MERVI JUMPPANEN

Basal Cytokeratins and HER-2 Oncogene 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 auditorium of
Mediwest Health Technology Center, Koskenalantie 16, Seinäjoki,
on February 28th, 2007, at 12 o'clock.

(Simultaneous video conference connection
in the small auditorium of Building K,
Medical School of the University of Tampere,
Teiskontie 35, Tampere)

UNIVERSITY OF TAMPERE

ACADEMIC DISSERTATION
University of Tampere, Institute of Medical Technology
Seinäjoki Central Hospital, Department of Pathology
Tampere University Hospital
Tampere Graduate School in Biomedicine and Biotechnology (TGSBB)
Finland

Supervised by
Professor Jorma Isola
University of Tampere

Reviewed by
Docent Outi Monni
University of Helsinki
Docent Anita Naukkarinen
University of Kuopio

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

Tel. +358 3 3551 6055
Fax +358 3 3551 7685
taju@uta.fi
www.uta.fi/taju
http://granum.uta.fi

Cover design by
Juha Siro

Printed dissertation
Acta Universitatis Tamperensis 1208
ISBN 978-951-44-6847-6 (print)
ISSN 1455-1616

Electronic dissertation
Acta Electronica Universitatis Tamperensis 591
ISBN 978-951-44-6848-3 (pdf)
ISSN 1456-954X
http://acta.uta.fi

Tampereen Yliopistopaino Oy – Juvenes Print
Tampere 2007

Tyvisolukon sytokeratiinit ja HER-2-syöpägeeni rintasyövässä

Rintasyöpä on monimuotoinen sairaus, joka voidaan jaotella useisiin biologisesti ja kliinisesti toisistaan poikkeaviin alaryhmiin. Epiteelin tyvisolukon tukirankaproteiineja (sytokeratiineja) ilmentävä ns. basaalityypin rintasyöpä on yksi näistä alaryhmistä. Se on useimmiten hormonaaliselle lääkehoidolle reagoimaton (hormonireseptorinegatiivinen) ja solukon mikroskooppisten piirteiden perusteella huonosti erilaistunut. Toinen tärkeä rintasyövän alaryhmä on HER-2-syöpägeeniä yli-ilmentävät kasvaimet, joissa on kyseisen solukalvolla sijaitsevan epidermaalista kasvutekijäreseptoria koodaavan geenin mutaatio (geenimonistuma).

Tässä tutkimuksessa kehitettiin yksinkertainen immunohistokemiallinen tunnistusmenetelmä basaalityypin rintasyövälle. Menetelmän avulla selvitettiin kyseisen rintasyöpätyypin yleisyyttä ja tyyppipiirteitä sekä taudin ennustetta. Sporadisista eli satunnaisesti ilmenevistä invasiivisista tiehytperäisistä karsinoomista noin 10 % oli basaalityyppiä. Basaalityypin rintasyöpä oli yleisempi perinnöllisissä kasvaimissa, joissa on syövälle altistava BRCA1-rintasyöpägeenin mutaatio. Näistä jopa 78 % ilmensi tyvisolukon sytokeratiineja. Kaikki sporadiset basaalityypin rintasyövät ilmensivät erittävälle rauhasepiteelille tyypillisiä luminaalisia sytokeratiineja, kun taas osassa perinnöllisistä basaalityypin rintasyövistä näitä sytokeratiineja ei esiintynyt lainkaan. Tämän lisäksi basaalityypin rintasyövän osoitettiin ilmentävän usein vimentiini-tukirankaproteiinia, c-kit-tyrosiinikinaasireseptoria**,** p53-kasvurajoitegeeniä ja epidermaalista kasvutekijäreseptoria (EGFR). Mielenkiintoista oli se, että vaikka basaalityypin kasvaimessa saattaa olla HER-2-geenimonistuma, nämä ominaisuudet esiintyivät pääasiassa erillisinä.

Tutkimuksessa saatiin selville, että basaalityypin kasvaimet ovat tyypillisesti aggressiivisia, histologisesti huonosti erilaistuneita, hormonireseptorinegatiivisia ja nopeasti jakautuvia. Tutkimuksessa saatujen tulosten perusteella basaalityypin kasvaimilla on taipumus uusiutua aikaisessa vaiheessa. Kuitenkaan basaalityypin rintasyövän ennuste ei eroa muiden

estrogeenireseptorinegatiivisten kasvainten ennusteesta. Tästä huolimatta komplementaarisella DNA sirutekniikalla (cDNA microarray) saatujen tulosten perusteella basaalityypin rintasyöpä eroaa biologisesti muista estrogeenireseptorinegatiivisista rintasyövistä.

Tutkimuksessa osoitettiin ensimmäistä kertaa, että basaalityypin rintasyövät jakautuvat kahdeksi erilliseksi alaryhmäksi, joilla on erilaiset biologiset ja ennusteelliset ominaisuudet. Basoluminaalisiksi nimetyt kasvaimet ilmentävät tyvisolukon sytokeratiineja vain osassa kasvainsoluista, kun taas basaalisiksi nimetyt rintasyövät ilmentävät näitä sytokeratiineja kauttaaltaan lähes kaikissa syöpäsoluissa. Nopea jakautumisaktiivisuus sekä vimentiini- ja c-kit-proteiinien ilmentyminen olivat merkittävästi yleisempiä basaalisessa syöpäalatyypissä. Basaalisen syöpäalatyypin ennuste oli myös parempi kuin basoluminaalisen. Merkittävää oli se, että lähes kaikki basaalityypin kasvaimet, joissa oli HER-2-geenimonistuma, olivat basoluminaalista alatyyppiä.

Tämän tutkimuksen yhtenä tavoitteena oli kehittää HER-2-geenimonistuman tunnistamiseksi kaksivärinen kromogeeninen in situ hybridisaatio -menetelmä (dc-CISH), jossa sekä HER-2-syöpägeeni että kromosomi 17 tunnistetaan samanaikaisesti. Menetelmällä saatavat tulokset olivat yhteneväisiä Yhdysvaltain elintarvike- ja lääkeviraston (FDA) hyväksymän fluoresenssi in situ hybridisaatio -menetelmän kanssa. Dc-CISH on uusi geenikopiolukujen testaamiseen soveltuva menetelmä, joka on yksinkertainen ja samalla luotettava valomikroskooppisen tulkinnan mahdollistava tekniikka.

CONTENTS

LIST OF ORIGINAL COMMUNICATIONS

This thesis is based on the following articles referred to in the text by their Roman numerals:

I Laakso M, Loman N, Borg Å, and Isola J (2005). Cytokeratin CK5/14-positive breast cancer: true basal phenotype confined to BRCA1 tumors. Modern Pathology 18:1321-1328.

II Jumppanen M*, Gruvberger-Saal S*, Kauraniemi P, Tanner M, Bendahl P-O, Lundin M, Krogh M, Kataja P, Borg Å, Fernö M, and Isola J (2006). Basal-like phenotype is not associated with patient survival in estrogen receptor negative breast cancers. Breast Cancer Res, in press.

III Laakso M, Tanner M, and Isola J (2006). Dual-colour chromogenic in situ hybridization for testing of *HER-2* oncogene amplification in archival breast tumours. J Pathol 210:3-9.

IV Laakso M, Tanner M, Nilsson J, Wiklund T, Erikstein B, Kellokumpu-Lehtinen P, Malmström P, Wilking N, Bergh J, and Isola J (2006). Basoluminal carcinoma: a new biologically and prognostically distinct entity between basal and luminal breast cancer. Clin Cancer Res 12:4185-4191.

* These authors contributed equally.

ABBREVIATIONS

Bcl-2	B-cell lymphoma/leukemia-2
BRCA1	breast cancer susceptibility gene 1
BRCA2	breast cancer susceptibility gene 2
CD	cluster of differentiation
cDNA	complementary DNA
CEACAM1	Carcinoembryonic antigen-related cell adhesion molecule 1
CGH	comparative genomic hybridization
CISH	chromogenic in situ hybridization
CK	cytokeratin
c-kit	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
CTCb	high-dose cyclophosphamide, thiotepa, and carboplatin supported by autologous bone marrow support
DAB	diaminobenzidine
dc-CISH	dual-colour chromogenic in situ hybridization
dUTP	deoxyuridine triphosphate
EASE	expression analysis systematic explorer
EDTA	ethylene diamine tetraacetic acid
EGFR	epidermal growth factor receptor
EMA	epithelial membrane antigen
ER	estrogen receptor
ERBB2	avian erythroblastic leukemia viral oncogene homolog 2
ESA	epithelial specific antigen
EVA1	Epithelial V-like antigen 1
FDA	food and drug administration
FDR	false discovery rate
FEC	5-fluorouracil, epirubicin, and cyclophosphamide
FISH	fluorescent in situ hybridization
GATA-3	GATA binding protein 3
HER	human epidermal growth factor receptor
ID4	inhibitor of DNA binding 4
IHC	immunohistochemistry
Ki-67	cell cycle related nuclear protein
mRNA	messenger ribonucleic acid
p21	protein 21
p53	protein 53
p63	protein 63
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PR	progesterone receptor
RT	room temperature
SBG 9401	Swedish Breast Cancer Group trial 9401
Skp-2	S-phase kinase-associated protein 2
SLC2A1	solute carrier family 2
SSC	standard saline citrate
TDLU	terminal ductal-lobular unit
TMA	tissue microarray
TMB	tetramethyl benzidine
TNM	tumor, node, metastasis

TRIM29	tripartite motif-containing 29
TTF1	trefoil factor 1
XBP1	X-box binding protein 1

ABSTRACT

Breast cancer is no longer regarded as a single disease group but it represents a variety of subgroups with unique biological and clinical characteristics. Basal phenotype breast cancers form one of the subgroups with a unique gene expression profile and cytokeratin expression pattern. Another important breast cancer subgroup determinant is HER-2 oncogene amplification, which is both a prognostic and predictive factor in breast cancer.

In this study, basal phenotype breast cancer was studied by both immunohistochemistry and cDNA microarrays. Immunohistochemical identification of basal phenotype breast cancer was conducted using a triple antibody cocktail of p63 and basal cytokeratins 5 and 14. Since there was no association between p63 and basal cytokeratin expression, breast cancer was regarded as basal phenotype if basal cytokeratin expression was present. The basal phenotype was seen in ~10% of sporadic invasive ductal, sporadic HER-2 amplified, and hereditary BRCA2 germ-line mutated tumors, while BRCA1 germ-line mutated tumors showed 78% prevalence of the basal phenotype. Sporadic basal phenotype tumors always expressed luminal cytokeratins, which supports the progenitor cell origin of basal phenotype tumors. In contrast, a small number of CK5/14-positive BRCA1 germ-line mutated hereditary tumors lacked luminal cytokeratin expression.

The most prominent features associated with basal phenotype breast cancer were histologic grade III and steroid hormone receptor negative status. Other features characteristic of basal phenotype breast cancers were high proliferation activity, immunopositivity for vimentin, c-kit, p53, EGFR immunopositivity, and negativity for Bcl-2. Additionally, this study showed that even if HER-2 amplification and the basal phenotype can co-exist, they are inversely associated within the ER-negative tumor subgroup. Interestingly, this study showed for the first time that basal phenotype tumors are divided into two microscopically distinguishable tumor subtypes, which differ in the biological features typical of basal phenotype tumors. "Basoluminal" tumors show

heterogeneous basal cytokeratin expression in lower proportion of the tumor cells while "basal" subtype tumors stain uniformly with basal cytokeratins. High proliferation activity and c-kit and vimentin immunopositivity were seen more frequently in the basal than in the basoluminal tumor subtype. Conversely, basoluminal tumors were larger in size than basal tumors and included almost all of the HER-2 amplified CK5/14 expressing tumors.

When basal phenotype tumors were studied as one group for prognosis, it was shown that they have significantly worse survival estimates during the first years of follow-up than the basal cytokeratin negative tumors. Although, when survival was studied within the ER-negative subgroup, no adverse survival was seen at any stage for basal phenotype tumors identified either by immunohistochemistry or cDNA microarrays. Still, the basal phenotype tumors differ biologically from other ER-negative tumors since they display a unique gene expression profile within the ER-negative tumor subgroup. Further, the newly defined basal phenotype breast cancer subtypes differed in prognosis. The basoluminal subtype tumors had worse survival estimates than the basal subtype breast cancers. This difference was not explained by more frequent HER-2 amplification in the basoluminal subtype.

In this study a dual-color chromogenic in situ hybridization (dc-CISH) method was developed for the simultaneous detection of probes against the HER-2 oncogene and the chromosome 17 centromere. Distinct probe detection was achieved by using two enzymes and a combination of green and red chromogens. The HER-2/chomosome 17 centromere ratios obtained by dc-CISH were highly concordant with the results gained by the FDA-approved fluorescence in situ hybridization (FISH) method. dc-CISH combines the good qualities of both FISH and CISH and represents a simple and reliable method for gene copy number assessment in conjunction with good morphological examination and brightfield microscopy.

INTRODUCTION

Breast cancer has a major impact on the health of women worldwide. In Finland, 3903 new breast cancer cases were diagnosed during the year 2004, 31% of all diagnosed cancers in females that year (www.cancerregistry.fi). The age-adjusted incidence of breast cancer has increased in Finland during the past 50 years, and currently every ninth woman will be affected by breast cancer during their lifetime (Pukkala et al. 2003). Mortality has remained almost statistically unchanged over time, and has actually started to decrease slightly since the late 1990s (www.cancerregistry.fi). Of Northern European countries, Finland, along with Sweden and Iceland, has the lowest ratio of mortality versus incidence (around 0.3, www.cancerregistry.fi).

Epidemiological studies have found many risk factors for breast cancer and aside from age and family history, most relate to longer exposure to endogeneous steroid hormones (Brewster and Bondy 2005). Early menarche, late menopause, and most likely obesity contribute to the cumulative life-time hormone exposure (Brewster and Bondy 2005). On the other side, young age at first birth, a high number of pregnancies, and lactation lower the life-time breast cancer probability (Brewster and Bondy 2005). It is thought that the protective effect of pregnancy is caused by terminal differentiation of the breast epithelium lowering the number of undifferentiated epithelial cells, which may be more vulnerable to carcinogenesis than the differentiated ductal and lobular structures (Brewster and Bondy 2005). It is also accepted that physical activity and intake of fruits and vegetables lower breast cancer risk (Brewster and Bondy 2005). The most compelling question in breast cancer epidemiology lately has been the role of hormone replacement therapy for postmenopausal women and birth control pills in breast cancer risk. It is now accepted that hormone replacement therapy causes an elevation of risk, especially when administered over five years and as a combined therapy of estrogen and progesterone (Brewster and Bondy 2005). The impact of birth control pills is still controversial, but it seems that there is no association with breast cancer risk (Brewster and Bondy 2005). In addition to above-mentioned risk factors, some hereditary

predispositions elevate breast cancer risk. It is estimated that almost one third of breast cancers might have some hereditary predisposition, but familial breast cancer with high-penetrance genetic changes account only for 5% to 10% of all breast tumors. Germ-line mutations in BRCA1 and BRCA2 genes are associated with families with breast and ovarian cancer and account for a portion of the hereditary breast tumors (Brewster and Bondy 2005, Kiechle and Meindl 2005). A woman with a BRCA1 or BRCA2 mutation has an 80-90% risk of acquiring breast cancer by the age of 80 (Kiechle and Meindl 2005).

Breast cancers are categorized according to histological types and are clinically characterized according to tumor size, axillary lymph node and distant metastasis (TNM) status, as well as the histologic grade of differentiation at diagnosis. Prognosis and treatment decisions are linked to TNM staging, grade, and histological type, but also very closely to the hormone receptor and HER-2 oncogene amplification status of the tumor (Fitzgibbons et al. 2000, Ross and Harbeck 2005). Hormone receptors and HER-2 status are determined for all primary breast cancers. Tumors positive for hormone receptors will be treated with hormonal therapy and breast cancers positive for HER-2 will be given trastuzumab (Herceptin^{TM}) antibody therapy against the HER-2 receptor (Fitzgibbons et al. 2000, Ross and Harbeck 2005). With regards to targeted treatment for patients with hormone receptor negative and HER-2 non-amplified tumors, chemotherapy remains the only effective option after surgery and radiation therapy (Ross and Harbeck 2005).

This study aims to describe the clinicopathological, biological, oncogenic, prognostic, and predictive characteristics of basal phenotype breast cancer and was undertaken in order to clarify the classification, origin, and aggressive nature of this hormone receptor negative breast cancer subgroup. One additional goal of this research was to move the chromogenic in situ hybridization method used in HER-2 oncogene amplification assessment to the next level by enabling the detection of two simultaneously hybridized DNA probes from a single sample.

REVIEW OF THE LITERATURE

1. Human mammary gland epithelium

The mammary gland is a continuously developing organ from the embryonic stage through puberty, menstrual cycles, and pregnancies until atrophy after menopause. The human mammary gland forms during embryonic development to contain the nipple, aerola, and the underlying ductal tree with 15-25 major ducts arranged in a segmental or radial manner constructed from secretory glandular cell and myoepithelial cell layers (Elston and Ellis 1998, Parmar and Cunha 2004). The mammary gland develops during puberty when the estrogen production by ovaries induces growth and maturation (Osborne 1996, Elston and Ellis 1998). During this time, the breasts enlarge due to the deposition of connective and fat tissue around the developing parenchyme (Osborne 1996). The ducts elongate and form a branching ductal tree extending to individual lobes with branching ductules leading to separate lobules (Osborne 1996, Parmar and Cunha 2004). After development in puberty, the mammary gland structure fluctuates with the menstrual cycle according to hormonal changes and develops into a fully differentiated lactating organ in pregnancy, again due to hormonal regulation (Dickson 1996, Osborne 1996).

Both the ducts and lobes are lined throughout by the two epithelial layers, the luminal glandular and basal myoepithelial layers (Figure 1), which are separated from the stroma by a basement membrane (Jolicoeur 2005). A third cell population present in the mammary gland epithelium is formed by the stem/progenitor cells of the luminal and basal cell layers (see paragraph 1.2). The use of the term "basal" is complex, it can mean localization above the basement membrane and in this way can be used as a synonym to the mammary gland myoepithelium. Alternatively, it can refer to the basal cytokeratin positive cells in the mammary gland described in detail in paragraph 1.1 (Gusterson et al. 2005). In the functional terminal ductal-lobular unit (TDLU), the secretory or glandular epithelium facing the lumen and thus called the "luminal

epithelium" will develop during pregnancy to produce milk components including lipids, proteins, and lactose (Osborne 1996, Parmar and Cunha 2004). The luminal cells contain specialized apical and basolateral cell membranes expressing sialomucin and adhesion molecules, respectively (Rønnov-Jessen et al. 1996). Luminal cells are columnar in the ducts but tend to be more cuboidal in the secretory TDLU (Elston and Ellis 1998). The expression of epithelial specific antigen (ESA) and epithelial membrane antigen (EMA) has long been associated with luminal cells (Nagle et al. 1986), but the most widely used method for luminal cell recognition is their unique cytokeratin expression pattern (See paragraph 1.1).

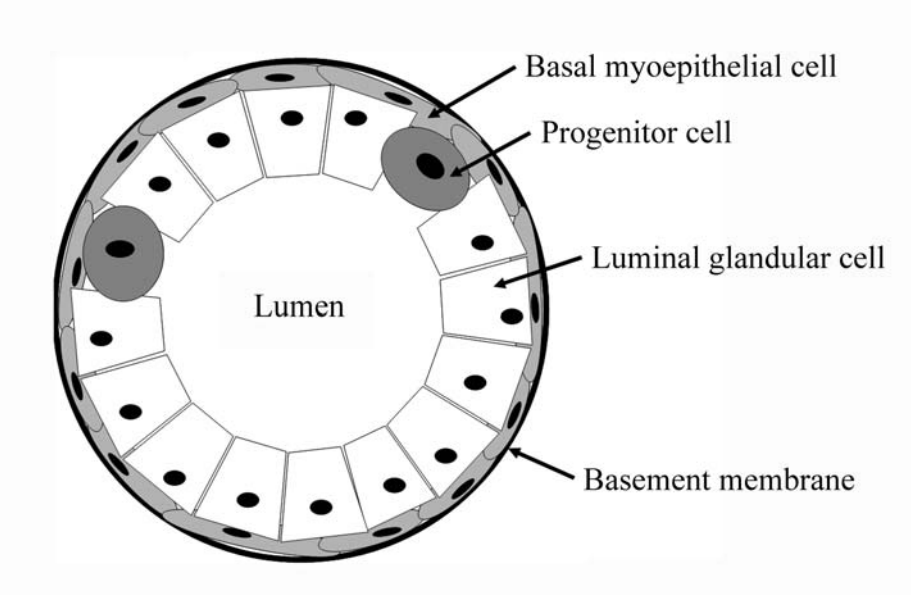

Figure 1. The structure of normal human mammary gland epithelium with the two well characterized cell layers, luminal glandular and basal myoepithelial, and a third cell population comprising the mammary gland stem/progenitor cells (modified from Birnbaum et al. 2004)

The myoepithelial or basal cells are situated between the luminal cell layer and the basement membrane and they function in structural maintenance, extracellular matrix remodeling, milk ejection induced by oxytocin, and control of luminal cell physiology (Osborne 1996, Lakhani and O'Hare 2001, Deugnier et al. 2002, Gudjonsson et al. 2002, Jolicoeur 2005). Myoepithelial cells are flattened or spindle-shaped, especially in the TDLU, and they form a continuous cell layer above the basement membrane (Osborne 1996, Deugnier et al. 2002, Gusterson et al. 2005, Polyak and Hu 2005). They form a connection to the underlying basement membrane through myofibrils

and hemidesmosomes (Osborne 1996, Deugnier et al. 2002) and are the only resource of laminin-1, a basement membrane protein in the mammary gland (Gudjonsson et al. 2002). This trait is essential for basement membrane deposition, epithelial cell polarity and double layer formation in ducts and lobules (Gudjonsson et al. 2002). Myoepithelial cells show a low proliferative activity in the adult mammary gland while luminal cells divide and increase in number in pregnancy and during mentrual cycles (Joshi et al. 1986, Suzuki et al. 2000, Bánkfalvi et al. 2004). Myoepithelial cells are most often defined by their smooth muscle actin expression (Taylor-Papadimitriou et al. 1989, Rønnov-Jessen et al. 1996). In addition, vimentin, glial fibrillary acid protein, CD10, p63, smooth muscle myosin, S100, calponin, and caldesmon can also be used for myoepithelial cell identification (Gould et al. 1990, Yaziji et al. 2000). The cytokeratin expression pattern of myoepithelial cells differs from that of the luminal secretory cells, but this pattern is not unique only to the myoepithelial cells in the mammary gland (Böcker et al. 2002).

