

Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H and Brown EL (1996): Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol 14:1675-1680.

Love RR, Duc NB, Havighurst TC, Mohsin SK, Zhang Q, DeMets DL and Allred DC (2003): Her-2/neu overexpression and response to oophorectomy plus tamoxifen adjuvant therapy in estrogen receptor-positive premenopausal women with operable breast cancer. J Clin Oncol 21:453-457.

Lucito R, West J, Reiner A, Alexander J, Esposito D, Mishra B, Powers S, Norton L and Wigler M (2000): Detecting gene copy number fluctuations in tumor cells by microarray analysis of genomic representations. Genome Res 10:1726-1736.

Lucito R, Healy J, Alexander J, Reiner A, Esposito D, Chi M, Rodgers L, Brady A, Sebat J, Troge J, West JA, Rostan S, Nguyen KCQ, Powers S, Ye KQ, Olshen A, Venkatraman E, Norton L and Wigler M (2003): Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. Genome Res 13:2291-2305.

Luoh SW (2002): Amplification and expression of genes from the 17q11 approximately q12 amplicon in breast cancer cells. Cancer Genet Cytogenet 136:43-47.

Marks P, Rifkind RA, Richon VM, Breslow R, Miller T and Kelly WK (2001): Histone deacetylases and cancer: causes and therapies. Nat Rev Cancer 1:194-202.

Martin KJ, Kritzman BM, Price LM, Koh B, Kwan CP, Zhang X, Mackay A, O'Hare MJ, Kaelin CM, Mutter GL, Pardee AB and Sager R (2000): Linking gene expression patterns to therapeutic groups in breast cancer. Cancer Res 60:2232-2238.

Massion PP, Kuo WL, Stokoe D, Olshen AB, Treseler PA, Chin K, Chen C, Polikoff D, Jain AN, Pinkel D, Albertson DG, Jablons DM and Gray JW (2002): Genomic copy number analysis of non-small cell lung cancer using array comparative genomic hybridization: implications of the phosphatidylinositol 3-kinase pathway. Cancer Res 62:3636-3640.

Maxwell PJ, Longley DB, Latif T, Boyer J, Allen W, Lynch M, McDermott U, Harkin DP, Allegra CJ and Johnston PG (2003): Identification of 5-fluorouracil-inducible target genes using cDNA microarray profiling. Cancer Res 63:4602-4606.

Meden H, Beneke A, Hesse T, Novophashenny I and Wischnewsky M (2001): Weekly intravenous recombinant humanized anti-P185HER2 monoclonal antibody (herceptin) plus docetaxel in patients with metastatic breast cancer: a pilot study. Anticancer Res 21:1301-1305.

Menard S, Valagussa P, Pilotti S, Gianni L, Biganzoli E, Boracchi P, Tomasic G, Casalini P, Marubini E, Colnaghi MI, Cascinelli N and Bonadonna G (2001): Response to cyclophosphamide, methotrexate, and fluorouracil in lymph node-positive breast cancer according to HER2 overexpression and other tumor biologic variables. J Clin Oncol 19:329-335.

McPherson K, Steel CM and Dixon JM (2000): ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. BMJ 21:624-628.

Miles DW, Harris WH, Gillett CE, Smith P and Barnes DM (1999): Effect of c-erbB(2) and estrogen receptor status on survival of women with primary breast cancer treated with adjuvant cyclophosphamide/methotrexate/fluorouracil. Int J Cancer 84:354-359.

Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W, Bell R, Rosenthal J, Hussey C, Tran T, McClure M, Frye C, Hattier T, Phelps R, Haugen-Strano A, Katcher H, Yakumo K, Gholami Z, Shaffer D, Stone S, Bayer S, Wray C, Bodgen R, Dayananth P, Ward J, Tonin P, Narod S, Bristow PK, Norris FH, Helvering L, Morrison P, Rosteck P, Lai M, Barrett JC, Lewis C, Neuhausen S, Cannon-Albright L, Goldgar D, Wiseman R, Kamb A and Skolnick MH (1994): A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 266:66-71.

Monni O, Bärlund M., Mousses S, Kononen J, Sauter G, Heiskanen M, Paavola P, Avela K, Chen Y, Bittner ML and Kallioniemi A (2001): Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. Proc Natl Acad Sci U S A 98:5711-5716.

Montemurro F, Choa G, Faggiuolo R, Sperti E, Capaldi A, Donadio M, Minischetti M, Salomone A, Vietti-Ramus G, Alabiso O and Aglietta M (2003): Safety and activity of docetaxel and trastuzumab in HER2 overexpressing metastatic breast cancer: a pilot phase II study. Am J Clin Oncol 26:95-97.

Moog-Lutz C, Tomasetto C, Regnier CH, Wendling C, Lutz Y, Muller D, Chenard MP, Basset P and Rio MC (1997): MLN64 exhibits homology with the steroidogenic acute regulatory protein (STAR) and is over-expressed in human breast carcinomas. Int J Cancer 71:183-191.

Mousses S, Bittner ML, Chen Y, Dougherty ER, Baxevanis A, Meltzer PS and Trent JM (2000): Gene expression analysis by cDNA microarrays. In: F.J. Livesey and S. P. Hunt (eds.), Functional Genomics, pp. 113-137. Oxford: Oxford University Press.

Muss HB, Thor AD, Berry DA, Kute T, Liu ET, Koerner F, Cirrincione CT, Budman DR, Wood WC, Barcos M and Henderson CI (1994): c-erbB-2 expression and response to adjuvant therapy in women with node-positive early breast cancer. N Engl J Med 330:1260-1266.

Muss HB (2001): Role of adjuvant endocrine therapy in early-stage breast cancer. Semin Oncol 28:313-321.

Nephew KP and Huang TH (2003): Epigenetic gene silencing in cancer initiation and progression. Cancer Lett 190:125-133.

Newby JC, Johnston SR, Smith IE and Dowsett M (1997): Expression of epidermal growth factor receptor and c-erbB2 during the development of tamoxifen resistance in human breast cancer. Clin Cancer Res 3:1643-1651.

Niskanen E, Blomqvist C, Franssila K, Hietanen P and Wasenius VM (1997): Predictive value of c-erbB-2, p53, cathepsin-D and histology of the primary tumour in metastatic breast cancer. Br J Cancer 76:917-922.

Nunes RA and Harris LN (2002): The HER2 extracellular domain as a prognostic and predictive factor in breast cancer. Clin Breast Cancer 3:125-135.

Olayioye MA, Graus-Porta D, Beerli RR, Rohrer J, Gay B and Hynes NE (1998): ErbB-1 and ErbB-2 acquire distinct signaling properties dependent upon their dimerization partner. Mol Cell Biol 18:5042-5051.

Olayioye MA, Neve RM, Lane HA and Hynes NE (2000): The ErbB signaling network: receptor heterodimerization in development and cancer. EMBO J 19:3159-3167.

Ormandy CJ, Musgrove EA, Hui R, Daly RJ and Sutherland RL (2003): Cyclin D1, EMS1 and 11q13 amplification in breast cancer. Breast Cancer Res Treat 78:323-335.

Paik S, Bryant J, Park C, Fisher B, Tan-Chiu E, Hyams D, Fisher ER, Lippman ME, Wickerham DL and Wolmark N (1998): erbB-2 and response to doxorubicin in patients with axillary lymph node-positive, hormone receptor-negative breast cancer. J Natl Cancer Inst 90:1361-1370.

Park T, Yi SG, Lee S, Lee SY, Yoo DH, Ahn JI and Lee YS (2003): Statistical tests for identifying differentially expressed genes in time-course microarray experiments. Bioinformatics 19:694-703.

Pawlowski V, Revillion F, Hebbar M, Hornez L and Peyrat JP (2000):. Prognostic value of the type I growth factor receptors in a large series of human primary breast cancers quantified with a real-time reverse transcription-polymerase chain reaction assay. Clin Cancer Res 6:4217-4225.

Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP and Fodor SP (1994): Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc Natl Acad Sci U S A 91:5022-5026.