1.1 Cytokeratin expression in normal human mammary gland

Cytokeratins belong to 7-11 nm intermediate filaments together with vimentin, desmin, neurofilaments, and glial fibrillary acidic protein (Moll et al. 1982). Cytokeratins are expressed in epithelial cells as a large and complex family of proteins (Moll et al. 1982). The cytokeratin genes either in chromosomes 12 and 17 are highly conserved as are the protein chain structures which have a central α-helical rod-like domain, an N-terminal head, and a carboxyl terminal tail (Chu and Weiss 2002). Cytokeratins are divided into acidic (type I) and neutral to basic (type II) groups based on their characteristic isoelectric points and into high or low molecular weight cytokeratins (Moll et al. 1982). Cytokeratins 1-8 belong to type II and cytokeratins 9-20 to type I (Moll et al. 1982). The cytokeratin polymers prefer pairing with partners from the opposite cytokeratin type, forming parallel heteropolymers of types I and II (Moll et al. 1982, Chu and Weiss 2002). Epithelial cells hence express at least two different types of cytokeratin polymers. The heteropolymers align

antiparallel to each other to form tetramers and continue to form the filament by end to end assembly, further alignment, and winding (Chu and Weiss 2002). The main function of cytokeratins is to provide mechanical support for the epithelial cells, but they may also have other functional properties. It is known that posttranslational modifications occur and most likely have regulatory effects on cytokeratin functions (Omary et al. 1998).

Luminal glandular and basal myoepithelial cells of the normal mammary gland can be distinguished in tissue sections on the basis of their cytokeratin (CK) expression patterns. To understand the basis for nomenclature, it should be noted that the name "basal cytokeratin" was originally given to the CK5/14/17 expressed in basal cells of the stratified epithelium of the skin and is now widely used when referring to these cytokeratins (Gusterson et al. 2005). In a normal mammary gland epithelium, the cytokeratin pattern is broad including CK5, 7, 8, 13, 14, 15, 17, 18, and 19 (Moll et al. 1982, Taylor-Papadimitriou et al. 1989, Rønnov-Jessen et al. 1996, Chu and Weiss 2002). Cytokeratins 7, 8, 18, and 19 are expressed in the luminal epithelium while basally located myoepithelial cells have a different cytokeratin pattern negative for the luminal but positive for the basal cytokeratins 5, 14, and 17 (Moll et al. 1982, Dairkee et al. 1986, Nagle et al. 1986, Taylor-Papadimitriou et al. 1989, Rønnov-Jessen et al. 1996, Moll 1998, Böcker et al. 2002). The cytokeratin pattern can vary between structural units of the normal mammary gland (Table 1) and the distribution of a particular cytokeratin is not always restricted only to the luminal glandular or to basal myoepithelial cells (Taylor-Papadimitirou et al. 1989, Rønnov-Jessen et al. 1996). Luminal CK8 and CK19 are not expressed in a portion of the luminal cells in the TDLU, leading to a heterogeneous staining pattern (Taylor-Papadimitriou et al. 1989, Rønnov-Jessen et al. 1996). Also the basal myoepithelial cells of the TDLU, detected by smooth muscle actin positivity, might be at least partly negative for CK5/14/17 (Böcker et al. 2002). Luminal CK7/19 can also be weakly expressed in the myoepithelial cells, especially in larger ducts while basal CK5/14/17 can be occasionally found in luminally located cells in the human mammary gland (Nagle et al. 1986,

Taylor-Papadimitriou et al. 1989, Moll 1998, Böcker et al. 2002). In fact, it is now widely accepted that the mammary gland epithelium holds a third cell population in addition to the luminal and basal cell layers. These cells are the tissue specific stem/progenitor cells (Figure 1), which are capable of producing the more differentiated glandular and myoepithelial cells (Kordon and Smith 1998, Stingl et al. 2006). One candidate for the mammary gland progenitor cell is the occasional more luminally located CK5-positive cell (Böcker et al. 2002).

Table 1. Major cytokeratins in normal human mammary gland epithelium. Negative (neg), weakly positive (weak), heterogeneously positive (heter), and strongly/uniformly positive (strong) expression is indicated separately in the luminal and myoepithelial layers in the TDLU and ducts.

Cytokeratins	**TDLU**		**DUCTS**	
	Luminal epithelium	***Myoepithelium***	***Luminal epithelium***	***Myoepithelium***
CK7	strong	weak	strong	weak
CK8	heter	neg	heter	neg
CK18	strong	neg	strong	neg
CK19	heter	neg	strong	weak
CK5	heter	heter	heter	strong
CK14	neg	weak	heter	strong
CK17	neg	neg	heter	strong

1.2 Mammary gland stem/progenitor cells

A tissue specific stem cell is a quiescent long-lived cell with the ability to self-renew and produce progenitor cells (Smalley and Ashworth 2003). Progenitor cells, also called transit amplifying cells, are multipotent undifferentiated cells, which are able to proliferate and produce the functional cell types of a particular tissue (Smalley and Ashworth 2003, Kalirai and Clarke 2006). The differentiation towards distinct cell lineages happens in a continuum of stages, and thus there are progenitor cells in different stages towards terminal differentiation with intermediate phenotypes (Behbod and Rosen 2004, Kalirai and Clarke 2006). The presence of a third epithelial cell

population in the mammary gland, stem/progenitor cells, is now widely accepted since it has been shown that the functional gland can be regenerated from a single cell (Kordon and Smith 1998, Stingl et al. 2006). Still, there are contradictory results about the method of identification, especially about the proteins, which would best specify the stem/progenitor cells in mammary epithelium (Kalirai and Clarke 2006).

Stem/progenitor cells have been studied by their ability to pump chemicals out of the cytoplasm due to the action of multidrug resistance proteins (Kalirai and Clarke 2006). This method identifies the so-called side population enriched in stem/progenitor cells (Alvi et al. 2003, Clarke RB et al. 2004). These cells are double negative for the mammary gland epithelium lineage specific markers EMA and CD10 (Clayton et al. 2004). They are also capable of differentiation to both luminal and myoepithelial cell types and regeneration of the mouse mammary gland upon transplantation (Alvi et al. 2003, Clarke RB et al. 2004). By labeling the mammary gland epithelial cells with DNA precursors, which will be retained in the stem cells as they persist throughout life and rarely divide, a similar cell population has been detected (Welm et al. 2002, Clarke RB et al. 2004). Another approach has been the identification of mammary gland epithelial cells with certain stem cell properties using flow cytometry or immunostaining (Kalirai and Clarke 2006). There are many markers such as stem cell antigen-1 (Welm et al. 2002, Deugnier et al. 2006), CK5 (Böcker et al. 2002), CK6 (Welm et al. 2003), CK19, Musashi-1, p21 (Clarke RB et al. 2004), EMA, CD10, and ESA (Stingl et al. 1998), which can be used solely or in combination to identify cell populations capable of generating both the epithelial cell phenotype and possibly the functional mammary gland. The proportion of stem/progenitor cells varies greatly between the different detection methods (Welm et al. 2002, Alvi et al. 2003) and it is most likely that the detected cell populations are overlapping, containing stem/progenitor cells but also more differentiated cells in varying amounts (Welm et al. 2002). Thus, there is no definite consensus about the exact phenotype of the biologically functional stem/progenitor cell population in the human mammary gland. Even

the expression of the estrogen receptor (ER) in the stem/progenitor cells is still controversial (Kalirai and Clarke 2006). One model suggests that ER-negative progenitors would be more primitive and produce ER-positive progenitors capable of differentiation towards ER-positive and ER-negative glandular cells (Dontu et al. 2004). The ER-negative stage of stem/progenitor cells is supported by the ER-negativity of the side population cells derived from the mammary gland (Clayton et al. 2004).

One of the mammary gland progenitor cell models located CK5-positive cells in the suprabasal position of the mammary gland epithelium by using a double immunofluorescence technique. This model also showed that these cells can differentiate (Figure 2) into smooth muscle actin and CK5 expressing precursors of myoepithelial cells and into CK8/18 and CK5 expressing precursors of luminal glandular cells (Böcker et al. 2002, Boecker et al. 2002, Boecker and Buerger 2003). When these cells mature to fully developed myoepithelium and glandular secretory cells, they lose CK5 expression and show only a smooth muscle actin or CK8/18-positive immunophenotype, respectively (Böcker et al. 2002). Böcker and coworkers (2002) thus concluded that CK5-positive cells display stem/progenitor cell properties in the mammary gland. This theory is supported by the results from mammosphere studies (Dontu et al. 2003). Mammospheres are cell cultures supporting the undifferentiated state of the epithelial cells and are derived from a single cell grown in conditions not allowing for adherence to the substratum (Dontu et al. 2003). The mammospheres are enriched in early progenitor cells and are able to form mammary gland epithelial cell lines and functional structures (Dontu et al. 2003). Most importantly, these cells expressed CK5, CD10, and α6 integrin, suggesting that these markers would be specific for early progenitor cells of the mammary epithelium (Dontu et al. 2003). The results of Deugnier et al. (2002b) also support CK5 expression in breast progenitor cells as the cell line BC44, which is positive for CK5 can form bilayered alveolus-like structures after injection to a cleared mammary fat pad.

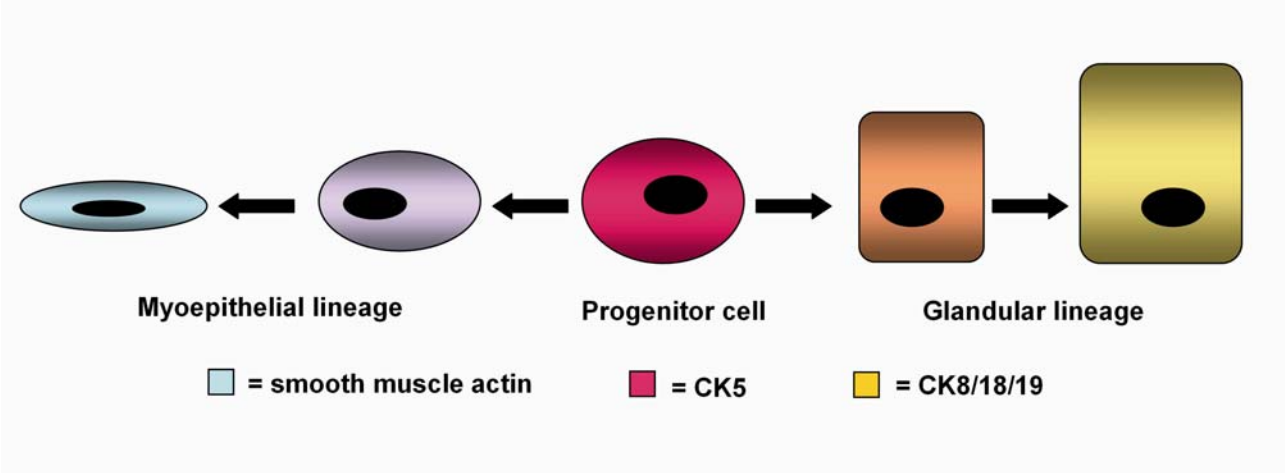

Figure 2. A suggested model of CK5-positive breast stem/progenitor cells capable of gradual differentiation towards both the luminal glandular lineage positive for CK8/18 and the basal myoepithelial lineage positive for smooth muscle actin (modified from Boecker et al. 2002).

2. Breast cancer stem cells

Cancer cells and stem/progenitor cells possess many similar characteristics. Both have the abilities of self-renewal, limitless proliferation potential, and tissue specific differentiation (Reya et al. 2001, Smalley and Ashworth 2003, Ponti et al. 2006). Moreover, regulation of these features is similar in cancer and stem cells including telomerase activity, inhibition of apoptosis (Ponti et al. 2006), and molecular pathways important in self-renewal like Wnt (Li et al. 2003, Behbod and Rosen 2004, Brennan and Brown 2004), Notch (Dontu et al. 2004), and Hedgehog (Liu et al. 2006). In fact, these molecular pathways often show dysfunction in cancers (Liu et al. 2005). Stem cells are long-lived and are thus subject over long periods of time to genetic mutations and environmental stress leading more easily to the accumulation of errors in the genome than in differentiated cells with a limited life span (Al-Hajj et al. 2004, Ponti et al. 2006). Also, as the capacity for self-renewal is already available, it is possible that fewer alterations are needed for the stem/progenitor cell to transform to a cancer cell than would be needed for a more differentiated cell (Burkert et al. 2006). Since it seems easier to retain the self-renewal property than to regain it and reactivate immortality after differentiation, it has been suggested that tumors including breast cancers preferentially arise from

transformed tissue-specific stem/progenitor cells called cancer stem cells (Reya et al. 2001, Smalley and Ashworth 2003, Ponti et al. 2006).

This model is supported by the results of Al-Hajj et al. (2003), who showed that only a minority of breast tumor cells have the capacity for self-renewal and differentiation. These cells could form the small cancer stem cell population producing the more differentiated limited proliferative life span cancer cells, which form the mass of the tumor. The cancer stem cells found by Al-Hajj et al. (2003) expressed ESA and CD44 and were negative for CD24, and when transplanted, they formed a heterogeneous tumor phenotypically similar to the original tumor. The existence of CD44+ CD24- breast cancer stem/progenitor cells has been confirmed by others (Ponti et al. 2005). It was recently reported that most early disseminated tumor cells in the bone marrow are enriched in this putative CD44+CD24- breast cancer stem cell phenotype (Balic et al. 2006) supporting their central role in metastatic dissemination. Additionally, breast cancer cell lines, which possess a high proportion of cells with the cancer stem cell phenotype CD44+CD24- show basal, myoepithelial, or mesenchymal characteristics according to gene expression profiling and marker expression (Sheridan et al. 2006). This research group concluded that these cells are highly invasive and express frequently CK5, CD10, and vimentin (Sheridan et al. 2006).

Given the fact that cancer stem cells exist, they are ideal targets for therapy. Cancer recurrence could be prevented by eradicating the cancer stem cells, since the re-seeding of the tumor would become impossible or greatly limited (Reya et al. 2001, Smalley and Ashworth 2003, Al-Hajj et al. 2004). It has been suggested that chemotherapies in use would preferentially eradicate the more differentiated tumor cells with limited proliferation potential as the cancer stem cells might have resistance to chemotherapeutics and divide rarely similarly as the stem/progenitor cells do (Reya et al. 2001, Smalley and Ashworth 2003, Al-Hajj et al. 2004). This would lead to recurrence since the cancer stem cells would be retained in the tissues (Reya et al. 2001, Smalley and Ashworth 2003, Al-Hajj et al. 2004).

3. Basal phenotype breast cancer

Basal phenotype or basal-like breast cancer has been under intensive research throughout the time of this study. In light of the literature published by fall 2006, this constitutes a unique and aggressive breast cancer subtype (Sørlie et al. 2001, Abd El-Rehim et al. 2004) suggested to originate from epithelial stem/progenitor cells of the breast (Boecker et al. 2002, Böcker et al. 2002, Boecker and Buerger 2003). For these reasons, the basal phenotype has gained a great deal of interest in the field of breast cancer investigation.

Immunohistochemical expression of basal cytokeratins was shown in the breast as early as 1982 (Moll et al. 1982). Still, the knowledge about these CK5/14/17-positive tumors has accumulated only recently from immunohistochemical and gene expression microarray studies. Since these cytokeratins are normally expressed in the basal cell layer of the breast epithelium, these breast carcinomas have been given the suffix "basal". This terminology has not yet been well established, since terms basal-like, basal phenotype, basal epithelial phenotype, basal cell phenotype, and basal-type are used in the literature. The term basal-like is most often used when referring to the cDNA microarray based classification and the other terms are in use in immunohistochemical studies. In this work, the terms basal phenotype and basal-like will be used to refer to this specific breast cancer subtype.

3.1 Identification of basal phenotype breast cancer with immunohistochemistry

It is generally accepted that the expression of cytokeratins remains stable during carcinogenesis, thereby allowing for indirect conclusions to be made regarding the cell of origin (Moll et al. 1982). In general, breast cancers show strong and wide expression of CK8/18/19, which is the main reason for the assumption that breast cancers arise from differentiated luminal epithelial cells expressing these cytokeratins. By showing CK14/17 expression in breast cancer, Moll et al. (1982) formed the basis for the classification of breast cancers using intermediate filament immunohistochemistry.

Years later, antibodies specifically detecting CK5 were added to the equation (Nagle et al. 1986, Gould et al. 1990). It should be noted, however, that the antibody commonly used in these studies (clone D5/16 B4) also detects CK6, so many studies cannot separate the impact of these cytokeratins. Also, CK4 has been used in some researches but its independent role in basal phenotype breast cancer is controversial (Nagle et al. 1986, Malzahn et al. 1998). CK5/14/17 expression is currently the most widely used immunohistochemical method for the distinction of basal phenotype breast cancer from luminal CK8/18/19 expressing tumors (Dairkee et al. 1987a, Wetzels et al. 1991, Malzahn et al. 1998, Korsching et al. 2002, van de Rijn et al. 2002, Reis-Filho et al. 2003, Abd El-Rehim et al. 2004, Birnbaum et al. 2004, Clarke et al. 2005, Gusterson et al. 2005, Potemski et al. 2005, Banerjee et al. 2006, Rakha et al. 2006). In general, breast cancers have been thought to express either luminal or basal cytokeratins (Wetzels et al. 1991, Birnbaum et al. 2004), but evidence of co-expression has been growing (Malzahn et al. 1998, Böcker et al. 2002, Boecker and Buerger 2003, Abd El-Rehim et al. 2004, Kim et al. 2006). The separation of basal and non-basal tumors should thus be conducted without luminal cytokeratin involvement, solely classifying tumors by basal cytokeratin expression or absence, respectively.

As basal cytokeratins are normally expressed in the basal myoepithelial cell layer in the breast, there have been attempts to find new basal phenotype breast cancer markers among myoepithelium specific proteins like the nuclear transcription factor and the p53 homolog p63 (Barbareschi et al. 2001, Nylander et al. 2002, Wang et al. 2002, Ribeiro-Silva et al. 2003). The prevalence of p63-positive breast cancer is suggested to be as high as 10-12% (Reis-Filho et al. 2003, Ribeiro-Silva et al. 2003, Makretsov et al. 2004), although some investigators have not been able to detect p63 immunopositivity in invasive breast carcinomas at all (Barbareschi et al. 2001, Wang et al. 2002). Aside from p63, the well-characterized myoepithelium differentiation marker smooth muscle actin has been suggested as a possible basal/myoepithelial phenotype indicator in breast cancer. It is rarely expressed in breast carcinomas (Lakhani and O'Hare 2001), although some

researchers have reported results showing smooth muscle actin immunopositivity in breast cancers expressing basal cytokeratins (Hungermann et al. 2005, Jacquemier et al. 2005, Livasy et al. 2006, Rakha et al. 2006). Further study is necessary to determine whether the smooth muscle actin expressing tumors are part of the basal phenotype or form a distinct breast cancer subtype.

Immunohistochemical studies are affected not only by the antibodies used but also by the variability in the interpretation of the staining results. Many studies have used "any cytoplasmic staining" as the minimum level of basal cytokeratin positive tumor cell staining needed for a tumor to be regarded as the basal phenotype (Korsching et al. 2002, van de Rijn et al. 2002, Abd El-Rehim et al. 2004, Nielsen et al. 2004, Potemski et al. 2005, Banerjee et al. 2006, Rodrígues-Pinilla et al. 2006a). Some researchers have used different percentage cut-offs ranging from 1% (Kim et al. 2006) to 6% (Malzahn et al. 1998) or 10% (Otterbach et al. 2000, Rakha et al. 2006). Histoscores for semiquantitative evaluation has also been used (Putti et al. 2005). When using immunohistochemistry as the detection method, this kind of variation in limits for positivity should be minimized in order to get comparable results between studies and to get a clear picture of the associations with other biological and clinicopathological parameters.

3.2 Identification of basal phenotype breast cancer with gene expression microarrays

cDNA microarrays have become a widely used tool to search prognostically and biologically meaningful tumor subtypes among breast cancers. The tumors are classified to distinct clusters by gene expression profiles, which differ from each other significantly or by finding a typical signature for tumor characteristics such as poor prognosis. When Perou et al. published in 2000 an intrinsic gene expression signature of breast tumors, one of the tumor subtypes expressed many of the genes characteristic of breast basal epithelial cells including cytokeratins 5 and 17, and the uniqueness and existence of basal phenotype breast cancer was confirmed. Gene expression microarray studies have since distinguished the basal-like tumors from the luminal (A, B, and C), ERBB2, and normal

breast-like tumor subtypes with different gene expression signatures and prognosis (Perou et al. 2000, Sørlie et al. 2001, Sotiriou et al. 2003, Sørlie et al. 2003, Rouzier et al. 2005, Calza et al. 2006, Fan et al. 2006, Hu et al. 2006, Sørlie et al. 2006). The most typical feature of the intrinsic basal-like breast cancer is the lack of expression of ER and genes usually co-expressed with ER (Perou et al. 2000, Sørlie et al. 2001, Sotiriou et al. 2003, Sørlie et al. 2003). cDNA microarray studies have shown that the most powerful indicator in determining the gene expression profiles and prognostic groups of breast cancer is ER and ER-related genes (Gruvberger et al. 2001, West et al. 2001, van't Veer et al. 2002, Pusztai et al. 2003, Gruvberger et al. 2002). As the basal-like intrinsic signature is so tightly related to ER, the role of this breast cancer subtype should be studied more specifically in relation to the appropriate ER-negative reference group.

3.3 Prevalence of basal phenotype tumors in sporadic breast cancers

Regardless of the method of detection, the basal phenotype seems to account for approximately one tenth of all sporadic breast cancers. The prevalence of CK14-positive tumors was among the first to be described and has been reported to vary from 4% to 18% (Dairkee et al. 1987a, Wetzels et al. 1991, Heatley et al. 1995, Abd El-Rehim et al. 2004). The frequency of solely CK5 or CK17 expressing tumors varies from 8% to 20% (Otterbach et al. 2000, Korching et al. 2002, Reis-Filho et al. 2003, Abd El-Rehim et al. 2004, Makretsov et al. 2004, Nielsen et al. 2004) and from 5% to 31%, respectively (Wetzels et al. 1991, Reis-Filho et al. 2003, Nielsen et al. 2004). Among tumors with CK5 and/or CK14 expression, the prevalence has been ~19% (Wetzels et al. 1991, Rakha et al. 2006), and for the combination of CK5/17 the occurrence is reported to be roughly 15% (van de Rijn et al. 2002, Nielsen et al. 2004). If all three cytokeratins are taken into account, positivity is seen in 17% of tumors according to Banerjee et al. (2006). Potemski et al. (2005) have reported a significantly higher occurrence of basal phenotype tumors in invasive ductal carcinomas from 25% to 37% using immunohistochemical methods.