Pegram MD, Lipton A, Hayes DF, Weber BL, Baselga JM, Tripathy D, Baly D, Baughman SA, Twaddell T, Glaspy JA and Slamon DJ (1998): Phase II study of receptor-enhanced chemosensitivity using recombinant humanized anti-p185HER2/neu monoclonal antibody plus cisplatin in patients with HER2/neu-overexpressing metastatic breast cancer refractory to chemotherapy treatment. J Clin Oncol 16:2659-2671.

Pegram MD and Slamon DJ (1999): Combination therapy with trastuzumab (Herceptin) and cisplatin for chemoresistant metastatic breast cancer: evidence for receptor-enhanced chemosensitivity. Semin Oncol 26:89-95.

Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO and Botstein D (2000): Molecular portraits of human breast tumours. Nature 406:747-752.

Petit T, Borel C, Ghnassia JP, Rodier JF, Escande A, Mors R and Haegele P (2001): Chemotherapy response of breast cancer depends on HER-2 status and anthracycline dose intensity in the neoadjuvant setting. Clin Cancer Res 7:1577-1581.

Phillips JL, Hayward SW, Wang Y, Vasselli J, Pavlovich C, Padilla-Nash H, Pezullo JR, Ghadimi BM, Grossfeld GD, Rivera A, Linehan WM, Cunha GR and Ried T (2001): The consequences of chromosomal aneuploidy on gene expression profiles in a cell line model for prostate carcinogenesis. Cancer Res 61:8143-8149.

Pinkas-Kramarski R, Soussan L, Waterman H, Levkowitz G, Alroy I, Klapper L, Lavi S, Seger R, Ratzkin BJ, Sela M and Yarden Y (1996): Diversification of Neu differentiation factor and epidermal growth factor signaling by combinatorial receptor interactions. EMBO J 15:2452-2467.

Pinkas-Kramarski R, Alroy I and Yarden Y (1997): ErbB receptors and EGF-like ligands: cell lineage determination and oncogenesis through combinatorial signaling. J Mammary Gland Biol Neoplasia 2:97-107.

Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW and Albertson DG (1998): High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet 20:207-211.

Pinto AE, Andre S, Pereira T, Nobrega S and Soares J (2001): C-erbB-2 oncoprotein overexpression identifies a subgroup of estrogen receptor positive (ER+) breast cancer patients with poor prognosis. Ann Oncol 12:525-533.

Plass C (2002): Cancer epigenomics. Hum Mol Genet 11:2479-2488.

Platzer P, Upender MB, Wilson K, Willis J, Lutterbaugh J, Nosrati A, Willson JK, Mack D, Ried T and Markowitz S (2002): Silence of chromosomal amplifications in colon cancer. Cancer Res 62:1134-1138.

Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D and Brown PO (1999): Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat Genet 23:41-46.

Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, Tibshirani R, Botstein D, Borresen-Dale AL and Brown PO (2002): Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proc Natl Acad Sci U S A 99:12963-12968.

Ponder BA (2001): Cancer genetics. Nature 411:336-341.

Prenzel N, Fischer OM, Streit S, Hart S and Ullrich A (2001): The epidermal growth factor receptor family as a central element for cellular signal transduction and diversification. Endocr Relat Cancer 8:11-31.

Pukkala E, Sankila R and Vertio H (1997): Syöpä Suomessa. Suomen Syöpärekisteri, Helsinki.

Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, Ladd C, Reich M, Latulippe E, Mesirov JP, Poggio T, Gerald W, Loda M, Lander ES and Golub TR (2001): Multiclass cancer diagnosis using tumor gene expression signatures. Proc Natl Acad Sci U S A 98:15149-15154.

Ramaswamy S, Ross KN, Lander ES and Golub TR (2003): A molecular signature of metastasis in primary solid tumors. Nat Genet 33:49-54.

Reed W, Hannisdal E, Boehler PJ, Gundersen S, Host H and Marthin J (2000): The prognostic value of p53 and c-erb B-2 immunostaining is overrated for patients with lymph node negative breast carcinoma: a multivariate analysis of prognostic factors in 613 patients with a follow-up of 14-30 years. Cancer 88:804-813.

Relogio A, Schwager C, Richter A, Ansorge W and Valcarcel J (2002): Optimization of oligonucleotide-based DNA microarrays. Nucleic Acids Res 30:e51.