By conducting classification with cDNA microarrays, the proportion of basal-like breast cancer is found to be between 9% and 18%, quite close to the immunohistochemically attained numbers (Perou et al. 2000, Sørlie et al. 2001, Sotiriou et al. 2003, Sørlie et al. 2003, Fan et al. 2006). When the immunohistochemical marker pattern (ER/HER-2-negative and CK5/6-positive and/or EGFR-positive) created to identify cDNA microarray classified basal-like tumors (Nielsen et al. 2004) is used, the prevalence of the basal phenotype varies from 11% to 15% (Carey et al. 2006, Kim et al. 2006, Rodrígues-Pinilla et al. 2006a, Rodrígues-Pinilla et al. 2006b). This phenotype was recently detected in high grade in situ carcinoma lesions of breast and was suggested to be a precursor lesion for the invasive basal phenotype breast cancer (Bryan et al. 2006). Further, by using the same IHC panel (ER, HER-2, CK5/6, and EGFR) for basal phenotype identification, Carey et al. (2006) showed a higher prevalence of the basal phenotype in premenopausal African American women (36%) than in premenopausal non-African American women (16%). No other ethnic associations have been reported.

3.4 Association of the basal phenotype with BRCA1 in hereditary and sporadic tumors

It was noted years ago in the microarray studies by Sørlie et al. (2003) that hereditary breast cancer possesses similarity with sporadic basal-like breast cancers. Hereditary breast cancers account for 5-10% of all tumors in the breast, and are caused by protein truncating germ-line mutations in the BRCA1 or BRCA2 gene in about 25-30% of the cases with familial predisposition (Honrado et al. 2006). The rest of the familial breast cancers are caused by other high-penetrance susceptibility genes not well identified at this time. BRCA1 and BRCA2 gene products function in DNA repair, cell cycle control, differentiation, and regulation of gene expression and they are tumor suppressors, which respond to cellular damage and maintain genomic stability (Honrado et al. 2006).

At the time when this study was initiated, no data had been reported on basal cytokeratin expression in BRCA2 germ-line mutated tumors. In contrast, the association between the BRCA1 mutation and the basal phenotype has been established (Foulkes et al. 2003). In the gene expression microarray based classification of BRCA1 tumors, they have been shown to belong to the basal-like tumor cluster (Sørlie et al. 2003). The prevalence of the basal phenotype in BRCA1 germ-line mutated tumors has since been ascertained to be high, but the variation of occurrence is still quite extensive, ranging from 53% to more than 80% (Foulkes et al. 2003, Sørlie et al. 2003, Foulkes et al. 2004, Lakhani et al. 2005, Palacios et al. 2005, Rodrígues-Pinilla et al. 2006a, Rodrígues-Pinilla et al. 2006b).

As hereditary BRCA1 associated and sporadic tumors with the basal phenotype show such similar features and phenotype (Turner and Reis-Filho 2006), it has been suggested that BRCA1 protein dysfunction could also be involved in the evolution of sporadic basal phenotype tumors. It is not certain if epigenetic changes like methylation of the BRCA1 gene or the promoter would be causing the downregulation of the gene in sporadic basal phenotype tumors (Estellar et al. 2000, Rice et al. 2000, Matros et al. 2005). On the contrary, Matros et al. (2005) suggested that sporadic basal phenotype tumors would express high levels of the BRCA1 protein, a feature typical for tumors with a high mitotic rate. Even though, an inverse association between CK5 and BRCA1 protein expressions seems likely according to larger studies (Abd El-Rehim et al. 2004, Ribeiro-Silva et al. 2005). Additionally, it was recently reported that mRNA levels of BRCA1 are low in sporadic basal phenotype tumors (Turner et al. 2006). This was most likely caused by ID4, a negative regulator of BRCA1, which was overexpressed in basal phenotype tumors (Turner et al. 2006).

3.5 Morphological features and histological type of basal phenotype breast cancer

The most prominent morphological feature of basal phenotype tumors described in almost every study dealing with this breast cancer subgroup is high histological grade (Malzahn et al. 1998, Otterbach et al. 2000, Sotiriou et al. 2003, Abd El-Rehim et al. 2004, Foulkes et al. 2004, Potemski et al. 2005, Calza et al. 2006, Hu et al. 2006, Kim et al. 2006, Livasy et al. 2006, Rakha et al. 2006, Rodrígues-Pinilla et al. 2006a). High histological grade means poor differentiation of the tumor cells, nuclear pleomorphism, and high mitotic count, and it correlates with poor prognosis in breast cancer (Ross and Harbeck 2005). In addition to high histological grade, it has been shown that basal phenotype tumors show certain morphological features like a pushing border, lymphocyte infiltration, tumor necrosis, central scarring, and the presence of spindle cells more often than other breast cancers (Tsuda et al. 1999, Putti et al. 2005, Fulford et al. 2006, Livasy et al. 2006). Some of these features have also been described for BRCA1 mutated tumors, further illustrating the association between BRCA1 tumors and the basal phenotype (Honrado et al. 2006). Many of these morphological features also depict the medullary histotype of breast cancer, and it has been shown that half of medullary and atypical medullary tumors show basal phenotype characteristics like CK5/14 expression (Abd El-Rehim et al. 2004, Jacquemier et al. 2005, Carey et al. 2006, Fulford et al. 2006, Kim et al. 2006, Reis-Filho et al. 2006a). Otherwise, a majority of basal phenotype tumors are ductal with a not otherwise specified histotype, while lobular carcinomas very rarely show basal phenotype characteristics (Nagle et al. 1986, Abd El-Rehim et al. 2004, Carey et al. 2006, Kim et al. 2006, Livasy et al. 2006, Rakha et al. 2006). In addition to high frequency in medullary carcinomas most metaplastic carcinomas express basal cytokeratins (Livasy et al. 2006).

3.6 Clinicopathological and biological characteristics of basal phenotype breast cancer

Poorly differentiated basal phenotype breast carcinomas almost always show hormone receptor negative status (Malzahn et al. 1998, Perou et al. 2000, Sørlie et al. 2001, Korsching et al. 2002, Sotiriou et al. 2003, Abd El-Rehim et al. 2004, Foulkes et al. 2004, Nielsen et al. 2004, Potemski et al. 2005, Banerjee et al. 2006, Hu et al. 2006, Livasy et al. 2006, Rakha et al. 2006). Hormone receptor negativity, a known prognostic factor in breast cancer (Ross and Harbeck 2005), is the most essential feature of basal phenotype tumors. In addition to these features of aggressive behavior, basal phenotype tumors are highly proliferative (Korsching et al. 2002, Abd El-Rehim et al. 2004, Carey et al. 2006, Rakha et al. 2006, Rodrígues-Pinilla et al. 2006a) and in some studies are also associated with larger tumor size (Abd El-Rehim et al. 2004, Foulkes et al. 2004, Hu et al. 2006, Kim et al. 2006, Rakha et al. 2006, Rodrígues-Pinilla et al. 2006a) and younger age at diagnosis (Abd El-Rehim et al. 2004, Foulkes et al. 2004, Calza et al. 2006, Carey et al. 2006). Besides these factors, most of which would predict poor prognosis, there is limited knowledge about the association of the basal phenotype with axillary lymph node status. There are reports indicating that basal phenotype tumors would more often be lymph node negative (Abd El-Rehim et al. 2004, Rakha et al. 2006) while others have not found any association with lymph node status (van de Rijn et al. 2002, Foulkes et al. 2003, Calza et al. 2006, Carey et al. 2006), and some have reported a positive association (Ribeiro-Silva et al. 2006). These results might at least indicate that there is no clear tendency towards quick spreading to lymph nodes. Like in all studies associating biomarkers with lymph node status, fortified diagnostic procedures like mammography screening may be a confounding factor. Despite of the unknown metastatic propensity to lymph nodes, basal phenotype tumors are suggested to metastasize rarely to bone but have a tendency to local recurrence, visceral, and brain metastasis (Tsuda et al. 2000, Albiges et al. 2005, Hicks et al. 2006, Rodrígues-Pinilla et al. 2006a). The propensity to visceral metastasis is supported by Minn et al.

(2005) who reported that the basal-like expression profile overlaps with an expression profile specific for lung metastasis in breast cancer. Overall, CK14-positive breast cancer cell lines have been shown to possess a high invasive potential (Gordon et al. 2003).

Once the basal phenotype tumor subtype became widely acknowledged, a number of studies have addressed correlations between the basal phenotype and other biomarkers. It has been adequately proven that basal cytokeratin expression is strongly associated with positive EGFR, vimentin, and c-kit (CD117) status (Korsching et al. 2002, Nielsen et al. 2004, Korsching et al. 2005, Tsuda et al. 2005a, Kim et al. 2006, Livasy et al. 2006, Rakha et al. 2006). On the contrary, Bcl-2 expression is more likely to be low or absent in tumors expressing basal cytokeratins (Korsching et al. 2002). Overexpression of p53 is a frequent finding in basal phenotype breast cancer (Korsching et al. 2002, Foulkes et al. 2004, Kim et al. 2006, Rakha et al. 2006) caused by the p53 mutation (Calza et al. 2006, Carey et al. 2006). Frequent positivity of cyclin E has been detected in basal phenotype tumors (Foulkes et al. 2004, Potemski et al. 2005), although contradictory results exist indicating an association with cyclin A but not with cyclin E or D1 expression (Korsching et al. 2002). Further, an inverse association between basal cytokeratin expression and cyclin D1 overexpression and gene amplification has been indicated (Reis-Filho et al. 2006b). Placental cadherin, which is often expressed in medullary breast cancer (Palacios et al. 1995), has been suggested to associate with the basal phenotype in sporadic and BRCA1 mutated tumors (Arnes et al. 2005**,** Matos et al. 2005, Rakha et al. 2006). Fascin expression, which is related to lung metastasis in breast cancer, has recently been immunohistochemically associated with the basal phenotype (Rodrígues-Pinilla et al. 2006a) as was vascular endothelial growth factor (VEGF) overexpression, which may function as an autocrine growth factor and promote proliferation (Ribeiro-Silva et al. 2006). Additionally, associations with nerve growth factor receptor expression (Reis-Filho et al. 2006c) and with 14-3-3 sigma (Simpson et al. 2004) have been established for basal phenotype tumors.

Microarray studies have provided data of hundreds of genes associated with the basal-like phenotype. One with a strong association is αB-Crystallin, which is highly expressed in basal phenotype breast cancer (Perou et al. 2000) and in some breast cancer cell lines (Charafe-Jauffret et al. 2005). Recent data indicates that αB-Crystallin is an oncoprotein associated with poor clinical outcome (Moyano et al. 2006). Other proteins recognized as myoepithelial markers and associated with basal phenotype breast cancer are osteonectin or SPARC (Jones et al. 2004a, Lakhani et al. 2005) and caveolin-1 (Charafe-Jauffret et al. 2005, Rodrígues-Pinilla et al. 2006b). Osteonectin also has prognostic relevance as a short survival predictor (Jones et al. 2004a). Other genes found differently expressed in basal-like tumors by gene expression microarray studies include laminin, fatty acid binding protein 7, caveolin-2, annexin A8, TRIM29, and Skp-2 (Perou et al. 2000, Sørlie et al. 2001, Signoretti et al. 2002, Sotiriou et al. 2003, Sørlie et al. 2003, Sørlie et al. 2006).

3.7 Association of basal phenotype breast cancer with HER-2 status

Basal-like and HER-2 oncogene overexpressing breast cancers are exclusively classified to separate clusters according to gene expression signatures (Perou et al. 2000, Sørlie et al. 2001, Sotiriou et al. 2003, Sørlie et al. 2003). The results of some immunohistochemical studies support this and suggest that on average, basal phenotype tumors may express less HER-2 protein (Korsching et al. 2002). When immunohistochemical methods were developed to detect microarray identified basal-like tumor group, the result was that a tumor has to be HER-2 negative to be classified as basal-like (Nielsen et al. 2004, Carey et al. 2006, Kim et al. 2006, Reis-Filho et al. 2006a, Rodrígues-Pinilla et al. 2006a). In a clear contrast, there is evidence that basal cytokeratin expression and HER-2 oncogene amplification can co-exist (van de Rijn et al. 2002, Birnbaum et al. 2004, Tanner et al. 2004, Kim et al. 2006). Whether these features are mutually exclusive or not has to be elucidated and related to prognostic and predictive studies.

3.8 Genetic alterations in basal phenotype tumors

In addition to clinicopathological characteristics, basal phenotype breast cancers have been studied for cytogenetic alterations, mainly by comparative genomic hybridization (CGH). Korsching et al. (2002) found a higher number of genetic alterations in basal cytokeratin expressing tumors than in non-expressing tumors. Contradictory results indicating a lower number of genetic alterations in basal phenotype tumors within the histological grade III tumor subgroup have been presented (Jones et al. 2001, Jones et al. 2004b). Jones et al. (2004b) further suggested that losses of 16p, 17q, 19q, and Xp would be typical of basal cytokeratin expressing tumors. Further, Wang et al. (2004) found the loss of heterozygosity in 5q and 4p typical of basal-like tumors, which were identified by gene expression microarray. Interestingly, evidence was recently presented suggesting that both the sporadic and hereditary BRCA1 mutated basal phenotype tumors would lack an inactive X chromosome (Richardson et al. 2006). This would lead to overexpression of a small subset of X chromosomal genes and was concluded to have a role in basal phenotype breast cancer pathogenesis.

3.9 Prognosis of basal phenotype breast cancer

The unfavorable prognosis of basal cytokeratin expressing breast cancer was first described in 1987 (Dairkee et al. 1987b). Thereafter, it has been shown in many immunohistochemical studies that basal cytokeratin expressing tumors associate with poor survival (van de Rijn et al. 2002, Abd El-Rehim et al. 2004, Foulkes et al. 2004, Makretsov et al. 2004, Nielsen et al. 2004, Potemski et al. 2005, Banerjee et al. 2006, Carey et al. 2006, Rakha et al. 2006, Rodrígues-Pinilla et al. 2006a). Still, there are contradictory results of the prognostic effect within lymph node negative and lymph node positive subgroups. Some studies show significantly worse prognosis for basal phenotype tumors within the node negative tumor subgroup but not within node positive tumors (van de Rijn et al. 2002, Potemski et al. 2005). Others on the contrary show poor prognosis for basal phenotype

breast cancer within node positive tumors but not within the node negative subgroup (Malzahn et al. 1998, Nielsen et al. 2004). Recently, survival has been studied by cDNA microarrays and classification based on the so called intrinsic gene expression signature generated by Perou et al. (2000). The microarray studies show adverse survival for basal-like breast cancer when compared to luminal and normal-like clusters, although the ERBB2 tumor cluster seems to have very similar prognosis to basal-like breast cancer (Sørlie et al. 2001, Sotiriou et al. 2003, Sørlie et al. 2003, Calza et al. 2006, Hu et al. 2006). Many microarray studies have searched for prognostic gene expression profiles and gene sets for breast cancer (van de Vijver et al. 2002, van't Veer et al. 2002, Ma et al. 2004, Paik et al. 2004, Chang et al. 2005). A recent work by Fan et al. (2006) studied if the basal-like tumors identified by the intrinsic gene signature (Sørlie et al. 2001) would also be classified as poor prognosis tumors by other gene expression-based prognostic models. The 70-gene profile (van de Vijver et al. 2002, van't Veer et al. 2002), wound response model (Chang et al. 2005), and recurrence score model (Paik et al. 2004) all confirmed the poor prognosis of the basal-like intrinsic signature (Fan et al. 2006) as described originally by Sørlie et al. (2001). There is one opposing study of the prognosis of basal-like breast cancer where in the Korean population no significant survival difference was found between basal-like and non-basal tumors (Kim et al. 2006).

While univariate significance of the basal phenotype on survival has been found, there is little knowledge about the effect in a multivariate setting. Two studies using multivariate statistics show independence of the basal cytokeratin immunopositivity as a prognostic factor from tumor size, lymph node status, grade, and vascular invasion (Abd El-Rehim et al. 2004, Rakha et al. 2006). In addition, it was recently shown that the basal-like intrinsic gene signature seems to retain its prognostic relevance in comparison with the "luminal A" signature in a multivariate setting including age, ER status, lymph node status, tumor grade, and tumor size (Hu et al. 2006). In addition to the intrinsic signature, tumor size and histological grade were found to be additional

independent prognostic factors in this study (Hu et al. 2006). On the contrary, by fall 2006 there have been two multivariate prognostic studies suggesting that CK5/17 immunopositivity is not an independent prognostic factor (van de Rijn et al. 2002, Potemski et al. 2005). Van de Rijn et al. (2002) showed that the prognostic effect of CK5/17 immunopositivity was not independent of tumor size, lymph node status, or tumor grade. Still, they could show that within the lymph node negative tumor subgroup, CK5/17 retained prognostic independence from tumor size, tumor grade, HER-2, ER, and GATA-3 status (van de Rijn et al. 2002). In the study of Potemski et al. (2005), CK5/17 status, tumor size, node status, tumor grade, ER status, HER-2 status, cyclin E, and Ki-67 expression were included in the multivariate analysis. The independent prognostic factors found were lymph node status, HER-2 status, and cyclin E expression. The poor prognosis of basal phenotype tumors was suggested to be determined by ER status and cyclin E expression (Potemski et al. 2005).

3.10 Treatment of basal phenotype breast cancer

In order to have applicability in clinical practice, breast tumor classification should predict the efficacy or lack of efficacy of post-surgical therapy. Treatment of basal phenotype breast cancer is problematic as these tumors are invariably hormone receptor negative, ruling out the use of anti-estrogens and aromatase inhibitors. In addition, antibody therapy towards the HER-2 receptor will not be applicable for the majority of basal phenotype breast cancer patients. These tumors are thus treated most often by chemotherapy alone. To the best of my knowledge, there are only a few studies reporting the responsiveness of basal phenotype breast cancer to chemotherapy, and all are retrospective in nature. In a study of the preoperative setting (Rouzier et al. 2006), treatment with paclitaxel followed by 5-fluorouracil, doxorubicin, and cyclophosphamide chemotherapy resulted in the highest rates of pathologic complete response in basal-like and ERBB2 tumors (both 45%). Luminal and normal-like tumor subtypes had only 6% and 0% rates of complete response,

respectively. In this study, the basal-like subtype was defined by gene expression microarrays. Furthermore, the molecular mechanism of the chemotherapy sensitivity of the basal-like and ERBB2 tumors seems to be different since authors were unable to find overlap between the genes associated with complete response in basal-like and ERBB2 tumor subtypes (Rouzier et al. 2006). In an adjuvant treatment setting, the basal phenotype was associated with poor prognosis only among non-treated patients (Rodrígues-Pinilla et al. 2006a). In the group of patients treated with cyclophosphamide-methotrexane-5-fluorouracil, there was no difference in disease-specific survival between basal and non-basal tumors (Rodrígues-Pinilla et al. 2006a). These results would suggest that basal phenotype tumors are more sensitive to chemotherapy than non-basal tumors. On the other hand, when Banerjee et al. (2006) studied the treatment response of basal phenotype tumors to anthracycline-based adjuvant chemotherapy (5-fluoruracil and cyclophosphamide and either epirubicin or adriamycin), they found that the survival of patients with basal phenotype tumors was significantly shorter after therapy than the survival of patients with non-basal age, node, and grade matched tumors. This is supported by the results of Sørlie et al. (2003) who used adjuvantly treated patients and still associated basal-like tumors with poor prognosis. In a study by Calza et al. (2006), the authors concluded that basal-like breast cancers respond to endocrine therapy. In this series, 45% of basal-like tumors were estrogen receptor positive, which is in contradiction to almost all published reports (Perou et al. 2000, Sørlie et al. 2001, Sotiriou et al. 2003, Sørlie et al. 2003, Nielsen et al. 2004, Hu et al. 2006, Kim et al. 2006, Rodrígues-Pinilla et al. 2006a). These varied results show that there is no clear consensus on the treatment response of basal phenotype tumors.

4. HER-2 oncogene in breast cancer

The oncogene HER-2, also called ERBB2 or neu, belongs to the epidermal growth factor receptor family with HER-1 (EGFR), HER-3, and HER-4. The HER-2 gene is located in the chromosomal region 17q12 and it codes for a transmembrane tyrosine kinase receptor functioning in signal transduction (Ross et al. 2004). Receptor activation is achieved after ligand binding by homo- or

heterodimerization of the receptors with the family members and by subsequent autophosphorylation of cytoplasmic tyrosine residues (Stern 2000, Ross et al. 2004). Other family members have numerous growth factor partners, but HER-2 is an orphan receptor without a known ligand (Rubin and Yarden 2001, Cardoso et al. 2005). HER-2 is the preferable dimerization partner among the receptor family, and it forms more stable dimers and leads to more prominent signaling than other family members (Tzahar et al. 1996, Yarden and Sliwkowski 2001, Cardoso et al. 2005). Signaling of epidermal growth factor receptor family members leads to various effects in cell proliferation, survival, motility, and adhesion and is mediated through pathways such as the mitogen activated protein kinase (MAPK), phosphatidylinositol 3-kinase (PI3K)/Akt, Janus kinase, and phospholipase Cγ (PLC-γ) pathways (Yarden and Sliwkowski 2001, Cardoso et al. 2005).

In breast cancer, HER-2 oncogene amplification can be found in 10-34% of tumors (Ross et al. 2004), and amplification almost invariably leads to protein overexpression and excess signaling activity. HER-2 status is stable in metastasis (Simon et al. 2001, Tanner et al. 2001, Gong et al. 2005a) and is associated with high histological grade, hormone receptor negative status, and a ductal histotype (Ross et al. 2004, Cardoso et al. 2005). In addition, HER-2 amplified tumors show an extensive ductal carcinoma in situ (DCIS) component, larger tumor size, high proliferation activity, p53 expression and mutation, topoisomerase IIα amplification and lymph node metastasis more often than non-amplified tumors (Isola et al. 1992, Bilous et al. 2003a, Ross et al. 2004, Ariga et al. 2005, Cardoso et al. 2005, Huang et al. 2005, Tanner et al. 2006). Many of these features point towards poor prognosis and since 1987, when it was first shown that HER-2 is a prognostic factor in breast cancer (Slamon et al. 1987), many studies have come to the same conclusion (Ross et al. 2004). Also, cDNA microarray studies have shown that HER-2 positive breast tumors form a tumor cluster with a unique gene expression pattern and poor prognosis (Perou et al. 2000). Adding to the prognostic relevance of the receptor, HER-2 is a predictive factor in breast cancer. Positive HER-2 status is associated with a favorable response to antracyclines, possibly caused by frequent co-

amplification with topoisomeraseIIα (Slamon et al. 2001, Yamauchi et al. 2001, Cardoso et al. 2005, Gonzalez-Angulo et al. 2006, Tanner et al. 2006). HER-2 also associates with a poor response to hormonal therapy (Yamauchi et al. 2001, Cardoso et al. 2005). Most importantly, these tumors respond to antibody therapy towards the HER-2 receptor (Cardoso et al. 2005, Gonzalez-Angulo et al. 2006).

Trastuzumab (Herceptin^{TM}), a monoclonal IgG class anti-HER-2-humanized murine antibody (Carter et al. 1992) developed by Genentech (South San Francisco, CA), was launched and FDA-approved in 1998 and has since been used in treatment of HER-2 positive breast cancer in metastatic settings (Jackisch 2006). The treatment leads to a significantly longer time to progression and longer survival times for the HER-2 amplified breast cancer patients, and is most effective when used in combination with chemotherapy (Cardoso et al. 2005, Jackisch 2006). The main limiting factor in the use of trastuzumab has been its cardiac toxicity, especially when used in combination with anthracyclines (Gonzalez-Angulo et al. 2006, Jackisch 2006). Recently, the use of the trastuzumab antibody has been broadened to adjuvant settings with promising results (Piccart-Gebhart et al. 2005, Romond et al. 2005, Gonzalez-Angulo et al. 2006, Joensuu et al. 2006), and the most essential question is now how long trastuzumab should be administered to achieve the best treatment response in the adjuvant setting.

4.1 Methods in the assessment of HER-2 status

As treatment decisions and the prognosis of breast cancer patients depend so significantly on HER-2 status among other prognostic and predictive factors like hormone receptor status, the assessment of HER-2 status is already a critical step at the time of primary breast cancer diagnosis. The assessment of HER-2 can be viewed from three angles; to detect the protein, the gene, or the mRNA. The protein expression is mostly visualized by IHC from tumor tissue slides, but enzyme-linked immunosorbent assay (ELISA) can also be used to detect protein from fresh samples (Ross et

al. 2003). The examination of protein overexpression by IHC has been widely used for HER-2 assessment, as it is easy to perform and is a routine technique in pathology laboratories (Cardoso et al. 2005). Recently, evidence has been growing to emphasize the lack of standardization in the IHC staining procedure and problems in the evaluation leading to variability in the final result (Gancberg et al. 2002, Ross et al. 2003, Sidoni et al. 2006). HER-2 oncogene amplification has been studied mainly by the fluorescence in situ hybridization (FISH) method, which is easier to standardize and control than IHC (Cardoso et al. 2005). FISH is often used in combination with IHC to confirm the equivocal IHC results of the HER-2 protein overexpression. More recently, chromogenic in situ hybridization (CISH) has become a new alternative to FISH for gene copy number assessment (Tanner et al. 2000, Isola et al. 2004). While gene expression microarray analyses are used successfully for wider examination of prognostic and predictive tumor groups in breast cancer, it is too expensive and laborious to conduct routine HER-2 assessment at the mRNA level (Ross et al. 2003). There have also been attempts to determine the HER-2 status from mRNA using PCR, but the limitation of this method is the impact of the normal breast cells on the final result (Ross et al. 2003).

4.1.1 Detection of HER-2 gene amplification by FISH

The fluorescence in situ hybridization (FISH) method is used to determine gene copy numbers by introducing a fluorescently-labeled DNA probe to a tissue section after appropriate pretreatments. The technique has remained almost unchanged since it was first described (Pinkel et al. 1986, Kallioniemi et al. 1992) and it has been adapted to formalin-fixed paraffin-embedded tissue samples (Hyytinen et al. 1994, Pauletti et al. 1996). When using FISH for HER-2 status assessment, a probe for the HER-2 gene and a control probe for the chromosome 17 centromere labeled with distinct fluorophores are cohybridized to the sample tissue. This approach enables simultaneous detection of the gene of interest and the control chromosome from a single sample slide and is valuable when distinguishing chromosomal aneuploidy from a low amplification of the gene (Kallioniemi et al.

1992, McCormick et al. 2002, Loring et al. 2005, Saez et al. 2006). The evaluation is done by determining the average ratio of HER-2 to chromosome 17 copy number per cell (Kallioniemi et al. 1992), and is thus less prone to subjective interpretation and variability than IHC. The controlling of this method is easy and effective as normal copy number must be seen in the surrounding normal tissue to approve the procedure as successful and reliable. If the HER-2/chromosome 17 ratio is more than two, the sample is considered amplified (Kallioniemi et al. 1992). This cut-off for amplification has not been confirmed and relies solely on the first publication of HER-2 assessment conducted by FISH (Kallioniemi et al. 1992). Earlier, more tumor cells were counted in HER-2 FISH assessment from tissue samples (Hyytinen et al. 1994, Pauletti et al. 1996, Tanner et al. 2000, McCormick et al. 2002, Loring et al. 2005), but the number of cells used has declined, for example only 20 tumor cells are analyzed in the FDA-approved Pathvysion test.

Fluorescent methods are distinct from other diagnostic tests routinely used in pathology laboratories, and they also possess other problematic characteristics. The technique requires expertise in fluorescence applications and especially in fluorescence microscopy and evaluation. Notably, the histologic view of the sample is poor when using fluorochrome-labeled probes, potentially leading to evaluation of non-cancerous areas. Furthermore, the probes and the fluorescence microscopes are expensive, and the sample is preserved no more than a few months and thus can be filed to the archives only as photos. These problems have been the inspiration to generate an enzyme based chromogenic in situ hybridization method for gene copy number assessment (Tanner et al. 2000).

4.1.2 Detection of HER-2 gene amplification by CISH

The chromogenic in situ hybridization (CISH) method mostly corresponds with FISH, as in both methods DNA probes are hybridized to and visualized from a sample tissue. In CISH, detection is conducted by hapten labeled probes recognized by peroxidase- or alkaline phosphatase-labeled reporter antibodies that are visualized using an enzymatic reaction and chromogen (Tanner et al.

2000). The current CISH HER-2 assessment is based on single-color detection of a digoxigenin-labeled HER-2 probe. In equivocal cases, additional control hybridization with biotin- or digoxigenin-labeled chromosome 17 centromere probe is conducted on a separate sample slide (Tanner et al. 2000, Isola et al. 2004).

The gene copy number limit for amplification most widely used in CISH has been 6 or more copies of the HER-2 oncogene in >50% of the tumor cells, but in equivocal cases when the HER-2/chromosome 17 ratio is used, the limit for amplification is the same as in FISH (Tanner et al. 2000, Isola et al. 2004). Chromosome 17 centromere detection is estimated to be necessary in 7-21% of tumors (Isola et al. 2004, Saez et al. 2006), causing extra work when a second hybridization has to be done on a separate sample slide. The only drawback in the widespread use of CISH has been this limitation to a single-probe application, as there is no doubt about the accuracy of CISH in HER-2 assessment. Numerous studies have reported an excellent concordance between FISH and CISH in formalin-fixed paraffin-embedded samples (Tanner et al. 2000, Arnould et al. 2003, Gupta et al. 2003, Park et al. 2003, Isola et al. 2004, Hauser-Kronberger and Dandachi 2004, Vera-Román et al. 2004, Wixom et al. 2004, Bhargava et al. 2005, Gong et al. 2005b, Loring et al. 2005, Hanna and Kwok 2006, Saez et al. 2006) and in cytological specimens (Lin et al. 2005). Despite the single-probe limitation, CISH has gained a lot of attention and approval as it can be viewed with a standard brightfield microscope along with a hematoxylin counterstain and a good histologic view making the evaluation more convenient and familiar to pathologists (Tanner et al. 2000). Further, the ability to archive the CISH slides is almost indefinite, and the method is easily applied to routine immunohistochemical staining protocols and robots in use in almost every pathology laboratory.

AIMS OF THE STUDY

The aims of the present study were

- To establish a simple immunohistochemical detection method for basal phenotype breast cancer.
- To study the prevalence, phenotypic stability, and luminal CK8/18 expression of basal phenotype tumors.
- To determine the clinicopathologic, biologic, and oncogenic associations of basal phenotype tumors.
- To investigate the prognosis and gene expression profile of basal phenotype breast cancer, especially within the estrogen receptor negative tumor group.
- To explore the heterogeneity of basal cytokeratin expression and its biological and prognostic associations.
- To generate a dual-color CISH method for simultaneous detection of the HER-2 oncogene and its reference probe on the chromosome 17 centromere.

MATERIALS AND METHODS

1. Tumor samples and cell lines

1.1 Study I

The study material consisted of a population-based cohort of 288 consecutive sporadic invasive ductal breast cancers derived from the archives of the Department of Pathology at Seinäjoki Central Hospital. Histopathological information was collected for each patient sample including grade, tumor size, lymph node metastasis, ER, PR, and HER-2. Separate sets of 27 tumors from BRCA1 germ-line mutation carriers, 15 tumors from BRCA2 germ-line mutation carriers (both from the University of Lund, Department of Oncology, Lund, Sweden), 141 HER-2 amplified cancers (Seinäjoki Central Hospital and the Institute of Medical technology, University of Tampere), and 38 pairs of primary and metastatic carcinomas (Seinäjoki Central Hospital and the Institute of Medical technology, University of Tampere) were also studied. Mutation analyses of BRCA1 and BRCA2 have been described previously (Loman et al. 2001).

To optimize and validate the basal phenotype immunostaining method, a subset (n=101) of the invasive ductal tumors was prepared as five tissue microarray (TMA) blocks, each containing cylinders (1 mm diameter) of carcinoma-containing tumor tissue. All samples were routinely formalin-fixed and paraffin-embedded.

1.2 Study II

The tumor cohort comprised 445 primary stage II breast cancers collected from the South Sweden Health Care Region with approval from the Lund University Hospital ethics committee and was described earlier in more detail (Chebil et al. 2003). In this study, patients treated with tamoxifen for two years with complete follow-up data and uniform method for hormone receptor analysis were

included. The formalin-fixed paraffin-embedded sample material was provided as eight TMAs containing three sample cylinders (diameter 0.6 mm) for each primary tumor. A selected cohort of 100 tumors containing an equal number of ER-positive and ER-negative tumors was used for the cDNA microarray analysis.

1.3 Study III

Routine formalin-fixed and paraffin-embedded tissue samples (n=40) from newly diagnosed invasive breast carcinomas were used for the study. The samples were selected to represent HER-2 non-amplified, amplified, and equivocal border-line copy number cases according to the HER-2 result obtained in routine diagnostics using CISH. Four breast cancer cell lines with normal (MCF-7), border-line (MDA-453), moderately amplified (JIMT-1), and highly amplified (BT-474) HER-2 copy number were included in the study. The cultured cells were formalin-fixed and embedded in paraffin as cell pellets using the thrombin clotting technique.

1.4 Study IV

A population-based cohort of 506 primary invasive breast cancers was studied. Of these, 53 tumors showing at least 5% of CK5/14 positive tumor cells were studied in detail. The CK5/14 negative control group consisted of 45 consecutive CK5/14-negative invasive ductal breast cancers and 22 additional CK5/14-negative and estrogen receptor negative tumors (16 of these were grade III) from the entire patient cohort. All except two of the basal cytokeratin expressing tumors were invasive ductal breast cancers. One CK5/14-positive tumor was diagnosed as a metaplastic carcinoma and the other as a medullary carcinoma. To study the persistence of CK5/14 heterogeneity, we studied four pairs of CK5/14-positive primary tumors and their metachronous metastases. The tumors and clinicopathological data were collected from the archives of the Department of Pathology at

Seinäjoki Central Hospital, Seinäjoki, Finland, with permission from the ethical committee of Seinäjoki Central Hospital.

To evaluate the prognostic impact of the intratumoral heterogeneity of CK5/14 and the treatment response of basal phenotype tumors to adjuvant chemotherapy, we studied a separate cohort of 382 tumors from a randomized adjuvant chemotherapy trial of high-risk breast cancer patients (trial SBG 9401, Bergh et al. 2000, Tanner et al. 2006). In brief, this cohort comprised high-risk breast cancer patients with eight or more positive lymph nodes or five or more involved lymph nodes and negative hormone receptor status, and either nuclear anaplasia grade II-III or a high S-phase fraction. The patients were adjuvantly treated with either nine courses of dose-escalated 5-fluorouracil, epirubicin, and cyclophosphamide (FEC) or 3-4 courses of standard FEC followed by high-dose cyclophosphamide, thiotepa, and carboplatin supported by autologous bone marrow support (CTCb). All tumor samples in this study were routinely fixed with formalin, embedded in paraffin and sections of three to five micrometer thickness were obtained.

2. Immunohistochemistry (I,II,IV)

For immunohistochemical (IHC) staining the sample slides were first deparaffinized and rehydrated and subsequently pre-treated. Pre-treatment for all antibodies (Table 1) was carried out in an autoclave at 103 °C for 5 min with 0.05 M Tris-HCl buffer, pH 9.0 containing 0.001 M EDTA. For CK8, CK18, and CK19 antibodies, a subsequent pre-treatment with proteinase K (ChemMate Proteinase K, DakoCytomation, Glostrub, Denmark) was done at room temperature (RT) for 10 min.

Immunostainings were carried out with a Techmate 500+ (DakoCytomation) automated immunostainer. Endogenous peroxidase was blocked with hydrogen peroxide (ChemMate Peroxidase-Blocking Solution, DakoCytomation) for 3 × 2 min 30 sec followed by 30 min primary antibody incubation. Information on the primary antibodies used in this study is presented in Table 3. A Powervision+ polymer kit (PowerVision+^{TM}, Immunovision Technologies

Co., Brisbane, CA) was used for detection. To prevent overstaining of p53, this antibody was detected by the less sensitive avidin-biotin based detection method Vectastain Universal ABC kit (Vector Laboratories, Burlingame, CA). The reactions were visualized with diaminobenzidine (DAB) chromogen (2 × 5 min, Liquid DAB+, DakoCytomation) and enhanced with 0.5% $CuSO_4$ for 5 min. Hematoxylin (Chemmate hematoxylin, DakoCytomation) was used as a counter stain.

Table 2. The primary antibodies used.

Antibody	Clone	Distributor	Dilution
CK5	XM26	Novocastra	1:300
CK14	LL002	Novocastra	1:300
p63	4A4+Y4A3	Neomarkers	1:1500
CK5/CK14/p63	above clones	above firms	1:400/1:400/1:1500
CK8/18	5D3	Novocastra	1:400
CK8	TC1	Neomarkers	1:800
CK18	DC10	DakoCytomation	1:300
CK19	RCK108	Eurodiagnostica	1:200
CK17	E3	LabVision	1:50
smooth muscle actin	1A4	DakoCytomation	ready to use
Bcl-2	124	DakoCytomation	1:700
c-kit	polyclonal	DakoCytomation	1:200
EGFR	EGFR.113	Novocastra	1:100
Ki-67	MIB-1	DakoCytomation	1:1000
p53	DO-7	Novocastra	1:500
vimentin	3B4	DakoCytomation	1:1000

2.1 The optimization of basal phenotype breast cancer detection (I)

In order to identify basal phenotype tumors easily and efficiently, we created an antibody cocktail immunostaining for mammary gland basal cells. We tested antibodies against: CK5, clone XM26

(Novocastra, Newcastle upon Tyne, UK); CK14, clone LL002 (Novocastra); CK5/6, clone D5/16B4 (DakoCytomation), and p63, clone 4A4+Y4A3 (Neomarkers, Fremont, CA). Five different antigen retrieval pre-treatments were compared for all the antibodies using adjacent TMA sections: protease treatment at 37 °C for 3 min 30 sec (protease from Bacillus Licheniformis, Sigma), heat-treatment in an autoclave at 103 °C for 5 min with 0.05 M Tris-HCl buffer, pH 9.0, containing 0.001 M EDTA and with Antigen Retrieval AR-10 Solution (Biogenex, San Ramon, CA), and the same autoclave pre-treatment with the two above-mentined buffers both followed by proteinase K treatment as described above. After optimization of the pretreatment for each of the antibodies, the optimization continued by finding the optimal dilutions for each antibody. A triple antibody cocktail (CK5/CK14/p63, Table 3) was found to be most suitable to detect basal cells. Basal cell carcinoma, skin, and normal breast tissue were used as positive controls for CK5/CK14/p63 immunostaining. For negative controls the primary antibodies were omitted.

To ascertain the effectiveness of the CK5/CK14/p63 antibody cocktail in the detection of basal phenotype tumors, we also studied whether cytokeratin 17 could detect additional basal phenotype breast cancers among 110 invasive ductal cancers with known CK5/CK14/p63 status.

2.2 Double immunostaining of CK5/14 and Ki-67 (IV)

A sequential two-color immunostaining was used to characterize the proliferative activity of CK5/14-positive and CK5/14-negative tumor cells among tumors that were heterogeneously positive for CK5/14. The slides were first immunostained with Ki-67 (as described above) by using DAB as a chromogen (brown reaction product) and subsequently with the antibody cocktail CK5/14 (as described above for CK5 and CK14) by using 3-amino 9-ethyl-carbazole (AEC, red reaction product) as a chromogen. The pretreatment was conducted prior to the first immunostaining as described for CK5, CK14, and Ki-67. Both antibodies were detected using the PowerVision+

detection method as described above. The enhancement of DAB with $CuSO_4$ was omitted in the double immunostaining.

2.3 Interpretation of the immunohistochemical stainings (I,II,IV)

The specimens were interpreted as positive for CK5 and/or CK14 if more than 20% of the neoplastic cells showed cytoplasmic staining and positive for p63 when the staining was nuclear. The 20% limit for positivity was also used for CK8/18, CK8, CK18, CK19, CK17, Bcl-2, c-kit, vimentin, and p53 antibodies. In the cohort of primary stage II breast cancers (II), the stained slides were scanned for virtual microscopy and analyzed through internet (Lundin et al. 2004).

When studying the heterogeneity of basal cytokeratin expression (IV), the percentage of CK5/14-positive malignant epithelial cells from the tumors regarded as basal phenotype (≥5% CK5/14-positive tumor cells was used as a limit for positivity in study IV) was defined using an Olympus BX61 microscope and AnalySIS image analysis software (Soft Imaging System GmbH, Münster, Germany). At least 100 tumor cells were counted from two to five visually selected fields. The Ki-67 labeling index was defined using the same method. In study IV, CK17 and CK8/18 were classified as negative (<5%), heterogeneously positive (5%-69%), or uniformly positive (≥70% of immunopositive tumor cells). EGFR immunohistochemistry was scored on a four-step scale (-, +, ++, and +++) with scores of ++ and +++ regarded as overexpression. For the survival and treatment response study on the 382 high-risk breast cancer patients (IV), the CK5/14 staining was classified as negative (<5%), heterogeneously positive (5%-69%), or uniformly positive (≥70% of immunopositive tumor cells).

3. In situ hybridization (I-IV)

The amplification of the HER-2 and EGFR oncogenes were studied using the chromogenic in situ hybridization (CISH) method as described earlier (Isola et al. 2004, Järvelä et al. 2006,

respectively). HER-2 and EGFR oncogenes were considered amplified when CISH revealed six or more gene copies per cell in at least 10% of the tumor cells.

3.1 Dual-color chromogenic in situ hybridization (dc-CISH, III)

3.1.1 Pretreatment

The slides for dual-color chromogenic in situ hybridization (dc-CISH) were de-paraffinized and incubated in 0.01 M Tris-HCl with 0.001 M EDTA at pH 9.0 in an immunostaining pre-treatment incubator (PT-Module, LabVision, Fremont, CA) at 98 °C-99 °C for 15 min, followed by a cooling period of 20 min. After a wash with phosphate-buffered saline (PBS), enzymatic digestion was carried out by applying 100 µl of digestion enzyme (Digest-All III solution, Zymed Inc., South San Francisco, CA) to the slides for 1-2 min at RT. The slides were then washed with PBS and dehydrated with graded ethanols. For cases that were underdigested in this short pretreatment protocol, an alternative pretreatment was carried out according to the FISH pretreatment protocol (see below).

3.1.2 Hybridization

A bacterial artificial chromosome (BAC) clone specific to the HER-2 DNA (Invitrogen, Carlsbad, CA) was labelled with dUTP-digoxigenin (Roche Biochemicals, Mannheim, Germany) and the chromosome 17 centromere probe (p17H8) was labelled with dUTP-biotin (Roche Biochemicals) using nick-translation (Hyytinen et al. 1994). The probe mixture (1.5 µl nick-translated HER-2 probe and 0.5 µl nick-translated chromosome 17 centromere probe [both 250 ng/µl], 1.0 µl placental DNA [1 µg/µl, Sigma], 0.5 µl human Cot-1 DNA [1 µg/µl, Roche Biochemicals], and 6.5 µl hybridization buffer containing 15% w/v dextran sulphate and 70% formamide in standard saline citrate [SSC] at pH 7.0) was applied to the slides, and they were then covered with 18 × 18 mm

cover slips and sealed with rubber cement. The sections were denatured on a thermal plate (3 min at 94 °C) and hybridized overnight at 42 °C. After hybridization, the cover slips were removed and the slides were washed with 0.5 × SSC (3 min at 72 °C), followed by another wash (1 min at RT) with the same buffer.

3.1.3 Detection of the hybridized probes

The probe for the chromosome 17 centromere was detected first by means of sequential incubations with mouse anti-biotin antibody (Z021, 1:300, Zymed Inc.) and reagents from the Powervision+ alkaline phosphatase polymer kit using New Fuchsin as a chromogen (Immunovision Technologies Co.). After the enzymatic reaction, the slides were washed with distilled water, and the detection of the digoxigenin-labelled HER-2 probe was conducted as follows. The slides were incubated with anti-digoxigenin antibody (1.71.256, 1:300, Roche Biochemicals) and reagents from the Powervision+ horseradish peroxidase polymer kit (Immunovision Technologies Co.). A ready-to-use tetramethyl benzidine (TMB) solution was used as a chromogen (Research Diagnostics, Inc. division of Fitzgerald Industries International, Inc., Concord, MA). The tissue sections were counterstained with hematoxylin, cleared, and dehydrated with graded ethanol and xylene and then embedded.

3.2 Fluorescence in situ hybridization (III)

Fluorescence in situ hybridization (FISH) was performed using a commercially available probe mixture of HER-2 (SpectrumGreen), chromosome 17 centromere (SpectrumAqua), and topoisomerase IIα (SpectrumOrange) using the recommended protocol (Vysis Inc., Abbott Laboratories, Des Plaines, IL). The slides were pretreated with 0.2 M HCl for 20 min and subsequently with a 0.01 M citric acid buffer with 0.05% citraconic anhydride at 98 °C for 15 min.

Enzymatic digestion was conducted at 37 °C for 20-25 min with the Digest-All III solution. The slides were post-fixed with 10% formaldehyde for 10 min and dehydrated with graded ethanols.

3.3 Scoring of dc-CISH and FISH samples (III)

The dc-CISH hybridizations were evaluated with an Olympus BX61 microscope using a 40× objective. For dc-CISH, three different tumor areas were visually chosen and at least 100 tumor cells were scored. Counting was performed using a CCD camera live image and the TouchCount mode of the AnalySIS imaging system (Soft Imaging Systems GmbH). In the TouchCount mode, the copy number is assessed by clicking on each gene copy with the mouse, which counts and marks each object on a live camera image with an overlayed graphic symbol. FISH scoring was conducted in the same manner except that a 60× oil-immersion objective was used. SpectrumGreen and SpectrumAqua fluorescence (for HER-2 and chromosome 17 centromere, respectively) were inspected with the appropriate single-band pass filters. SpectrumOrange fluorescence (for topoisomerase II-alpha, included in the probe) was ignored in this study. A minimum of 20 cells was counted for each FISH sample. All dc-CISH hybridizations were evaluated by an observer unaware of the results of the FISH assays.