Rice JC and Allis CD (2001): Histone methylation versus histone acetylation: new insights into epigenetic regulation. Curr Opin Cell Biol 13:263-273.

Riese DJ 2nd and Stern DF (1998): Specificity within the EGF family/ErbB receptor family signaling network. Bioessays 20:41-48.

Rosen PP, Lesser ML, Arroyo CD, Cranor M, Borgen P and Norton L (1995): Immunohistochemical detection of HER2/neu in patients with axillary lymph node negative breast carcinoma. A study of epidemiologic risk factors, histologic features,

and prognosis. Cancer 75:1320-1326.
Rosenthal SI, Depowski PL, Sheehan CE and Ross JS (2002): Comparison of HER-2/neu oncogene amplification detected by fluorescence in situ hybridization in lobular and ductal breast cancer. Appl Immunohistochem Mol Morphol 10:40-46.
Ross JS and Fletcher JA (1998): The HER-2/neu oncogene in breast cancer: prognostic factor, predictive factor, and target for therapy. Stem Cells 16:413-428.
Ross JS and Fletcher JA (1999): The HER-2/neu oncogene: prognostic factor, predictive factor and target for therapy. Semin Cancer Biol 9:125-138.
Ross JS, Fletcher JA, Linette GP, Stec J, Clark E, Ayers M, Symmans WF, Pusztai L and Bloom KJ (2003): The Her-2/neu gene and protein in breast cancer: biomarker and target of therapy. Oncologist 8:307-325.
Rozan S, Vincent-Salomon A, Zafrani B, Validire P, De Cremoux P, Bernoux A, Nieruchalski M, Fourquet A, Clough K, Dieras V, Pouillart P and Sastre-Garau X (1998): No significant predictive value of c-erbB-2 or p53 expression regarding sensitivity to primary chemotherapy or radiotherapy in breast cancer. Int J Cancer 79:27-33.
Rubin I and Yarden Y (2001): The basic biology of HER2. Ann Oncol 12:S3-S8.
Sainsbury JR, Anderson TJ and Morgan DA (2000): ABC of breast diseases: breast cancer. BMJ 321:745-750.
Savelyeva L and Schwab M (2001): Amplification of oncogenes revisited: from expression profiling to clinical application. Cancer Lett 167:115-23.
Schmid P, Wischnewsky MB, Sezer O, Bohm R and Possinger K (2002): Prediction of response to hormonal treatment in metastatic breast cancer. Oncology 63:309-316.
Schmidt-Kittler O, Ragg T, Daskalakis A, Granzow M, Ahr A, Blankenstein TJ, Kaufmann M, Diebold J, Arnholdt H, Muller P, Bischoff J, Harich D, Schlimok G, Riethmuller G, Eils R and Klein CA (2003): From latent disseminated cells to overt metastasis: genetic analysis of systemic breast cancer progression. Proc Natl Acad Sci U S A 100:7737-7742.
Schena M, Shalon D, Davis RW and Brown PO (1995): Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467-470.
Schwab M (1999): Oncogene amplification in solid tumors. Semin Cancer Biol 9:319-325.
Sen S, Zhou H and White RA (1997): A putative serine/threonine kinase encoding gene BTAK on chromosome 20q13 is amplified and overexpressed in human breast cancer cell lines. Oncogene. 14:2195-2200.
Sgroi DC, Teng S, Robinson G, LeVangie R, Hudson JR Jr and Elkahloun AG (1999): In vivo gene expression profile analysis of human breast cancer progression. Cancer Res 59:5656-5661.
Shalon D, Smith SJ and Brown PO (1996): A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res 6:639-645
Shedden KA, Taylor JM, Giordano TJ, Kuick R, Misek DE, Rennert G, Schwartz DR, Gruber SB, Logsdon C, Simeone D, Kardia SL, Greenson JK, Cho KR, Beer DG, Fearon ER and Hanash S (2003): Accurate molecular classification of human cancers based on gene expression using a simple classifier with a pathological tree-based framework. Am J Pathol 163:1985-1995.
Shen TL, Han DC and Guan JL (2002): Association of Grb7 with phosphoinositides and its role in the regulation of cell migration. J Biol Chem 277:29069-29077.
Sjogren S, Inganas M, Lindgren A, Holmberg L and Bergh J (1998): Prognostic and predictive value of c-erbB-2 overexpression in primary breast cancer, alone and in combination with other prognostic markers. J Clin Oncol 16:462-469.
Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A and McGuire WL (1987): Human

breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science 235:177-182.