4. Statistical methods (I-IV)

Fisher's exact test and the Chi-square test were used to test the significance of the cross-tabulated data (using Stata 9.2 [Stata Corporation, College Station, TX], MedCalc [MedCalc Software, Mariakerke, Belgium], and GraphPad Instat [GraphPad Software, San Diego, CA] statistical software). The comparisons of Ki-67 labeling indices between tumor groups were done with the Mann-Whitney test and between CK5/14-positive and CK5/14-negative tumor cells in basoluminal tumors with the Wilcoxon matched pairs test. The correspondence of the Ki-67 labeling index of CK5/14-positive and CK5/14-negative tumor cells within a tumor was studied with linear regression. Survival analyses were calculated using Kaplan-Maier life table curves and the log-rank

test. Distant disease-free survival was calculated from the primary diagnosis to the date of an event (distant recurrence or death) or for event-free patients to the date of the most recent follow-up. Relapse-free survival was calculated from the primary diagnosis to the first reported breast cancer specific recurrence excluding contralateral breast cancer. Correlation of HER-2/17 centromere ratios between dc-CISH and FISH was counted using the Pearson correlation. All reported p-values are two-sided.

5. Gene expression microarray (II)

5.1 Array design

cDNA microrrays were manufactured in the SWEGENE Microarray Facility, Department of Oncology, Lund University. The gene set consisted of 24,301 sequence-verified IMAGE clones (Research Genetics, Huntsville, AL), and 1,296 internally-generated clones, together representing ≈16,000 Unigene clusters (build 180) and ≈1,200 unclustered expressed sequence tags (EST). The clones were PCR amplified using vector-specific primers essentially as previously described (Khan et al. 2002).

5.2 RNA isolation and microarray hybridization

A selected subset (n=100, of which 50 were ER-negative) from the total cohort of 445 primary stage II breast cancers was analyzed with microarrays. Nineteen of these tumors showed positive CK5/14 staining and the rest were negative. Only one of the CK5/14-positive tumors was ER-positive. Total RNA was extracted from grossly dissected frozen tissue samples (approximately 100 mg) by subsequent Trizol (Invitrogen, Carlsbad, CA) and the RNeasy kit (Qiagen, Hilden, Germany). For each hybridization, 15 μg of Universal Human Reference RNA (Stratagene, La Jolla, CA) was used to synthesize reference Cy5-labeled targets, and 25 μg of sample total RNA for Cy3-labeled targets using anchored oligo(dT) primers and the CyScribe indirect amino-allyl cDNA

synthesis and labeling protocol and GFX purification columns (Amersham Biosciences, Buckinghamshire, UK). Together with blocking agents (12 μg poly-d(A), 6 μg yeast tRNA, and 20 μg Cot-1 DNA), targets were hybridized to the microarrays for 18 hours under a glass coverslip using humidified Corning hybridization chambers at 42 °C and the Pronto Universal Hybridization System (Corning Inc., Corning, NY). Slides were scanned at 10 μm resolution in an Agilent DNA Microarray Scanner (Agilent Technologies, Palo Alto**,** CA) and the images were analyzed using GenePix Pro software (Axon Instruments, Union City, CA).

5.3 Microarray data analysis

The data was analyzed using the BASE (BioArray Software Environment) software (Saal et al. 2002). In brief, background-corrected intensities for sample and reference channels were calculated by subtracting the median local background signal from the median foreground signal for each spot. Filters were applied to remove all spots flagged during image analysis. Data within individual arrays were then normalized using an implementation of the lowess algorithm (Yang et al. 2002). Poorly-measured/expressed spots with a signal-to-noise ratio ≤3 in either the Cy3 or Cy5 channel were removed, and genes with missing data in more than 20 percent of all arrays or genes with a variation across arrays of ≤0.45 standard deviations of the $\log_2$(ratio) were filtered, leaving 10,479 informative genes. The expression ratios for each gene were then median-centered across all tumors.

To generate a gene list for the basal phenotype tumors, correlation scores were calculated between gene expression ($\log_2$(ratio)) for all reporters and the CK5/14 immunopositive tumors (Golub et al. 1999). To evaluate the significance of the expression signatures between the two annotation classes (CK5/14-positive and CK5/14-negative), 1,000 permutations were run where the samples were randomly given an annotation label and the p-value for a score was calculated as the average number of reporters exceeding the score in the permutation test, divided by the total

number of reporters in the gene list. The false discovery rate (FDR), i.e. the estimated number of genes per a given set of scored genes that could receive an equal or better score by chance, was calculated by random permutations and used as an indicator of the robustness of the gene expression profile. An FDR of 0 percent indicates no false positives whereas an FDR of 100 percent indicates a completely random signal. Gene expression profiles were analyzed with hierarchical clustering using a centered Pearson correlation and average linkage clustering (Eisen et al. 1998).

The ranked gene lists were subject to gene ontology annotation analysis using EASE (Expression Analysis Systematic Explorer; http://david.niaid.nih.gov/david/ease.htm) whereby only biological process ontology categories were included and the enrichment of categories in the gene list was evaluated by comparison to the total list of genes used for the microarray analysis. An EASE score of $p \leq 0.05$ was considered to be significant. The UniGene clusters representing the top 200 genes were annotated with subcellular location by cross-reference to two published microarray datasets (Eisen et al. 1998, Diehn et al. 2000) and to Swiss-Prot. The Swiss-Prot Subcellular Locations annotations were downloaded from the DRAGON database (Bouton and Pevsner 2000). A gene was considered to be membrane associated or secreted if the Swiss-Prot annotation contained one of the words “membrane”, “vesicle” or “secreted”, or if the membrane to cytosolic ratio in the polysome fraction study exceeded 2 or 1.08 in the studies by Diehn et al. (2000) or Stitziel et al. (2004), respectively. Primary expression data is available from the NCBI Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/).

RESULTS

1. Basal phenotype breast cancer

1.1 Optimization of basal phenotype breast cancer detection (I)

Tissue microarrays consisting of 97 analyzable primary breast tumors were used to optimize the identification of basal phenotype breast cancer. The two basal cytokeratins (CK5 and CK14) did not identify exactly the same tumors when used alone. Among the 97 tumors there were seven positive for CK5 and six positive for CK14. Two of the CK5-positive tumors were negative for CK14, and one tumor showed reverse results. The transcription factor p63 was expressed in the basal cells of normal breast ducts and was found in three CK5/14-negative tumors and in one CK5/14-positive tumor. To ensure detection of basal phenotype breast cancers with high sensitivity, a cocktail of the three antibodies was used in subsequent experiments. Use of the CK5/CK14/p63 antibody cocktail was further validated by comparing the results with those from tests with single antibodies. Only one tumor of the 97 showed a discordant result.

To ascertain that an additional basal CK17 antibody does not identify more basal tumors than found with the antibody cocktail CK5/CK14/p63, we stained 110 invasive ductal breast cancers with CK17. Only one of the 11 CK17-positive tumors was CK5/14-negative. Half of the CK5/14-positive tumors (10/20) showed a positive CK17 immunophenotype.

1.2 Prevalence of basal phenotype breast cancer (I,II,IV)

In this study 288 sporadic invasive ductal breast cancers were stained as whole tissue samples and 445 sporadic stage II breast cancers in TMAs (375 analyzable tumors) for CK5/CK14/p63. The limit was set to 20% of positive tumor cells for both cytoplasmic cytokeratin and nuclear p63 positivity. The prevalence of CK5/CK14 positive tumors was 9% and 13% and of p63-positive

tumors was 4% and 3.5%, respectively. Although CK5/14 and p63 were coexpressed in normal mammary gland basal cells, there was no association in malignant tumors. When high-risk breast cancer tumor samples (SBG 9401) were studied for the basal phenotype, 19% of the 382 tumors showed CK5/14 expression in at least 5% of the tumor cells, which was used as a cut-off for positivity in this cohort.

The prevalence of the basal phenotype was also studied in hereditary breast cancers. Of the 27 hereditary BRCA1 germ-line mutated tumors, 78% were positive for CK5/14 and one was also positive for p63. Only 1 of the 15 BRCA2-associated tumors showed CK5/14 expression.

1.3 Sub-stratification to basoluminal and basal subtypes (IV)

As basal cytokeratin expression often showed up as heterogeneous staining of the tumor cells, the exact proportion of CK5/14-positive tumor cells in basal phenotype tumors was studied to find out if this phenomenon shows any association with the clinicopathologic, biologic, or prognostic characteristics of the tumors.

For this reason 506 invasive breast tumors were screened by setting the cut-off at 5% of CK5/14-positive tumor cells, which according to our experience is the lowest fraction that can be defined reproducibly in an immunohistochemical staining. The fraction of CK5/14-positive tumor cells was defined for all 53 resulting CK5/14-positive tumors by image analysis. Based on the two-peak distribution of the CK5/14 positivity, we classified the tumors expressing CK5/14 into two subtypes by setting an arbitrary cut-off at 70% of the CK5/14-positive tumor cells. Approximately half of all CK5/14-positive tumors (58%) showed a microscopically distinct heterogeneous immunostaining (median, 32% of positively stained cells) and were called "basoluminal". On immunostaining, basoluminal tumors often showed a focal checkerboard pattern with CK5/14-negative and CK5/14-positive tumor cells located next to each other. The remaining tumors (42%)

stained uniformly or almost uniformly with CK5/14 (median, 94% of positively stained cells) and were called "basal".

To further validate the classification based on intratumoral heterogeneity of basal cytokeratin expression, we stained the same 53 tumors with an additional basal CK17 antibody. The CK17 staining type (negative, heterogeneous, uniform) correlated strongly with the CK5/14 staining type ($p<0.0001$).

1.4 Stability of the basal phenotype in metastasis (I,IV)

In order to ascertain the stability of the basal phenotype and its basoluminal and basal subtypes in metastasis, we studied matched pairs of primary tumors and metachronous metastases from 38 patients. The primary tumor and metastasis pairs were always concordant for CK5/14 expression. There were four CK5/14-positive primary tumor metastasis pairs. We further divided the four CK5/14-positive primary tumor and metastasis pairs into basoluminal and basal subtypes. All metastases showed the same CK5/14 staining type as was observed in the primary tumor. The only basal tumor metastasized to the chest wall and the three basoluminal tumors metastasized to the brain, subcutis, and ovary.

1.5 Expression of luminal cytokeratins in basal phenotype tumors (I,IV)

In order to study the origin and biology of tumors with the basal phenotype in more detail, the possible co-expression of luminal cytokeratins was examined. Luminal CK8/18 staining was conducted on 207 sporadic invasive ductal breast tumors, 116 HER-2 amplified, 26 BRCA1, and 15 BRCA2 tumors with known CK5/14 statuses. All cohorts except BRCA1 tumors showed strong CK8/18 immunopositivity in every tumor regardless of CK5/14 expression. Only BRCA1 germ-line mutated tumors showed the CK5/14-positive CK8/18-negative phenotype. Out of 20 CK5/14-positive BRCA1-associated tumors, five were CK8/18-negative and five showed weak staining

intensity. The rest were strongly positive for CK8/18. In accordance with the results from sporadic tumors, both basoluminal (n=31) and basal (n=23) tumor subtypes were always CK8/18-positive, and furthermore showed uniform CK8/18 immunostaining (≥70% positive tumor cells) in 90% and in 95% of the tumors, respectively.

While luminal glandular epithelial cell features were present in basal phenotype tumors, they did not show the characteristics of normal breast basal myoepithelial cells, as CK5/14-positive tumors showed the smooth muscle actin negative phenotype.

1.6 Clinicopathological and biological associations of basal phenotype breast cancer (I, II, IV)

The CK5/14 immunoreactivity did not correlate to patient age or tumor size. The CK5/14 expression showed no association to axillary lymph node metastasis in studies I and IV but an association to negative lymph node status was found in study II ($p=0.0005$). This discrepancy might be partly due to sample selection. A strong association between CK5/14 positivity and histological tumor grade III was found (I; $p=0.0007$, IV; $p=0.0002$) but even more striking was the association of CK5/14 with negative ER and PR status ($p<0.0001$ for both in I, II, and IV). In addition to the association with grade III and negative steroid hormone receptor status, CK5/14-positive tumors frequently overexpressed p53 (II; $p=0.003$, IV; $p=0.0002$) and EGFR (IV; $p<0.0001$). A strong association with negative Bcl-2 status and immunopositivity for c-kit and vimentin was also evident (IV; $p<0.0001$ for all). When the association of CK5/14 immunopositivity with p53 overexpression and negative lymph node status found in study II were investigated within the ER-negative tumor subgroup no significant associations were found.

In a similar comparison between the CK5/14-positive basoluminal and basal tumor subtypes, both tumor subtypes were predominantly hormone receptor negative and of a high histological grade. Basoluminal tumors were larger but less frequently vimentin- and c-kit-positive

than the basal tumors (IV; p=0.02, p=0.005, and p=0.02, respectively). There was no difference in EGFR, Bcl-2, or p53 expression between basoluminal and basal tumors. The majority of the basal cytokeratin positive tumors were of the ductal histotype (80%) and the rest were of the medullary or atypical medullary histotypes (20%) when studied from the ER-negative subgroup of 375 stage II breast cancers collected from Sweden (II).

1.7 Association of the basal phenotype with HER-2 amplification (I,II,IV)

To study if there are basal phenotype tumors with HER-2 amplification, CK5/14 immunopositivity was examined among 288 invasive breast cancers, 375 stage II breast cancers, and 141 HER-2 amplified tumors. A total of 12%, 10%, and 13% of the HER-2 amplified tumors in these cohorts showed CK5/14 immunopositivity, respectively. In other words, no association between CK5/14 expression and HER-2 amplification was seen. When the association between CK5/14 and HER-2 amplification was studied within the subgroup of ER-negative tumors among 288 invasive breast cancers or 375 stage II breast cancers, there was an inverse association, which was highly statistically significant (p=0.007 and p=0.01, respectively).

Between the basal (n=23) and basoluminal (n=31) tumor subtypes, there was a clear difference in the occurrence of HER-2 amplification. A significant association was observed between the basoluminal tumors and HER-2 oncogene amplification (p=0.009). In fact, the prevalence of HER-2 positivity decreased almost linearly with the increase in the proportion of CK5/14-positive tumor cells. There was no significant difference in the prevalence of HER-2 amplification between basoluminal tumors and luminal estrogen receptor negative tumors (35% versus 59%, p=0.11).

1.8 Proliferation activity of basal phenotype breast cancer (IV)

Proliferation activity of basal phenotype tumors (n=53) and the subtypes basoluminal (n=31) and basal (n=22) was compared to CK5/14-negative tumors (n=41) and separately to tumors, which were additionally ER-negative and histological grade III (n=22). Proliferation activity was defined by counting the exact percentage of Ki-67 immunopositive tumor cells. Tumors expressing basal cytokeratins showed a higher cell proliferation activity than the CK5/14-negative/ER-negative/grade III tumors (median 37% vs. 22%, p=0.003). Most interestingly, the basal tumor subtype showed a much higher proliferation activity than basoluminal tumors. The median Ki-67 labeling index was 33% in the basoluminal tumors in contrast to 58 % in the basal tumors (p=0.0014).

The cell proliferation activity of the CK5/14-positive and CK5/14-negative tumor cell populations from 25 basoluminal tumors was studied with an immunohistochemical double staining. The Ki-67 labeling indices in the two cell populations strongly correlated with each other in the individual tumors (r=0.60, p=0.0017), but the CK5/14-negative tumor cells had a slightly higher proliferation activity than the CK5/14-positive tumor cells (median 30% vs. 19%, p=0.04).

1.9 Gene expression profile of basal phenotype breast cancer (II)

While immunohistochemistry has long been used to detect basal cytokeratin positive breast cancer, a similar tumor group has been found using cDNA microarrays. The gene expression profile of basal cytokeratin immunopositive breast cancers was studied especially within the group of ER-negative tumors (n=50) in order to exclude the strong influence of ER to tumor classification by cDNA microarrays. In this subgroup, CK5/14-positive and CK5/14-negative tumors were associated with two distinct gene expression signatures (False Discovery Rate 6.7% per top 100 genes and 16.1% per top 500 genes). Hierarchical clustering analysis of the ER-negative tumors using the top 500 discriminatory genes for CK5/14 immunopositive tumors identified two separate

clusters. The basal-like cluster contained a large number of CK5/14-positive tumors (17/24) and additionally seven CK5/14-negative tumors. In the non-basal like cluster all but one of the tumors (25/26) were immunohistochemically CK5/14-negative and were frequently HER-2 oncogene amplified (18/26).

Next, the so called "intrinsic" gene set generated by Perou and co-workers (2000) was tested for its performance in this data set. Mapping of the intrinsic gene list from Sørlie et al. (2003) to this data using the Unigene Cluster ID as an identifier produced a list of 522 clones. Hierarchical clustering of the ER-negative tumor group generated a dendrogram with two major subgroups very similar to the hierarchical clustering analysis using our top 500 ranked basal genes (concordance 90%, p=0.0001). The basal-like cluster included the majority of the CK5/14-positive tumors and nine additional CK5/14-negative tumors, and the tumors in the non-basal subgroup showed frequent HER-2 amplification (17/27) and a predominantly CK5/14-negative immunophenotype (23/27). The basal phenotype classification by Sørlie's intrinsic gene set correlated strongly with basal cytokeratin immunohistochemistry (concordance 76%, p=0.0011). Interestingly, seven of the nine misclassified CK5/14-negative tumors by Sørlie´s intrinsic gene set were found to belong to the basal-like cluster when our top 500 CK5/14-associated genes were used in hierarchical clustering analysis.

A gene ontology annotation analysis was performed on the top 1,000 genes on our basal gene list (within ER-negative tumors). The biological process of epidermal differentiation & ectoderm development was the first annotation category for the genes up-regulated in basal-like tumors (with EASE score ≤0.05), including genes for basal cytokeratins 14 and 17. Genes down-regulated in basal phenotype tumors were characterized and shown to have functions in many signaling pathways.

1.10 Prognosis of basal phenotype breast cancer (II,IV)

Association of the basal status with patient prognosis was evaluated both by immunohistochemical and microarray based classification of basal phenotype tumors. First, we studied the prognosis of 375 stage II breast cancers by staining CK5/14 immunohistochemically. There were 48 CK5/14-positive basal phenotype tumors in this cohort. In the whole tumor set, the distant disease-free survival was significantly shorter for the CK5/14-positive tumors during the first years of follow-up (three years p=0.01 and five years p=0.04), but this difference was diminished by the end of the follow-up period (10 years p=0.19). An even more coherent prognostic association for CK5/14-positive and CK5/14-negative tumors was obtained when the cohort of 382 high-risk breast cancer patients from the SBG 9401 study containing 73 CK5/14-positive tumors was studied for relapse-free survival (p=0.52).

Next, the clinical outcome of basal phenotype tumors was studied within the ER-negative tumor subgroup (35/95 CK5/14-positive) from stage II breast cancers. The distant disease-free survival of immunohistochemically CK5/14-positive and CK5/14-negative tumors was identical. The same result was obtained when the basal-like versus non-basal like classification was based on gene expression microarrays within the ER-negative tumor entity (18/50 CK5/14-positive tumors, clustered either with our top 500 basal gene list or Sørlie's intrinsic gene list).

The relapse-free survival of CK5/14-positive tumor subtypes was studied from the cohort of 382 high-risk breast cancer patients. There were 28 basal and 45 basoluminal tumors in this cohort. The basoluminal tumors showed significantly shorter relapse-free survival than the basal tumors (p=0.01). The sub-stratification of the basoluminal group by HER-2 status (p>0.05) shows that the survival difference between basal and basoluminal tumors is not due to more frequent amplification of the HER-2 oncogene in the basoluminal subgroup.

1.11 Basal phenotype breast cancer and response to treatment (IV)

While there was no clear difference between the prognosis of basal phenotype and non-basal breast cancers, the groups were next studied in terms of their response to adjuvant chemotherapy. The treatment response of basal phenotype breast cancer was studied among 382 high-risk breast cancer patients from a randomized adjuvant chemotherapy trial SBG 9401. CK5/14-negative tumors showed very similar results between the chemotherapy regimes (tailored and dose-escalated FEC vs. standard FEC followed by CTCb along with bone marrow support) as noted in the original study (Bergh et al. 2000). FEC treatment gave a better response (p=0.03, Figure 3A). Despite the difference between treatment regimes within CK5/14-negative tumors, no difference was detected between the efficacy of the chemotherapy regimes in basal cytokeratin positive tumors (p=0.82, Figure 3B).

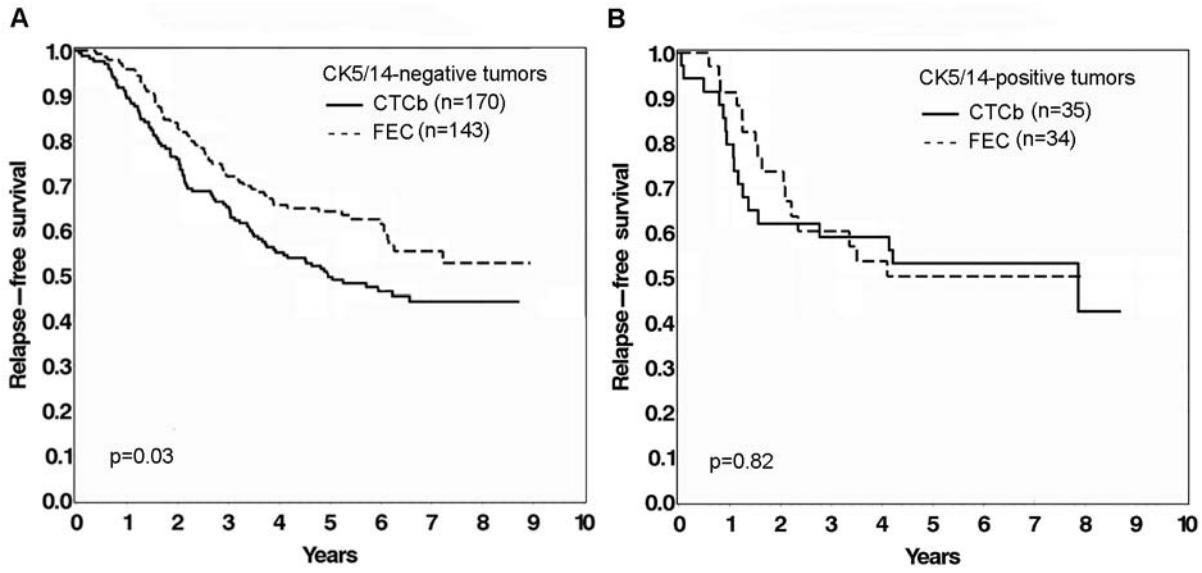

Figure 3. Relapse-free survival of CK5/14-negative (A) and CK5/14-positive (B) breast cancers treated either with nine courses of dose-escalated FEC (FEC) or 3-4 courses of standard FEC followed by high-dose CTC supported by autologous bone marrow support (CTCb). In CK5/14-negative tumors (A) the dose-escalated FEC treatment gave better results but there was no difference between the treatment regimens in basal phenotype CK5/14-positive tumors.