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

Snijders AM, Nowak N, Segraves R, Blackwood S, Brown N, Conroy J, Hamilton G, Hindle AK, Huey B, Kimura K, Law S, Myambo K, Palmer J, Ylstra B, Yue JP, Gray JW, Jain AN, Pinkel D and Albertson DG (2001): Assembly of microarrays for genome-wide measurement of DNA copy number. Nat Genet 29:263-264.

Stark GR (1993): Regulation and mechanisms of mammalian gene amplification. Adv Cancer Res 61: 87-113.

Stein D, Wu J, Fuqua SA, Roonprapunt C, Yajnik V, D'Eustachio P, Moskow JJ, Buchberg AM, Osborne CK and Margolis B (1994): The SH2 domain protein GRB-7 is co-amplified, overexpressed and in a tight complex with HER2 in breast cancer. EMBO J 13:1331-1340.

Stern DF (2000): Tyrosine kinase signalling in breast cancer: ErbB family receptor tyrosine kinases. Breast Cancer Res 2:176-183.

Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Dohner H, Cremer T and Lichter P (1997): Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. Genes Chromosomes Cancer 20:399-407.

Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lonning P and Borresen-Dale AL (2001): Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 98:10869-10874.

Squire JA, Pei J, Marrano P, Beheshti B, Bayani J, Lim G, Moldovan L and Zielenska M. High-resolution mapping of amplifications and deletions in pediatric osteosarcoma by use of CGH analysis of cDNA microarrays. Genes Chromosomes Cancer 38:215-225.

Stål O, Sullivan S, Wingren S, Skoog L, Rutqvist LE, Carstensen JM and Nordenskjold B (1995): c-erbB-2 expression and benefit from adjuvant chemotherapy and radiotherapy of breast cancer. Eur J Cancer 31A:2185-2190.

Su AI, Welsh JB, Sapinoso LM, Kern SG, Dimitrov P, Lapp H, Schultz PG, Powell SM, Moskaluk CA, Frierson HF Jr and Hampton GM (2001): Molecular classification of human carcinomas by use of gene expression signatures. Cancer Res 61:7388-7393.

Takahashi M, Yang XJ, Sugimura J, Backdahl J, Tretiakova M, Qian CN, Gray SG, Knapp R, Anema J, Kahnoski R, Nicol D, Vogelzang NJ, Furge KA, Kanayama H, Kagawa S and Teh BT (2003): Molecular subclassification of kidney tumors and the discovery of new diagnostic markers Oncogene 22:6810-6818.

Tanaka S, Mori M, Akiyoshi T, Tanaka Y, Mafune K, Wands JR and Sugimachi K (1998): A novel variant of human Grb7 is associated with invasive esophageal carcinoma. J Clin Invest 102:821-827.

Tay ST, Leong SH, Yu K, Aggarwal A, Tan SY, Lee CH, Wong K, Visvanathan J, Lim D, Wong WK, Soo KC, Kon OL and Tan P (2003): A combined comparative genomic hybridization and expression microarray analysis of gastric cancer reveals novel molecular subtypes. Cancer Res 63:3309-3316.

Taxman DJ, MacKeigan JP, Clements C, Bergstralh DT and Ting JP (2003): Transcriptional profiling of targets for combination therapy of lung carcinoma with paclitaxel and mitogen-activated protein/extracellular signal-regulated kinase kinase inhibitor. Cancer Res. 2003 63:5095-5104.

Telenius H, Carter NP, Bebb CE, Nordenskjold M, Ponder BA and Tunnacliffe A (1992): Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. Genomics 13:718-725.

Thor AD, Berry DA, Budman DR, Muss HB, Kute T, Henderson IC, Barcos M, Cirrincione C, Edgerton S, Allred C, Norton L and Liu ET (1998): erbB-2, p53, and efficacy of adjuvant therapy in lymph node-positive breast cancer. J Natl Cancer Inst 90:1346-1360.