2. Tissue microarrays (TMA) in immunohistochemical staining of heterogeneously expressed biomarkers

In study IV, the percentage of basal cytokeratin positive tumor cells was studied from the original tumor blocks as whole tissue samples. However, tissue cylinders had earlier been punched from some of these tumor blocks when making TMAs for previous needs. When the original tissue block was cut, the empty spots were clearly visible on the slide where ever a cylinder had been taken. In heterogeneously CK5/14-positive basoluminal tumors it could be seen from the stained slide that some tissue cylinders were taken from CK5/14-negative areas. If TMAs would have been used to detect basoluminal tumors in study IV, heterogeneity or even positivity of the sample could have been missed (Figure 4). Thus, when using TMAs the staining pattern of the antigen studied should be known beforehand and the original tissue block should be used if the staining pattern shows marked heterogeneity.

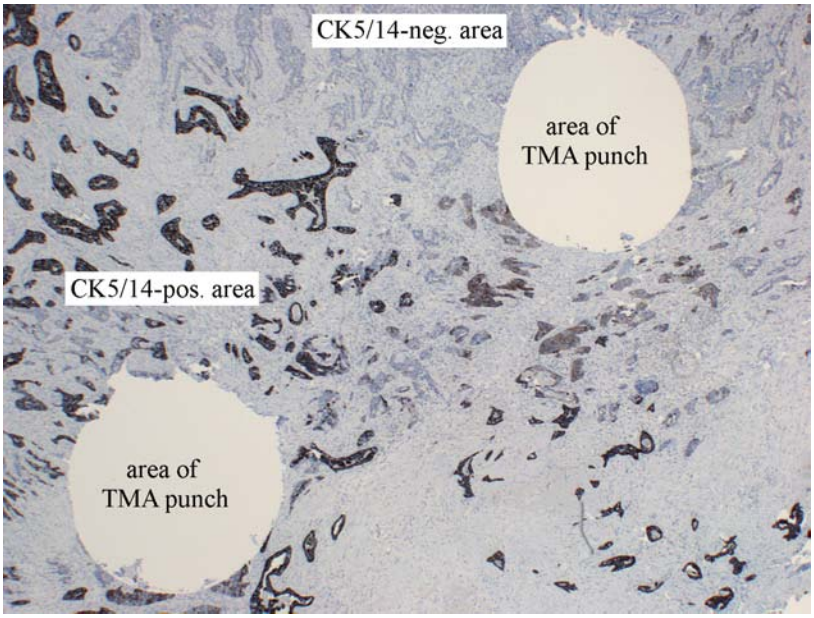

Figure 4. An immunohistochemical CK5/CK14/p63 triple antibody cocktail staining of a tumor block which previously had a tissue core sample punched for TMA construction. This tumor clearly shows a heterogeneous basal cytokeratin expression, and it can be seen that one tissue core was punched from the negative area and the other from an area showing more basal cytokeratin positive tumor cells. The heterogeneity of protein expression in tumors might lead to false negative results in TMAs if using only one tissue core or too small a diameter in the TMA construction.

3. HER-2 assessment by dc-CISH (III)

3.1 Dc-CISH versus FISH

The dc-CISH method resulted in clearly distinguishable signals for HER-2 (green) and the chromosome 17 centromere (red), and it allowed for the counting of separate signals without difficulty. The mean HER-2 and chromosome 17 centromere copy numbers and their ratios were highly concordant between dc-CISH and FISH in both breast cancer cell lines and paraffin-embedded tumor samples representing normal, border-line, and amplified HER-2 copy numbers. Neither the mean copy numbers of HER-2 or chromosome 17 centromere nor their ratio displayed any systematic shift to either direction in dc-CISH when compared to FISH (r=0.89, from all samples that were enumerable with both methods). When the generally accepted cut-off for the HER-2/17 centromere copy number ratio (2) was applied here, there was a 91% agreement between the two methods (Kappa coefficient 0.82). One cell line (MDA-453) showed HER-2 amplification under FISH but not under dc-CISH, and three breast cancers were identified as having a HER-2/17 centromere copy number ratio of 2 or more by dc-CISH but not by FISH. Of the 18 tumors with a highly amplified HER-2 oncogene, 14 showed a high number of gene copies that were clustered together making enumeration by dc-CISH unreliable. When critically scoring the copy numbers on the FISH slides, we considered HER-2 copy number enumeration to be impossible in 11 of these 18 tumors.

DISCUSSION

1. The basal cell specific antibody cocktail in basal phenotype breast cancer detection (I)

Antibody cocktails are becoming more popular in research and also in diagnostic pathology (Tacha et al. 2004). They are as easy to use as single antibodies but only one sample slide has to be stained. Additionally, the target is more sensitively recognized if all the antibodies give information about the same phenomenon. Furthermore, if nuclear and cytoplasmic antibodies are used in a cocktail the localization of the staining will still reveal the specificity if needed.

When optimizing and evaluating the best combination of the antibodies to detect basal phenotype breast cancer, it was seen that both exclusively CK5-positive and exclusively CK14-positive tumors existed. Additionally, p63 positivity, a suggested basal phenotype marker (Reis-Filho et al. 2003, Ribeiro-Silva et al. 2003, Makretsov et al. 2004), was seen in both CK5/14-negative and CK5/14-positive tumors. This led to the assumption that both of the known basal phenotype cytokeratin markers (Moll et al. 1982, Wetzels et al. 1991, Malzahn et al. 1998, Böcker et al. 2002, Chu et al. 2002, Boecker and Buerger 2003, Abd El-Rehim et al. 2004) and additionally p63 are needed for detection, since their expression is not totally corresponding. CK6 is often detected in combination with CK5, but since it is not expressed in the normal mammary gland (Moll et al. 1982, Böcker et al. 2002) and is not described as a basal phenotype marker, this antibody was not included in the cocktail. CK17 was also excluded as it was noted that there was only one CK17-positive CK5/14-negative tumor among 110 invasive ductal carcinomas.

2. Basal phenotype in sporadic and hereditary breast cancer (I,II,IV)

This study showed that the basal phenotype identified by basal cytokeratin immunohistochemistry accounts for 9% of sporadic breast tumors. Since p63 immunopositivity did not correlate to basal

cytokeratin positivity or other basal phenotype breast cancer characteristics, it was not used to define the basal phenotype. When the basal phenotype was studied within ER-negative or histologic grade III tumor entities, the prevalence was 42% and 17%, respectively. As could be expected, the studied cohort of high-risk breast cancer patients contained a higher number of basal phenotype tumors (19%) than the population based cohort. According to this study, the basal phenotype is found in about every tenth diagnosed breast cancer but is overrepresented in the most aggressive subgroup of breast carcinomas, which is in agreement with other studies (Malzahn et al. 1998, Otterbach et al. 2000, Nielsen et al. 2004, Putti et al. 2005, Rodrígues-Pinilla et al. 2006a).

Beyond sporadic tumors, our results support the correlation between the basal phenotype and BRCA1 germ-line mutated tumors (Foulkes et al. 2003, Sørlie et al. 2003). Further, the results show that by conducting the identification with a sensitive polymer-based IHC technology, a very high proportion, almost 80%, of hereditary BRCA1 tumors showed the basal phenotype. Similar figures have been reported by some researchers (Foulkes et al. 2003, Foulkes et al. 2004, Rodrígues-Pinilla et al. 2006a) but others propose a lower prevalence (Lakhani et al. 2005, Palacios et al. 2005, Rodrígues-Pinilla et al. 2006b). This study was among the first (Palacios et al. 2005, Rodrígues-Pinilla et al. 2006a) to show that the other known germ-line mutation causing hereditary breast cancer, BRCA2, is not associated with the basal phenotype.

The most important distinction between sporadic and hereditary basal phenotype tumors was luminal cytokeratin expression. This study is the first to report 100% prevalence of luminal cytokeratin expression in sporadic basal phenotype breast cancer. Others have found a lower proportion positive, and some have also proposed an inverse association between luminal and basal cytokeratin expression in breast cancer (Wetzels et al. 1991, Abd El-Rehim et al. 2004, Birnbaum et al. 2004, Rakha et al. 2006). In contradiction to the sporadic tumors, some basal phenotype BRCA1 germ-line mutated tumors were negative for luminal cytokeratins and hence more closely resemble the true basal cell phenotype of the mammary gland. Basal cytokeratin

expression has actually been assumed to be a good preselectional marker for BRCA1 mutation analysis (Lakhani et al. 2005). This study further shows that when a highly sensitive polymer-based detection method is used, a breast tumor with negative luminal cytokeratin and positive basal cytokeratin status is most likely BRCA1 mutated. Still, a majority of basal phenotype BRCA1 tumors co-expressed basal and luminal cytokeratins similarly to the sporadic basal phenotype breast cancers according to this study. These results are in contradiction with the results of Foulkes et al. (2004), who suggested that BRCA1 tumors would most often be CK8/18-negative.

It has been suggested that mammary gland epithelial cells expressing CK5 only would be stem/progenitor cells capable of differentiation towards luminal and myoepithelial lineages through intermediate phases (Boecker et al. 2002, Böcker et al. 2002, Boecker and Buerger 2003). The co-expression of luminal and basal cytokeratins in sporadic basal phenotype tumors corresponds to the cytokeratin profile of the CK5-positive breast progenitor cell committed to the luminal glandular lineage (Böcker et al. 2002). The commitment to the glandular lineage is further supported by the fact that none of the basal cytokeratin expressing tumors showed smooth muscle actin expression indicative of myoepithelial differentiation. The CK5/14+ CK8/18- phenotype found in some of the BRCA1 mutated tumors resemble the more primitive CK5 only expressing stem/progenitor cells of the mammary gland (Böcker et al. 2002). It has also been suggested that BRCA1 would be a stem cell regulator and promote differentiation towards the glandular lineage (Foulkes 2003, Furuta et al. 2005). The non-functional BRCA1 protein might thus have something to do with the more primitive breast progenitor cell phenotype seen in BRCA1 mutated tumors. These results support the proposal that basal phenotype tumors originate from the CK5-positive mammary gland stem/progenitor cells (Böcker et al. 2002). Still, the existence of breast progenitor cells expressing exclusively CK5 has been under doubt (Clarke CL et al. 2004). Additionally, there is a need to clarify the relationship between basal cytokeratins and smooth muscle actin expression in breast cancer, since it has been stated in contradiction to this study that these proteins might be

co-expressed in some tumors (Hungermann et al. 2005, Jacquemier et al. 2005, Livasy et al. 2006, Rakha et al. 2006).

3. Sub-stratification of CK5/14-positive tumors to basoluminal and basal subtypes (IV)

Although the basal phenotype has been generally regarded a uniform tumor subgroup (Sørlie et al. 2001, Abd El-Rehim et al. 2004, Birnbaum et al. 2004, Gusterson et al. 2005), Dairkee et al. (1987a) noticed already in 1987 that CK14 was expressed heterogeneously in some breast cancers. Still, this phenomenon has not been studied in detail during the past two decades. The same phenomenon of basal cytokeratin heterogeneity was noted in this study and for the first time the exact frequency of basal cytokeratin positive tumor cells was studied in breast cancer. Half of the basal phenotype tumors showed heterogeneous expression with totally positive or negative tumor areas with similar morphology beside each other as in a checkerboard. The other half of the tumors were uniformly basal cytokeratin positive. The naming of these tumors was based on the pattern of co-expression of basal and luminal cytokeratins in these tumor subtypes. Heterogeneously basal cytokeratins expressing tumors were named "basoluminal" since part of the cells in these tumors only showed luminal differentiation while the rest of the tumor cells co-expressed luminal and basal cytokeratins. Breast cancers that express basal cytokeratins uniformly in every tumor cell were named simply "basal" since they did not show clear luminal differentiation but co-expressed basal and luminal cytokeratins through out the tumor. Since the third basal epithelium cytokeratin 17 (Moll et al. 1982, Böcker et al. 2002) confirmed the division into basoluminal and basal subtypes, it seems likely that they are two distinct entities. Still, to confirm and to reproduce these results it is an important future goal to study the gene expression profile of basal and basoluminal tumors.

These results of the two distinct basal phenotype tumor subtypes are not without support, since some microarray studies have reported two separate basal-like breast cancer clusters

(Perou et al. 2000, Sotiriou et al. 2003). Additionally, heterogeneous basal cytokeratin expression has been reported also by other immunohistochemical researchers than Dairkee et al. (1987a). Malzahn et al. (1998) divided the basal cytokeratin immunopositive breast cancers into focal and diffuse subtypes with very similar results to this classification into basoluminal and basal subtypes. Rakha et al. (2006) and Jones et al. (2001) have also reported heterogeneous basal cytokeratin expression. These studies have not calculated the exact proportion of basal cytokeratin positive tumor cells or found any prognostic differences between the formed subgroups. In addition to these results, an interesting study of grade III breast cancers showed that CK14-positive breast cancers divide into two tumor groups with distinct profiles of genetic alterations and prognosis (Jones et al. 2004b). It is interesting to speculate whether the subdivision of the basal phenotype breast cancers into basoluminal and basal subtypes reflects the same tumor subgroups found by Jones et al. (2004b) or other suggested divisions of the basal phenotype breast cancers. It is definite that the relationship between the basal-like gene expression signatures and heterogeneous basal cytokeratin immunopositivity should be studied in detail in future.

The heterogeneous basal cytokeratin expression of the basoluminal subtype tumors gives another point of view to the suggested progenitor cell origin of basal phenotype tumors (Böcker et al. 2002). Since cancers are clonal and derived from one transformed cell, the cytokeratin profile of the tumor should reflect the phenotype of the cell of origin (Moll et al. 1983). The heterogeneous basal cytokeratin expression seen in basoluminal tumors would suggest that the tumor cells might be capable of differentiation towards the glandular lineage just like the normal breast progenitor cells are. This would explain the loss of CK5/14 expression from some of the tumor cells in basoluminal tumors. In basal tumors, which express CK5/14 in the whole tumor this differentiation would be somehow inhibited or prevented and thus the co-expression of luminal and basal cytokeratins typical for breast progenitor cells committed to the glandular lineage would be present throughout the tumor. Since the BRCA1 protein might be functioning in stem/progenitor

cell differentiation (Foulkes 2003, Furuta et al. 2005) and there is evidence of BRCA1 down-regulation in sporadic basal phenotype tumors (Abd El-Rehim et al. 2004, Ribeiro-Silva et al. 2005, Turner et al. 2006), it should be studied whether the downregulation of the BRCA1 protein differs between the basoluminal and basal tumor subtypes. It must be noticed still that the phenotypic variation might also be an effect of the genetic instability of the cancer (Jones et al. 2001).

4. Characteristics of basal phenotype breast cancer

This study shows that the most central features of basal phenotype breast cancer are ER-negative status and high histologic grade. This is consistent with earlier results (Perou et al. 2000, Sørlie et al. 2001, Korsching et al. 2002, Sørlie et al. 2003, Birnbaum et al. 2004, Malzahn et al. 2004, Nielsen et al. 2004). When these main characteristics were studied between the discovered basoluminal and basal tumor subtypes, they showed a high similarity. Basal phenotype breast cancers thus form a homogeneous tumor group in regard to the most prominent characteristics.

The same is not true for all of the biological characteristics associated with basal phenotype breast cancer such as vimentin and c-kit (Thomas et al. 1999, Nielsen et al. 2004, Korsching et al. 2005, Tsuda et al. 2005a). This study confirms the correlation of vimentin and c-kit immunopositivity to basal cytokeratin expression in breast cancer, but further shows that these proteins are more often seen in basal than in basoluminal subtype tumors. These results suggest that basoluminal and basal subtypes are biologically distinct entities. Since c-kit and vimentin expression has been associated with stem/progenitor cells in the mammary gland, these results also support the suggested progenitor cell origin of basal phenotype tumors (Korsching et al. 2005, Miettinen and Lasota 2005, Tsuda et al. 2005b, Sheridan et al. 2006). Other biological features found associated with the basal phenotype in this study were negative Bcl-2 status and overexpression of p53 and EGFR, which is in agreement with the earlier results (Korsching et al. 2002, Foulkes et al. 2004, Nielsen et al. 2004, Korsching et al. 2005, Tsuda et al. 2005a, Livasy et al. 2006, Rakha et al. 2006). It seems that the association between p53 and CK5/14 expression is

mainly due to negative ER status since within the ER-negative subgroup no such association was observed. The finding that basal phenotype tumors associate with EGFR overexpression might be a good starting point to study the efficacy of EGFR inhibitors and antibody therapies in breast cancer (Polychronis et al. 2005).

Even though Bcl-2, p53, and EGFR expression did not enrich to either the basoluminal or basal subtype, the findings from the Ki-67 labeling index strengthened the belief that the subtypes found are separate entities. The basal subtype tumors proliferated more actively than the basoluminal tumors. This proliferation difference seems to be correlated to the tumor subtype rather than to the basal cytokeratin prolifile of individual tumor cells since in basoluminal tumors, CK5/14-negative and CK5/14-positive tumor cells showed a similar proliferative activity. The basal phenotype tumors as a whole showed more prominent proliferation activity than non-basal ER-negative grade III tumors. This indicates that basal cytokeratin expression defines the tumor behavior beyond the ER status and grade. The high proliferation activity is most likely one reason for the recent finding that basal phenotype tumors are more frequent in interval than in screen detected tumors (Collett et al. 2005).

There have been many statements, mainly from microarray studies but also from IHC based classification that HER-2 and the basal phenotype show an inverse correlation in breast cancer (Perou et al. 2000, Sørlie et al. 2001, Korsching et al. 2002, Sørlie et al. 2003, Nielsen et al. 2004). Still, tumors and one cell line co-presenting these features have been found (van de Rijn et al. 2002, Tanner et al. 2004, Kim et al. 2006). This question was thoroughly studied in this work, and the results show that HER-2 amplified basal phenotype tumors exist. The prevalence of the basal phenotype in HER-2 oncogene amplified tumors (12%) is actually very similar to the prevalence in the population based cohort (9%). If the main characteristic shared by the HER-2 and basal phenotype tumors is eliminated and the correlation is studied within the ER-negative tumor subgroup, an inverse association is found. This would suggest that these tumor groups are mainly

distinct entities. HER-2 amplified basal phenotype tumors were found to be almost exclusively basoluminal in subtype, which strengthened the new definition of two biologically separate basal phenotype subtypes, basoluminal and basal. This finding is supported by the results of Rakha et al. (2006), who showed that negative HER-2 status was associated with uniform basal cytokeratin expression. Further, our results show that the lower the basal cytokeratin positive tumor cell proportion was, the higher the prevalence of HER-2 amplification grew. This is a very good indicator of a negative association between basal cytokeratin expression and HER-2 amplification, and shows that these two breast cancer groups are indeed mainly distinct entities. Based on these results, it can be speculated that gene expression microarrays might actually recognize tumors with a higher proportion of basal cytokeratin positive cells as basal-like and at least partly cluster the basoluminal tumors to other subgroups like ERBB2. Additionally, this clearly shows how the final results of immunohistochemical studies may be markedly affected by the limits chosen for positivity.

The evaluation of immunohistochemical stainings has been very variable, and this is also the case for basal phenotype identification by immunohistochemistry. Since in immunohistochemistry only one thin section of the tumor mass is studied and the basal cytokeratin expression might be seen in remnants of the normal basal cell layer, it would be preferable to see more than only a few stained cells in order to define a tumor basal phenotype. A percentage limit like 5%-20% would also be more reproducible in an immunohistochemical setting and give a more strict and reliable definition of the basal phenotype. The fluctuation in the requirements of the basal phenotype definition naturally causes changes also in the correlation to prognosis and to clinicopathological and biological markers as shown in this study. When the proportion of basal cytokeratin expressing tumor cells was studied with the cut-off set to 5%, the identified tumors were ER-negative, high grade, and had high proliferation activity, all known basal phenotype characteristics. Even though, these tumors also showed features like HER-2 amplification, which is

not a characteristic feature for basal phenotype breast cancers, at least according to gene expression microarray studies (Perou et al. 2000, Sørlie et al. 2001, Sotiriou et al. 2003, Sørlie et al. 2003, Sørlie et al. 2006). If the limit for positivity were set to 70%, the carcinomas in this group would form a quite consistent group more often showing vimentin and c-kit immunopositivity and no HER-2 amplification. According to this study, this subgroup would comprise only 4-5% of all invasive breast cancers. EGFR expression would be as common in basal phenotype tumors defined with either the 5% or 70% limit for positivity. Still, EGFR has been suggested to be a good immunohistochemical marker to detect tumors defined as basal-like by gene expression microarrays (Nielsen et al. 2004). A schematic presentation of basal phenotype breast cancer classification according to ER, HER-2 and basal cytokeratin expression is shown in Figure 5.

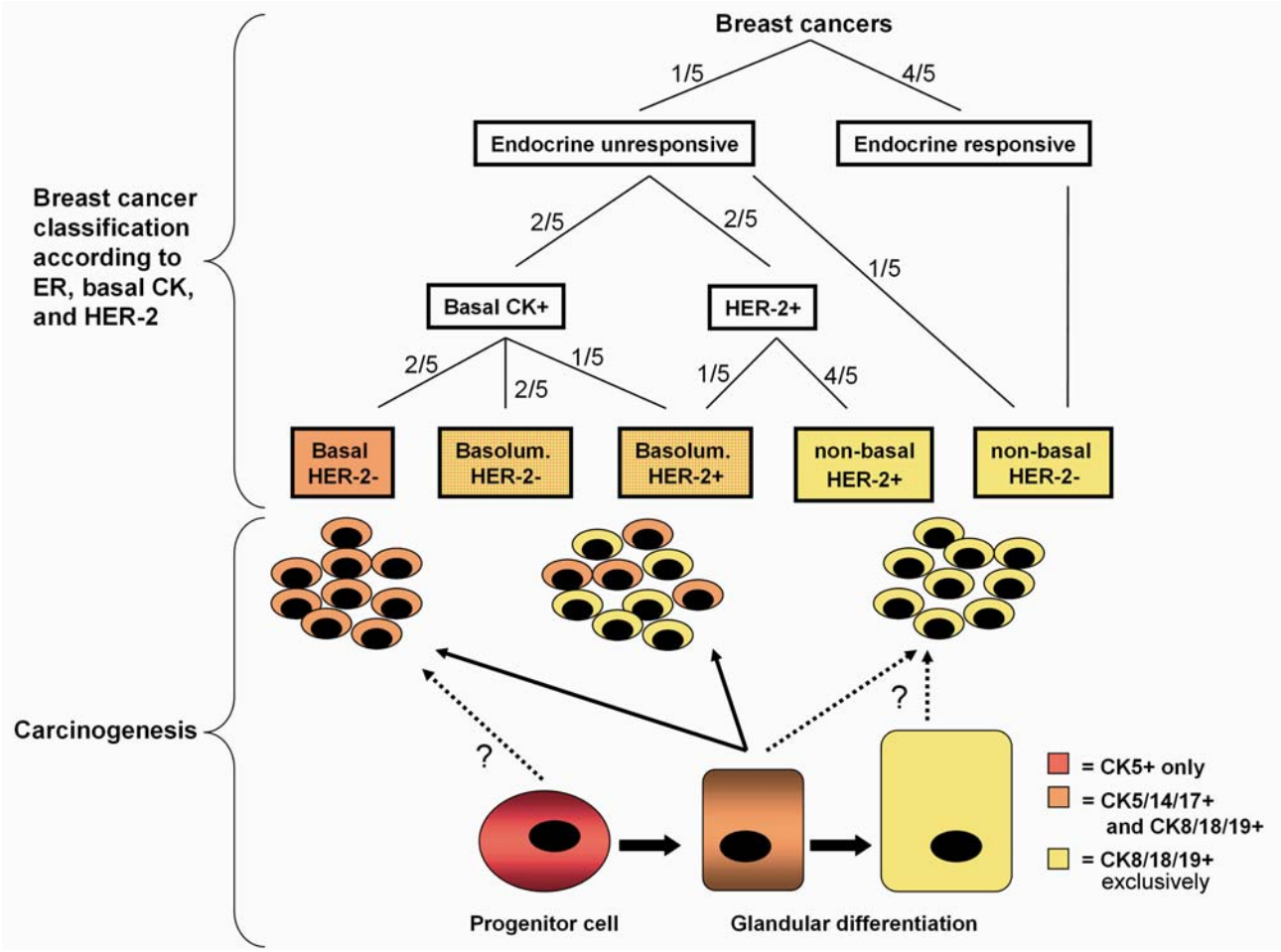

Figure 5. A schematic presentation of breast cancer classification according to ER, HER-2, and basal cytokeratin expression. The average proportions of different subgroups are indicated.

5. Unique gene expression profile of basal phenotype breast cancer (II)

The estrogen receptor has been shown to be a strong factor in the determination of prognosis and gene expression signature in breast cancer (Gruvberger et al. 2001, West et al. 2001, Gruvberger et al. 2002, Pusztai et al. 2003). Even though ER-negativity is also the main characteristic of basal phenotype tumors, there is a biological difference between basal phenotype and other ER-negative tumors. This is demonstrated by the unique gene expression profile found within the ER-negative tumor subgroup for immunohistochemically CK5/14-positive tumors in this study. These results confirm the biological uniqueness of basal phenotype breast cancer not only in the whole population of breast cancers but for the first time also within ER-negative tumors. These results further showed that the basal-like breast cancer cluster included the basal cytokeratin expressing tumors but in addition few immunohistochemically CK5/14-negative breast cancers. These results were confirmed when Sørlie's intrinsic gene list (Sørlie et al. 2001) was used to cluster the ER-negative tumors. The basal-like tumor cluster which resulted was very similar to our clustering and contained most of the immunohistochemically CK5/14-positive tumors and additionally the same CK5/14-negative tumors. This suggests that in addition to basal cytokeratins, other expressional features are involved in the formation of the basal-like cluster. One possibility might be the expression of EGFR, which has been suggested to be a required immunohistochemical marker in addition to basal cytokeratins to be able to define the microarray based basal-like cancer entity as a whole (Nielsen et al. 2004, Livasy et al. 2006). Other candidates are c-kit and vimentin, both of which are known to associate with basal phenotype breast cancer (Nielsen et al. 2004, Korsching et al. 2005). Altogether, gene expression microarray-based method and immunohistochemistry for CK5/14 showed around 75% concordance to identify basal-like breast cancer, but still around 25% of the tumours identified by microarrays would remain undetected by immunohistochemistry. It is thus clear that if the basal-like gene expression profile is the entity that is to be identified immunohistochemistry of CK5/14 is not enough. Still, the basal phenotype tumors identified by

basal cytokeratin immunohistochemistry or microarrays show very similar features and represent mostly the same tumor group.

These results of the basal-like gene expression signature give closer insight into the biology of the basal phenotype tumors within the ER-negative subgroup, and actually show that the influence of ER was not totally eliminated. ER status associated genes like XBP1 and TTF1 (Perou et al. 2000, Gruvberger et al. 2001, West et al. 2001) were found to be differentially expressed between basal-like and non-basal clusters, and these were among the top discriminatory genes. This suggests that there are probably differences in hormone-independence between basal-like and non-basal clusters within the ER-negative entity. The gene list included many other genes, which have been earlier found in the basal signature such as annexin A8, TRIM29, and αB-Crystallin (Perou et al. 2000, Sørlie et al. 2001, Sotiriou et al. 2003, Sørlie et al. 2003). From the therapeutic point of view it was noted that some of the top discriminatory genes were membrane associated (EVA1, SLC2A1, and CEACAM1), which might represent good targets for therapeutic interventions.

In addition to the above-mentioned biological characteristics, the gene list generated for the CK5/14-positive tumors was enriched in genes involved in the biological process of epidermal differentiation. The strong presence of epidermal differentiation in the gene expression profile of basal-like tumors might reflect the process of differentiation possibly evolving in these tumors. The tumor cells most probably reflect the differentiation potential of the cell of origin, which in this case might suggest an origin in a stem/progenitor cell of the mammary gland.

6. Prognostic and predictive implications of basal phenotype breast cancer (II,IV)

In this study, the basal phenotype was shown to associate with poor prognosis during the first years of follow-up in stage II tamoxifen treated primary breast cancers. This suggests that the basal phenotype would be an early relapse marker in breast cancer. Despite this, there was no indication

of a poor prognosis for basal phenotype tumors within the ER-negative subgroup if the basal phenotype was detected either with IHC or with gene expression microarrays. This is in agreement with the results of Potemski et al. (2005) and clearly indicates that whatever method is used for basal phenotype breast cancer identification, these tumors have a similar prognosis as the non-basal ER-negative breast cancers. These results are actually supported by the gene expression microarray studies, which do not show a significant difference between the prognosis of the two ER-negative tumor clusters, ERBB2 and basal-like (Sørlie et al. 2001, Sotiriou et al. 2003, Sørlie et al. 2003). However, Abd El-Rehim et al. (2004) and Rakha et al. (2006) have suggested that adjustment to steroid hormone receptor expression would not alter the adverse survival impact of the basal phenotype in breast cancer. The prognostic studies of basal phenotype breast cancer are presented in Table 3.

When high-risk breast cancer patients treated with intensive adjuvant chemotherapy were studied for survival, the results showed no difference between basal and non-basal phenotype breast cancers. The inability to detect a survival difference might be affected by the selection criteria of patients to have at least five metastatic axillary lymph nodes. Additionally, if the basal phenotype tumors are more sensitive to adjuvant chemotherapy treatment as suggested by some studies (Rodrígues-Pinilla et al. 2006a, Rouzier et al. 2006), the intensive treatment used in the original study (Bergh et al. 2000) might have obscured the difference between basal phenotype and non-basal breast cancer survival. Indirectly, this would support the suggestion that the highly proliferative basal phenotype breast cancers would be more sensitive to chemotherapy than non-basal tumors. One explanation to chemotherapy sensitivity in basal phenotype breast cancer might be their association with a low expression of Bcl-2, which has been linked with sensitivity to chemotherapy (Buchholz et al. 2003).

Table 3. Prognostic studies of basal phenotype breast cancer.

First Author	Year	No. of cases	Detection method	Univariate significance	Multivariate significance	Comments
Dairkee	1987b	118	IHC: CK14	yes	not studied	
Malzahn	1998	51	IHC: CK4,14,17	yes within N+	not studied	no adverse survival in the whole group or other subgroups (G1/G2, G3, ER+, ER-, N-)
Sørlie	2001	78	microarray	yes	not studied	the most adverse survival found in basal-like and ERBB2 subtypes
van de Rijn	2002	611	IHC: CK5,17	yes	yes within N-	adverse survival also in N- but not in the N+ tumors; no independence of size, N and grade; within N- independence of size, grade, HER-2, ER and GATA-3
Sotiriou	2003	99	microarray	yes	not studied	the most adverse survival found in basal-like and ERBB2 subtypes
Sørlie	2003	97	microarray	yes	not studied	the most adverse survival found in basal-like and ERBB2 subtypes
Abd El-Rehim	2004	1944	IHC: CK5/6 and CK14 separately	yes	yes	prognostic effect of CK5/6 independent of grade, node status and size
Foulkes	2004	247	IHC: CK5/6	yes	not studied	27 of the tumors were BRCA1 mutated
Jones	2004b	86 G III	IHC: CK14	no	not studied	no survival difference to other grade III breast cancer
Makretsov	2004	438	IHC and hierarchical clustering	yes	not studied	poor prognosis when compared to clusters resembling HER-2 and luminal subtypes, no univariate significance for CK5/6 alone
Nielsen	2004	930	IHC: CK5/6, 17	yes	not studied	significance found also in N+ but not in N- subgroup
Potemski	2005	195	IHC: CK5/6, 17	yes	no	significance found also in N- but not in the ER- or N+ subgroups, the poor prognosis was determined by ER and cyclin E status
Banerjee	2006	98	IHC: CK5/6, 14, 17	yes	not studied	adverse survival when compared to age-, node- and grade-matched IHC-negative tumors
Calza	2006	412	microarray	yes	not studied	adverse survival alike to luminal B and ERBB2 tumors
Carey	2006	496	IHC: ER/HER-2-, CK5/6+ and/or EGFR+	yes	not studied	adverse breast cancer-specific survival for basal-like and HER-2+/ER- subgroups
Fan	2006	295	microarray	yes	not studied	adverse survival according to recurrence score, wound response, and 70-gene profiles
Hu	2006	311	microarray	yes	yes	adverse survival for basal-like, ERBB2, and luminal B subtypes, independence of age, size, ER and node status and grade
Kim	2006	776	IHC: ER/HER-2-, CK5/14+ and/or EGFR+	no	no	no difference to HR+, HR+/HER-2+ or HR-/HER-2- tumors but better prognosis than for the HER-2+ tumors
Rakha	2006	1944	IHC: CK5/6, 14	yes	yes	in multivariate analysis independence of size, grade, node status and vascular invasion
Rodrígues-Pinilla	2006a	230 N-	IHC: ER/HER-2-, CK5/6+ and/or EGFR+	yes	not studied	adverse survival within N- tumors

G = histologic grade, ER = estrogen receptor, HR = hormone receptor, N = axillary lymph node status

In addition to the prognosis of basal phenotype breast cancer as a whole, the survival difference was studied between the basoluminal and basal subtypes. The independence of the new subtypes defined in this study was confirmed when it was found that basoluminal tumors show a worse prognosis than the basal subtype when studied in the high-risk breast cancer cohort. Since this was not caused by more frequent amplification of the HER-2 oncogene in basoluminal tumors, it indicates a variation in the natural aggressiveness of the basoluminal and basal subtypes. This leads to the conclusion that the separation of basoluminal and basal tumor subtypes would happen early in tumor development. Another possible cause for the survival difference between basoluminal and basal tumors might be the responsiveness to chemotherapy. This is less likely since the CK5/14 expression did not associate with relapse-free survival in patient groups randomized to receive CTCb or FEC treatments.

At the moment it seems that there are no clear benefits of the breast cancer classification to non-basal and basal since it does not influence the prognosis or treatment in comparison to other ER-negative tumors. As this study shows, basal phenotype tumors still differ biologically from other ER-negative breast cancers and additionally divide to two biologically and prognostically different entities. Additionally, there is growing evidence that basal phenotype tumors would be more sensitive to chemotherapy than non-basal tumors (Rodrígues-Pinilla et al. 2006a, Rouzier et al. 2006) and basal phenotype tumors carry many characteristics, which may be potential targets of drug development. It thus seems reasonable to believe that basal phenotype breast cancer and its subtypes will be important attributes in breast cancer classification in near future both in predictive and prognostic role.

7. dc-CISH in the assessment of HER-2 oncogene status (III)

In this study a CISH based method detecting two DNA probes simultaneously from a sample tissue was developed. The optimized dc-CISH method showed high concordance with the FDA-approved FISH method and turned out to be a practical and reliable application for HER-2 oncogene

amplification assessment. Both in FISH and in single-color CISH, normal copy numbers and high level amplifications are easy to detect and distinguish but the situation becomes problematic when 4-10 HER-2 gene copies are seen in tumor cell nuclei. These situations require the evaluation of the chromosome 17 copy number to be able to separate aneuploidy from low level amplification. In dc-CISH, this information is already present in the stained sample slide similarly as in FISH. Even if the sample selection overrepresented the borderline amplification cases, dc-CISH and FISH showed highly concordant results (91%) when the HER-2/chromosome 17 ratio of two was used as a limit for amplification (Kallioniemi et al. 1992). This indicates reliable separation of the signal colors in dc-CISH, the most critical step in the development of the method. The clear color separation was achieved by combining a red chromogen (New Fuchsin) with the clearly green signal of tetramethylbenzidine (TMB), which form a good contrast to each other and to the blue hematoxylin counter stain. The few discordant results between FISH and dc-CISH resulted in a ratio close to the amplification limit of two, reflecting the arbitrary nature of the cut-off (Kallioniemi et al. 1992). The limit for HER-2 amplification should be optimized and validated in the future to separate the trastuzumab responsive and unresponsive patients from each other.

The developed dc-CISH method combines the good properties of both FISH and CISH and is widely applicable to the detection of gene amplifications and deletions in paraffin-embedded tissue samples. Since in dc-CISH both the gene of interest and the control chromosome are detected simultaneously, the workload and the need for sample material are diminished. This further enables reliable interpretation of the gene copy number since the gene of interest and the control chromosome can be viewed simultaneously from exactly the same nucleus. dc-CISH also makes the interpretation and practice more convenient since normal brightfield microscopy familiar to pathologists can be used, a proper histopathologic view is achieved and the stained slides are preserved almost indefinitely (Tanner et al. 2000, Isola et al. 2004). Further, the laboratory procedure is similar to the single-color CISH method although it requires two staining steps. These

steps are easy to perform without extra workload by an automated immunostainer. The potential importance of the developed dc-CISH method was also acknowledged in the editorial associated with study IV (Shipley 2006).

The assessment of HER-2 oncogene status has become more important already at the time of primary breast cancer diagnosis since trastuzumab therapy has been indicated in an adjuvant setting (Piccard-Gebhart et al. 2005, Romond et al. 2005, Joensuu et al. 2006). When quick and accurate results are needed on a large scale, the detection method has to be simple and reliable. Immunohistochemistry is a simple method to separate trastuzumab responsive and non-responsive patients if correctly standardized, controlled, and evaluated (Ross et al. 2004). Since these qualifications are hard to achieve, there is need for confirmatory FISH tests in equivocal 2+ cases, which causes extra work for the laboratory (Gancberg et al. 2002, Bilous et al. 2003b, Ross et al. 2004). Since also immunohistochemically 3+ samples are sometimes non-amplified and occasional HER-2 amplifications are seen in 0/1+ tumors (Bartlett et al. 2001, Sauer et al. 2003, Ross et al. 2004, Joensuu et al. 2006, Saez et al. 2006), costly treatment might be misaddressed and patients might be exposed to ineffective treatments or left out of appropriate therapy. It has been suggested that HER-2 amplification status would predict the treatment response to trastuzumab better than the protein expression studied by IHC does (Mass et al. 2005). Conducting the HER-2 status assessment by an in situ hybridization method in the first place may be more convenient and cost effective (Elkin et al. 2006). Conducting all the HER-2 assays by FISH would require a great deal of time for evaluation by the pathology laboratory. The newly developed dc-CISH method provides a more convenient but still reliable approach to test gene copy number in paraffin-embedded tissue samples, which is a growing need in pathology laboratories.

SUMMARY AND CONCLUSIONS

The main conclusions from this study are:

1. Basal phenotype breast cancers detected by CK5/14 immunohistochemistry represent ~10% of sporadic and BRCA2 mutated hereditary breast cancers, but as much as 80% of hereditary BRCA1 mutated tumors show this phenotype. All sporadic basal phenotype tumors show co-expression of luminal cytokeratins. Distinctively, a small part of BRCA1 mutated basal phenotype tumors lack luminal cytokeratin expression. Basal phenotype tumors do not show myoepithelial differentiation.

2. Basal phenotype breast cancers are high grade hormone receptor negative breast cancers, which show high proliferation activity, vimentin and c-kit positivity, EGFR and p53 overexpression, and Bcl-2 negativity more often than other breast tumors. HER-2 oncogene amplified basal phenotype breast cancers exist, but an inverse association between these characteristics is seen within the ER-negative tumor subgroup.

3. Basal phenotype breast cancers show either uniform or heterogeneous basal cytokeratin expression. The two subtypes, basal and basoluminal, respectively, are distinct entities since they differ both prognostically and biologically. High proliferation activity and vimentin and c-kit immunopositivity are more often present in basal tumors while larger tumor size, HER-2 amplification, and worse survival estimates are found to be characteristic for basoluminal tumors. Similarly to basal phenotype tumors overall, both the basoluminal and basal subtypes seem to be stable in metastasis.

4. CK5/14-positive basal phenotype tumors have a unique gene expression profile within the ER-negative tumor subgroup and thus represent a biologically distinct entity, which differs from other breast cancers, most likely in the early phase of transformation.

5. The basal phenotype predicts an early breast cancer relapse. Despite this, there is no survival difference between basal phenotype breast cancers and other ER-negative tumors.

ACKNOWLEDGEMENTS

This study was performed in the Cancer Biology research group, Institute of Medical Technology, Tampere University and University Hospital and in the Department of Pathology, Seinäjoki Central Hospital during the years 2004-2007.

I am grateful to my supervisor Professor Jorma Isola, M.D., Ph.D., for the encouragement to start doctoral studies and for the opportunity to join his research group and finalize this way also my studies of medical cell biology. I would like to thank him also for all the questions answered, help given and especially for the possibility to work from Seinäjoki. During this time I have learned many new skills and got more interested about scientific work.

I would like to thank Seinäjoki Central Hospital and especially the head of the Pathology Department, Tapio Vartio, M.D., Ph.D., as well as Institute of Medical Technology and Tampere Graduate School in Biomedicine and Biotechnology for the opportunity to do research and for the great facilities for this work.

I express sincere thanks to the whole personnel of the Department of Pathology at the Seinäjoki Central Hospital for the very supporting and warm atmosphere. I'm grateful for all the help you have given me and for the patience you have showed. Antero Laasonen is especially thanked for endless enthusiasm in cell biology, excellent guidance to the techniques used in pathology laboratories and for the help during the whole process. Ritva Kujala, Helvi Salmela, Pirjo Pekkala and Päivi Kärki are acknowledges for their skillful laboratory work and especially for their help with the sample preparations. Thank you!

I would like to thank the Cancer Biology research group as a whole for creating a welcoming atmosphere when ever I arrived at Tampere and for all the support during better and worse times. Docent, Minna Tanner, M.D., Ph.D., and Päivikki Kauraniemi, Ph.D., are especially thanked for their scientific guidance during this process. Mrs. Sari Toivola and Mrs. Eeva Pesonen are gratefully acknowledged for their support and excellent work in the laboratory.

Docent Outi Monni, Ph.D., Helsinki University and docent Anita Naukkarinen, Ph.D., Kuopio University Hospital are gratefully acknowledged for reviewing this manuscript and giving good comments, which improved this work.

I would like to thank all the coauthors for their excellent work and co-operation.

Heidi, Johanna and Katri, our "Thesis Team of Seinäjoki", are thanked for the laughs, support and for the pizza days. It has been a pleasure to get to know you.

Finally I would like to thank my family for always believing in me even if they had only a clue what I was doing.

Aapo, you helped me in believing that I can do this. Thanks for the patience, support and the talks we had.

This study was financially supported by grants from the Pirkanmaa Hospital District Research Foundation, Medical Research Fund of Seinäjoki Central Hospital, Swedish Cancer Society, Swedish Research Council, Sigrid Juselius Foundation, Algol-Award, Oy Eli Lilly Finland Ab, and the Finnish Cancer Foundation.

Seinäjoki, January 2007

Mervi Jumppanen

REFERENCES

Abd El-Rehim DM, Pinder SE, Paish CE, Bell J, Blamey RW, Robertson JFR, Nicholson RI, and Ellis IO (2004): Expression of luminal and basal cytokeratins in human breast cancer. J Pathol 203:661-71.

Albiges L, Andre F, Balleyguier C, Gomez-Abuin G, Chompret A, and Delaloge S (2005): Spectrum of breast cancer metastasis in BRCA1 mutation carriers: highly increased incidence of brain metastasis. Ann Oncol 16:1846-1847.

Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, and Clarke MF (2003): Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci 100:3983-3988.

Al-Hajj M, Becker MW, Wicha M, Weissman I, and Clarke MF (2004): Therapeutic implications of cancer stem cells. Curr Opin Genet Dev 14:43-47.

Alvi AJ, Clayton H, Joshi C, Enver T, Ashworth A, Vivanco M, Dale TC, and Smalley MJ (2003): Functional and molecular characterization of mammary side population cells. Breast Cancer Res 5:R1-R8.

Ariga R, Zarif A, Korasick J, Reddy V, Siziopikou K, and Gattuso P (2005): Correlation of her-2/neu amplification with other prognostic and predictive factors in female breast carcinoma. Breast J 11:278-280.

Arnes JB, Brunet J, Stefansson I, Bégin LR, Wong N, Chappuis PO, and Foulkes WD (2005): Placental cadherin and the basal epithelial phenotype of BRCA1-related breast cancer. Clin Cancer Res 11:4003-4011.

Arnould L, Denoux Y, MacGrogan G, Penault-Llorca F, Fiche M, Treilleux I, Mathieu MC, Vincent-Salomon A, Vilain MO, and Couturier J (2003): Agreement between chromogenic in situ hybridization (CISH) ans FISH in the determination of HER2 status in breast cancer. Br J Cancer 88:1587-1591.

Balic M, Lin H, Young L, Hawes D, Giuliano A, McNamara G, Datar RH, and Cote RJ (2006): Most early disseminated cancer cells detected in bone marrow of breast cancer patients have a putative breast cancer stem cell phenotype. Clin Cancer Res 12:5615-5621.

Banerjee S, Reis-Filho JS, Ashley S, Steele D, Ashworth A, Lakhani SR, and Smith IE (2006). Basal-like breast carcinomas: clinical outcome and response to chemotherapy. J Clin Pathology 59:729-735.