Todd R and Wong DT (1999): Oncogenes. Anticancer Res 19:4729-4746.

Tomasetto C, Regnier C, Moog-Lutz C, Mattei MG, Chenard MP, Lidereau R, Basset P and Rio MC (1995): Identification of four novel human genes amplified and overexpressed in breast carcinoma and localized to the q11-q21.3 region of chromosome 17. Genomics 28:367-376.

Tzahar E, Waterman H, Chen X, Levkowitz G, Karunagaran D, Lavi S, Ratzkin BJ and Yarden Y (1996): A hierarchical network of interreceptor interactions determines signal transduction by Neu differentiation factor/neuregulin and epidermal growth factor. Mol Cell Biol 16:5276-5287.

Tzahar E, Pinkas-Kramarski R, Moyer JD, Klapper LN, Alroy I, Levkowitz G, Shelly M, Henis S, Eisenstein M, Ratzkin BJ, Sela M, Andrews GC and Yarden Y (1997): Bivalence of EGF-like ligands drives the ErbB signaling network. EMBO J 16:4938-4950.

Van de Vijver M, van de Bersselaar R, Devilee P, Cornelisse C, Peterse J and Nusse R (1987): Amplification of the neu (c-erbB-2) oncogene in human mammary tumors is relatively frequent and is often accompanied by amplification of the linked c-erbA oncogene. Mol Cell Biol 7:2019-2023.

Van de Vijver MJ, Peterse JL, Mooi WJ, Wisman P, Lomans J, Dalesio O and Nusse R (1988):. Neu-protein overexpression in breast cancer. Association with comedo-type ductal carcinoma in situ and limited prognostic value in stage II breast cancer. N Engl J Med 319:1239-1245.

Van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH and Bernards R (2002): A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347:1999-2009.

Van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R and Friend SH. Gene expression profiling predicts clinical outcome of breast cancer. Nature 415:530-536.

Varis A, Wolf M, Monni O, Vakkari ML, Kokkola A, Moskaluk C, Frierson H Jr, Powell SM, Knuutila S, Kallioniemi A and El-Rifai W (2002): Targets of gene amplification and overexpression at 17q in gastric cancer. Cancer Res 62:2625-2629.

Veltman JA, Fridlyand J, Pejavar S, Olshen AB, Korkola JE, DeVries S, Carroll P, Kuo WL, Pinkel D, Albertson D, Cordon-Cardo C, Jain AN and Waldman FM (2003): Array-based comparative genomic hybridization for genome-wide screening of DNA copy number in bladder tumors. Cancer Res 63:2872-2880.

Vesanto J, Himberg J, Alhoniemi E and Parhankangas J (2000): *SOM toolbox for Matlab 5. Technical report A57,* Helsinki University of Technology, Finland.

Virtaneva K, Wright FA, Tanner SM, Yuan B, Lemon WJ, Caligiuri MA, Bloomfield CD, de La Chapelle A and Krahe R (2001): Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. Proc Natl Acad Sci U S A 98:1124-1129.

Vogel C, Cobleigh MA, Tripathy D, Gutheil JC, Harris LN, Fehrenbacher L, Slamon DJ, Murphy M, Novotny WF, Burchmore M, Shak S and Stewart SJ (2001): First-line, single-agent Herceptin(trastuzumab) in metastatic breast cancer: a preliminary report. Eur J Cancer 37:S25-29.

Vogel CL, Cobleigh MA, Tripathy D, Gutheil JC, Harris LN, Fehrenbacher L, Slamon DJ, Murphy M, Novotny WF, Burchmore M, Shak S, Stewart SJ and Press M (2002): Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. J Clin Oncol 20:719-726.

Wahl GM (1989): The importance of circular DNA in mammalian gene amplification. Cancer Res 15: 1333-1340.

Weiss MM, Kuipers EJ, Postma C, Snijders AM, Siccama I, Pinkel D, Westerga J, Meuwissen SG, Albertson DG and Meijer GA (2003): Genomic profiling of gastric cancer predicts lymph node status and survival. Oncogene 22:1872-1879.

West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R, Zuzan H, Olson JA Jr, Marks JR and Nevins JR (2001): Predicting the clinical status of human breast cancer by using gene expression profiles. Proc Natl Acad Sci U S A 98:11462-11467.

Wilhelm M, Veltman JA, Olshen AB, Jain AN, Moore DH, Presti, Jr. JC, Kovacs G and Waldman FM (2002): Array-based comparative genomic hybridization for the differential diagnosis of renal cell cancer. Cancer Res 62:957-960.

Windle BE and Wahl GM (1992): Molecular dissection of mammalian gene amplification: new mechanistic insights revealed by analyses of very early events. Mutat Res 276:199-224.

Wistuba II, Behrens C, Milchgrub S, Syed S, Ahmadian M, Virmani AK, Kurvari V, Cunningham TH, Ashfaq R, Minna JD and Gazdar AF (1998): Comparison of features of human breast cancer cell lines and their corresponding tumors. Clin Cancer Res 4:2931-2938.

Wittwer CT, Ririe KM, Andrew RV, David DA, Gundry RA and Balis UJ (1997): The LightCycler: a microvolume multisample fluorimeter with rapid temperature control. Biotechniques 22:176-181.

Woelfle U, Cloos J, Sauter G, Riethdorf L, Janicke F, van Diest P, Brakenhoff R and Pantel K (2003): Molecular signature associated with bone marrow micrometastasis in human breast cancer. Cancer Res 63:5679-5684.

Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C and Micklem G (1995): Identification of the breast cancer susceptibility gene BRCA2. Nature 378:789-792.

Worthylake R, Opresko LK and Wiley HS (1999): ErbB-2 amplification inhibits down-regulation and induces constitutive activation of both ErbB-2 and epidermal growth factor receptors. J Biol Chem 274:8865-8874.

Wu G, Sinclair C, Hinson S, Ingle JN, Roche PC and Couch FJ (2001): Structural analysis of the 17q22-23 amplicon identifies several independent targets of amplification in breast cancer cell lines and tumors. Cancer Res 61:4951-4955.

Xu L and Li Y (2000): Positional cloning. Methods Mol Biol 136:285-296.

Yang Y, Dudoit S, Luu P and Speed T (2001): *Microarrays: Optical Technologies and Informatics*. Bittner M, Chen Y, Dorsel A and Dougherty E (eds.). SPIE, Society for Optical Engineering, San Jose, CA, pp 141-152.

Yarden Y and Sliwkowski M (2001): Untangling the ErbB signaling network. Nat Rev 2:127-137.

Yeang CH, Ramaswamy S, Tamayo P, Mukherjee S, Rifkin RM, Angelo M, Reich M, Lander E, Mesirov J and Golub T (2001): Molecular classification of multiple tumor types. Bioinformatics 17:S316-22.

Yu D and Hung MC (2000): Overexpression of ErbB2 in cancer and ErbB2-targeting strategies. Oncogene 19:6115-6121.

Zafarana G, Grygalewicz B, Gillis AJ, Vissers LE, van de Vliet W, van Gurp RJ, Stoop H, Debiec-Rychter M, Oosterhuis JW, van Kessel AG, Schoenmakers EF, Looijenga LH and Veltman JA. 12p-amplicon structure analysis in testicular germ cell tumors of adolescents and adults by array CGH. Oncogene 22:7695-7701.

Zhou Y, Gwadry FG, Reinhold WC, Miller LD, Smith LH, Scherf U, Liu ET, Kohn KW, Pommier Y and Weinstein JN (2002): Transcriptional regulation of mitotic genes by camptothecin-induced DNA damage: microarray analysis of dose- and time-dependent effects. Cancer Res 62:1688-1695.

Zhu Y, Qi C, Jain S, Le Beau MM, Espinosa R 3rd, Atkins GB, Lazar MA, Yeldandi AV, Rao MS and Reddy JK (1999): Amplification and overexpression of peroxisome proliferator-activated receptor binding protein (PBP/PPARBP) gene in breast cancer. Proc Natl Acad Sci U S A 96:10848-10853.

ORIGINAL COMMUNICATIONS