Bánkfalvi A, Ludwig A, de-Hesselle B, Buerger H, Buchwalow IB, and Boecker W (2004): Different proliferative activity of the glandular and myoepithelial lineages in benign proliferative and early malignant breast diseases. Mod Pathol 17:1051-1061.

Barbareschi M, Pecciarini L, Cangi MG, Macrì E, Rizzo A, Viale G, and Doglioni C (2001): p63, a p53 homologue, is a selective nuclear marker of myoepithelial cells of the human breast. Am J Surg Pathol 25:1054-1060.

Bartlett JMS, Going JJ, Mallon EA, Watters AD, Reeves JR, Stanton P, Richmond J, Donald B, Ferrie R, and Cooke TG (2001): Evaluating HER2 amplification and overexpression in breast cancer. J Pathol 195:422-428.

Behbod F and Rosen JM (2004): Will cancer stem cells provide new therapeutic targets? Carcinogenesis 26:703-711.

Bergh J, Wiklund T, Erikstein B, Lidbrink E, Lindman H, Malmström P, Kellokumpu-Lehtinen P, Bengtsson NO, Söderlund G, Anker G, Wist E, Ottosson S, Salminen E, Ljungman P, Holte H, Nilsson J, Blomqvist C, and Wilkin N (2000): Tailored fluorouracil, epirubicin, and cyclophosphamide compared with marrow-supported high-dose chemotherapy as adjuvant treatment for high-risk breast cancer: a randomised trial. Scandinavian Breast Group 9401 study. Lancet 356:1384-1391.

Bhargava R, Lal P, and Chen B (2005): Chromogenic in situ hybridization for the detection of HER-2/neu gene amplification in breast cancer with an emphasis on tumors with borderline and low-level amplification: does it measure up to fluorescence in situ hybridization? Am J Clin Pathol 123:237-243.

Bilous M, Ades C, Armes J, Bishop J, Brown R, Cooke B, Cummings M, Farshid G, Field A, Morey A, McKenzie P, Raymond W, Robbins P, and Tan L (2003a): Predicting the HER2 status of breast cancer from basic histopathology data: an analysis of 1500 breast cancers as part of the HER2000 international study. The breast 12:92-98.

Bilous M, Dowsett M, Hanna W, Isola J, Lebeau A, Moreno A, Penault-Llorca F, Ruschoff J, Tomasic G, and van de Vijver M (2003b): Current perspectives on HER2 testing: A review of national testing guidelines. Mod Pathol 16:173-182.

Birnbaum D, Bertucci F, Ginestier C, Tagett R, Jacquemier J, and Charafe-Jauffret E (2004): Basal and luminal breast cancer: Basic or luminous? Int J Oncol 25:249-58.

Boecker W and Buerger H (2003): Evidence of progenitor cells of glandular and myoepithelial cell lineages in the human adult female breast epithelium: a new progenitor (adult stem) cell concept. Cell Prolif 36:73-84.

Boecker W, Moll R, Dervan P, Buerger H, Poremba C, Diallo RI, Herbst H, Schmidt A, Lerch MM, and Buchwalow IB (2002): Usual ductal hyperplasia of the breast is a committed stem (progenitor) cell lesion distinct from atypical ductal hyperplasia and ductal carsinoma in situ. J Pathol 198:458-467.

Bouton CM and Pevsner J (2000): DRAGON: Database referencing of array genes online. Bioinformatics 16:1038-1039.

Böcker W, Moll R, Poremba C, Holland R, Van Diest PJ, Dervan P, Bürger H, Wai D, Ina Diaollo R, Brandt B, Herbst H, Schmidt A, Lerch MM, and Buchwallow IB (2002): Common adult stem cells in the human breast give rise to glandular and myoepithelial cell lineages: A new cell biological concept. Lab Invest 82:737-745.

Brennan KR and Brown AMC (2004): Wnt proteins in mammary development and cancer. J Mammary Gland Biol Neoplasia 9:119-131.

Brewster A and Bondy M (2005): Clinical and molecular epidemiology of breast cancer. In: Molecular oncology of breast cancer, pp. 34-47. Eds. Ross JS and Hortobagyi GN, Jones and Bartlett Publishers, Sudbury, MA

Bryan BB, Schnitt SJ, and Collins LC (2006): Ductal carcinoma *in situ* with basal-like phenotype: a possible precursor to invasive basal-like breast cancer. Mod Pathol 19:617-621.

Buchholz TA, Davis DW, McConkey DJ, Symmans WF, Valero V, Jhingran A, Tucker SL, Pusztai L, Cristofanilli M, Esteva FJ, Hortobagyi GN, Sahin AA (2003): Chemotherapy-induced apoptosis and Bcl-2 levels correlate with breast cancer responce to chemotherapy. Cancer J 9:33-41.

Burkert J, Wright NA, and Alison MR (2006): Stem cells and cancer: an intimate relationship. J Pathol 209:287-297.

Calza S, Hall P, Auer G, Bjöhle J, Klaar S, Kronenwett U, Liu ET, Miller L, Ploner A, Smeds J, Bergh J, and Pawitan Y (2006): Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients. Breast Cancer Res 8:R34-R42.

Cardoso F, Durbecq V, Sotiriou C, and Ross JS (2005): HER-2/neu gen and protein in breast cancer, pp. 232-255. Eds. Ross JS and Hortobagyi GN, Jones and Bartlett Publishers, Sudbury, MA.

Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S, Deming SL, Geradts J, Cheang MC, Nielsen TO, Moorman PG, Earp HS, and Millikan RC (2006): Race, breast cancer subtypes, and survival in the Carolina breast cancer study. JAMA 295:2492-2502.

Carter P, Presta L, Gorman CM, Ridgway JBB, Henner D, Wong WLT, Rowland AM, Kotts C, Carver ME, and Shepard HM (1992): Humanization of an anti-p185^{HER2} antibody for human cancer therapy. Proc Natl Acad Sci 89:4285-4289.

Chang HY, Nuyten DS, Sneddon JB, Hastie T, Tibshirani R, Sørlie T, Dai H, He YD, van't Veer LJ, Bartelink H, van de Rijn M, Brown PO, and van de Vijver MJ (2005): Robustness, scalibility, and integration of a wound response gene expression signature in predicting breast cancer survival. Proc Natl Acad Sci USA 102:3738-3743.

Charafe-Jauffret E, Ginestier C, Monville F, Finetti P, Adélaïde J Cervera N, Fekairi S, Xerri L, Jacquemier J, Birnbaum D, and Bertucci F (2005): Gene expression profiling of breast cell lines idenfies potential new basal markers. Oncogene 25:2273-2284.

Chebil G, Bendahl P-O, Idvall I, and Fernö M (2003): Comparison of immunohistochemical and biochemical assay of steroid receptors in primary breast cancer. Acta Oncologica 42:719-725.

Chu PG and Weiss LM (2002): Keratin expression in human tissues and neoplasms. Histopathology 40:403-439.

Clarke CL, Sandle J, Parry SC, Reis-Filho JS, O'Hare MJ, and Lakhani SR (2004): Cytokeratin 5/6 in normal human breast: Lack of evidence for a stem cell phenotype. J Pathol 204:147-152.

Clarke C, Sandle J, and Lakhani (2005): Myoepithelial cells: Pathology, cell separation and markers of myoepithelial differentiation. J Mammary Gland Biol Neopl 10:273-280.

Clarke RB, Spence K, Anderson E, Howell A, Okano H, and Potten CS (2004): A putative human breast stem cell population is enriched for steroid receptor-positive cells. Dev Biol 277:443-456.

Clayton H, Titley I, and Vivanco MdM (2004): Growth and differentiation of progenitor/stem cells derived from the human mammary gland. Exp Cell Res 297:444-460.

Collett K, Stefansson IM, Eide J, Braaten A, Wang H, Eide GE, Thoresen S, Foulkes WD, and Akslen LA (2005): A basal epithelial phenotype is more frequent in interval breast cancers compared with screen detected tumors. Cancer Epidemiol Biomarkers Prev 14:1108-1112.

Dairkee SH, Blayney-Moore CM, Smith HS, and Hackett AJ (1986): Concurrent expression of basal and luminal epithelial markers in cultures of normal human breast analyzed using monoclonal antibodies. Differentiation 32:93-100.

Dairkee SH, Ljung BM, Smith H, and Hackett A (1987a): Immunolocalization of a human basal epithelium specific keratin in benign and malignant breast disease. Breast Cancer Res Treat 10:11-20.

Dairkee SH, Mayall BH, Smith HS, and Hackett AJ (1987b): Monoclonal marker that predicts early recurrence of breast cancer. Lancet 1:514.

Deugnier M, Teulière J, Faraldo MM, Thiery JP, and Glukhova MA (2002): The importance of being a myoepithelial cell. Breast Cancer Res 4:224-230.

Deugnier M-A, Faraldo MM, Teulière J, Thiery JP, Medina D, and Glukhova MA (2006): Isolation of mouse mammary epithelial progenitor cells with basal characteristics from the Comma-Dβ cell line. Dev Biol 293:414-425.

Dickson RB (1996): Biochemical control of breast development. In: Diseases of the Breast, pp. 15-25. Eds. Harris JS, Lippmann ME, Morrow M, and Hellman S, Lippincott-Raven Publishers, Philadelphia.

Diehn M, Eisen MB, Botstein D, and Brown PO (2000): Large-scale identification of secreted and membrane associated gene products using DNA microarrays. Nature genetics 25:58-62.

Dontu G, Abdallah WM, Foley JM, Jackson KW, Clarke MF, Kawamura MJ, and Wicha MS (2003): In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes & Development 17:1253-1270.

Dontu G, Jackson KW, McNicholas E, Kawamura MJ, Abdallah WM, and Wicha MS (2004): Role of Notch signaling in cell-fate determination of human mammary stem/progenitor cells. Breast Cancer Res 6:R605-R615.

Eisen MB, Spellman PT, Brown PO, and Botstein D (1998): Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 351:1451-1467.

Elkin EB, Weinstein MC, Winer EP, Kuntz KM, Schnitt SJ, and Weeks JC (2006): HER-2 testing and trastuzumab therapy for metastatic breast cancer: A cost-effectiveness analysis. J Clin Oncol 22:854-863.

Elston CW and Ellis IO (1998): Normal structure and developmental abnormalities. In: The breast; systemic pathology, pp. 1-19. Eds. Elston CW and Ellis IO, Churchill Livingstone, Edinburgh, UK.

Estellar M, Silva JM, Dominguez G, Bonilla F, Matias-Guiu X, Lerma E, Bussaglia E, Prat J, Harkes IC, Repasky EA, Gabrielson E, Schutte M, Baylin SB, and Herman JG (2000): Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. J Natl Cancer Inst 92:564-569.

Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DSA, Nobel AB, van't Veer LJ, and Perou CM (2006): Concordance among gene-expression-based predictors for breast cancer. N Engl J Med 355:560-569.

Fitzgibbons PL, Page DL, Weaver D, Thor AD, Allred DC, Clark GM, Ruby SG, O´Malley F, Simpson JF, Connolly JL, Hayes DF, Edge SB, Lichter A, and Schnitt SJ (2000): Prognostic factors in breast cancer; College of American pathologists consensus statement 1999. Arch Pathol Lab Med 124:966-978.

Foulkes WD (2003a): BRCA1 functions as a breast stem cell regulator. J Med Genet 41:1-5.

Foulkes WD, Stefansson IM, Chappuis PO, Bégin LR, Goffin JR, Wong N, Trudel M, and Akslen LA (2003b): Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. J Natl Cancer Inst 95:1482-1485.

Foulkes WD, Brunet J-S, Stefansson IM, Straume O, Chappuis PO, Bégin LR, Hamel N, Goffin JR, Wong N, Trudel M, Kapusta L, Porter P, and Akslen LA (2004): The prognostic implication of the basal-like (Cyclin $E^{high}/p27^{low}/p53^{+}$/glomeruloid-microvascular-proliferation^{+}) phenotype of BRCA1-related breast cancer. Cancer Res 64:830-835.

Fulford LG, Easton DF, Reis-Filho JS, Sofronis A, Gillett CE, Lakhani SR, and Hanby A (2006): Specific morphological features in grade 3 invasive ductal carcinoma of breast. Histopathology 49:22-34.

Furuta S, Jiang X, Gu B, Cheng E, Chen P, and Lee W (2005): Depletion of BRCA1 impairs differentiation but enhances proliferation of mammary epithelial cells. Proc Natl Acad Sci USA 102:9176-9181.

Gancberg D, Järvinen T di Leo A, Rouas G, Cardoso F, Paesmans M, Verhest A, Piccard MJ, Isola J, and Larsimont D (2002): Evaluation of HER-2/NEU protein expression in breast cancer by immunohistochemistry: an interlaboratory study assessing the reproducibility of HER-2/NEU testing. Breast Cancer Res Treat 74:113-120.

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.

Gong Y, Booser DJ, and Sneige N (2005a): Comparison of HER-2 status determined by fluorescence in situ hybridization in primary and metastatic breast carcinoma. Cancer 103:1763-1769.

Gong Y, Gilcrease M, and Sneige N (2005b): Reliability of chromogenic in situ hybridization for detecting HER-2 gene status in breast cancer: comparison with fluorescence in situ hybridization and assessment of interobserver reproducibility. Mod Pathol 18:1015-1021.

Gonzalez-Angulo AM, Hortobágyi GN, and Estva FJ (2006): Adjuvant therapy with trastuzumab for HER-2/neu-positive breast cancer. The Oncologist 11:857-867.

Gordon LA, Mulligan KT, Maxwell-Jones H, Adams M, Walker RA, and Jones L (2003): Breast cell invasive potential relates to the myoepithelial phenotype. Int J Cancer 106:8-16.

Gould VE, Koukoulis GK, Jansson DS, Nagle RB, Franke WW, and Moll R (1990): Coexpression patterns of vimentin and glial filament protein with cytokeratins in the normal, hyperplastic and neoplastic breast. Am J Pathol 137:1143-1155.

Gruvberger S, Ringnér M, Chen Y, Panavally S, Saal LH, Borg Å, Fernö M, Peterson C, and Meltzer PS (2001): Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression pattern. Cancer Res 61:5979-5984.

Gruvberger SK, Ringnér M, Edén P, Borg Å, Fernö M, Peterson C, and Meltzer PS (2002): Expression profiling to predict outcome in breast cancer: the influence of sample selection. Breast Cancer Res 5:23-26.

Gudjonsson T, Rønnov-Jessen L, Villadsen R, Rank F, Bissell MJ, and Petersen OW (2002): Normal and tumor-derived myoepithelial cells differ in their ability to interact with luminal breast epithelial cells for polarity and basement membrane deposition. J Cell Sci 115:39-50.

Gupta D, Middleton LP, Whitaker MJ, and Abrams J (2003): Comparison of Fluorescence and Chromogenic In Situ Hybridization for Detection of HER-2/neu Oncogene in Breast Cancer. Am J Clin Pathol 119:381-387.

Gusterson BA, Ross DT, Heath VJ, and Stein T (2005): Basal cytokeratins and their relationship to the cellular origin and functional classification of breast cancer. Breast Cancer Res 7:143-148.

Hanna WM and Kwok K (2006): Chromogenic in-situ hybridization: a viable alternative to fluorescence in-situ hybridization in the HER2 testing algorithm. Mod Pathol 19:481-487.

Hauser-Kronberger C and Dandachi N (2004): Comparison of chromogenic in situ hybridization with other methodologies for HER2 status assessment in breast cancer. J Mol Histol 35:647-653.

Heatley M, Maxwell P, Whiteside C, and Toner P (1995): Cytokeratin intermediate filament expression in benign and malignant breast disease. J Clin Pathol 48:26-32.

Hicks DG, Short SM, Prescott NL, Tarr SM, Coleman KA, Yoder BJ, Crowe JP, Choueiri TK, Dawson AE, Budd T, Tubbs RR, Casey G, and Weil RJ (2006): Breast cancers with brain metastases are more likely to be estrogen receptor negative, express the basal cytokeratin CK5/6, and overexpress HER2 or EGFR. Am J Surg Pathol 30:1097-1104.

Honrado E, Benítez J, and Palacios J (2006): Histopathology of BRCA1- and BRCA2-associated breast cancer. Crit Rev Oncol Hematol 59:27-39.

Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, Livasy C, Carey LA, Reynolds E, Dressler L, Nobel A, Parker J, Ewend MG, Sawyer LR, Wu J, Liu Y, Nanda R, Tretiakova M, Ruiz Orrico A, Dreher D, Palazzo JP, Perreard L, Nelson E, Mone M, Hansen H, Mullins M, Quackenbush JF, Ellis MJ, Olopade OI, Bernard PS, and Perou CM (2006): The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 7:96-107.

Huang HJ, Neven P, Drijkoningen M, Paridaens R, Wildiers H, Van Limbergen E, Berteloot P, Amant F, Vergote I, and Christiaens MR (2005): Association between tumour characteristics and HER-2/neu by immunohistochemistry in 1362 women with primary operable breast cancer. J Clin Pathol 58:611-616.

Hungermann D, Buerger H, Oehlschlegel C, Herbst H, and Boecker W (2005): Adenomyoepithelial tumors and myoepithelial carcinomas of the breast - a spectrum of monophasic and biphasic tumors dominated by immature myoepithelial cells. BMC Cancer 5:92-100.

Hyytinen E, Visakorpi T, Kallioniemi A, Kallioniemi OP, and Isola JJ (1994): Improved technique for analysis of formalin-fixed, paraffin-embedded tumors by fluorescence in situ hybridization. Cytometry 16:93-99.

Isola J, Visakorpi T, Holli K, and Kallioniemi OP (1992): Association of overexpression of tumor suppressor protein p53 with rapid cell proliferation and poor prognosis in node-negative breast cancer patients. J Natl Cancer Inst 84:1109-1114.

Isola J, Tanner M, Forsyth A, Cooke TG, Watters AD, and Bartlett JM (2004): Interlaboratory comparison of HER-2 oncogene amplification as detected by chromogenic and fluorescence in situ hybridization. Clin Cancer Res 10:4793-4798.

Jackisch C (2006): HER-2-positive metastatic breast cancer: Optimizing trastuzumab-based therapy. The Oncologist 11(suppl 1):34-41.

Jacquemier J, Padovani L, Rabayrol L, Lakhani SR, Penault-Llorca F, Denoux Y, Fiche M, Figueiro P, Maisongrosse V, Ledoussal V, Martinez Penuela J, Udvarhely N, El Makdissi G, Ginestier C, Geneix J, Charafe-Jauffret E, Xerri L, Eisinger F, Birnbaum D, and Sobol H (2005): Typical medullary breast carcinomas have a basal/myoepithelial phenotype. J Pathol 207:260-268.

Joensuu H, Kellokumpu-Lehtinen P-L, Bono P, Alanko T, Kataja V, Asola R, Utriainen T, Kokko R, Hemminki A, Tarkkanen M, Turpeenniemi-Hujanen T, Jyrkkiö S, Flander M, Helle L, Ingalsuo S, Johansson K, Jääskeläinen A-S, Pajunen M, Rauhala M, Kaleva-Kerola J, Salminen T, Leinonen M, Elomaa I, and Isola J (2006): Adjuvant Docetaxel and virorelbine with or without Trastuzumab for breast cancer. N Engl J Med 354:809-820.

Jolicoeur F (2005): INtrauterine breast development and the mammary myoepithelial lineage. J Mammary Gland Biol Neoplasia 10:199-210.

Jones C, Nonni AV, Fulford L, Merrett S, Chaggar R, Eusebi V, and Lakhani SR (2001): CGH analysis of ductal carcinoma of the breast with basaloid/myoepithelial cell differentiation. British J of Cancer 85:422-427.

Jones C, Mackay A, Grigoriadis A, Cossu A, Reis-Filho JS, Fulford L, Dexter T, Davies S, Bulmer K, Ford E, Parry S, Budroni M, Palmieri G, Neville AM, O´Hare MJ, and Lakhani SR (2004a): Expression profiling of purified normal human luminal and myoepithelial breast cells: Identification of novel prognostic markers for breast cancer. Cancer Res 64:3037-3045.

Jones C, Ford E, Gillett C, Ryder K, Merrett S, Reis-Filho JS, Fulford LG, Hanby A, and Lakhani SR (2004b): Molecular cytogenetic identification of subgroups of grade III invasive ductal breast carcinomas with different clinical outcomes. Clin Cancer Res 10:5988-5997.

Joshi K, Smith JA, Perusinghe N, and Monoghan P (1986): Cell proliferation in the human mammary epithelium. Differential contribution by epithelial and myoepithelial cells. Am J Pathol 124:199-206.

Järvelä S, Helin H , Haapasalo J, Järvelä T, Junttila TT, Elenius K, Tanner M, Haapasalo H, and Isola J (2006): Amplification of the epidermal growth factor receptor in astrocytic tumors by chromogenic in situ hybridization: Association with clinicopathological features and patient survival. Neuropathol Appl Neurobiol 32:441-450.

Kalirai H and Clarke RB (2006): Human breast epithelial stem cells and their regulation. J Pathol 208:7-16.

Kallioniemi O-P, Kallioniemi A, Kurisu W, Thor A, Chen L-C, Smith HS, Waldman FM, Pinkel D, and Gray JW (1992): ERBB2 amplification in breast cancer analyzed by fluorescence in situ hybridization. Proc Natl Acad Sci USA 89:5321-5325.

Khan J, Saal LH, Bittner ML, Jiang Y, Gooden GC, Glatfelter AA, and Meltzer PS (2002): Gene expression profiling in cancer using cDNA microarrays. Methods Mol Med 68:205-222.

Kiechle M and Meindl A (2005): Predisposition for breast cancer. In: Molecular oncology of breast cancer pp. 48-59, Eds. Ross JS and Hortobagyi GN, Jones and Bartlett Publishers, Sudbury, MA.

Kim MK, Ro JY, Ahn S, Kim HH, Kim S, and Gong G (2006): Clinicopathological significance of the basal-like subtype of breast cancer: a comparison with hormone receptor and Her2/neu-overexpressing phenotypes. Hum Pathol 37:1217-1226.

Kordon EC and Smith GH (1998): An entire mammary gland may comprise the progeny from a single cell. Development 125:1921-1930.

Korsching E, Packeisen J, Agelopoulos K, Eisenacher M, Voss R, Isola J, van Diest PJ, Brandt B, Boecker W, and Buerger H (2002): Cytogenetic alterations and cytokeratin expression patterns in breast cancer: Integrating a new model of breast differentiation into cytogenetic pathways of breast carcinogenesis. Lab Invest 82:1525-1533.

Korsching E, Packeisen J, Liedtke C, Hungermann D, Wulfing P, van Diest PJ, Brandt B, Boecker W, and Buerger H (2005): The origin of vimentin expression in invasive breast cancer: epithelial-mesenchymal transition, myoepithelial histogenesis or histogenesis from progenitor cells with bilinear differentiation potential? J Pathol 206:451-457.

Lakhani SR and O'Hare MJ (2001): The mammary myoepithelial cell - Cinderella or ugly sister? Breast Cancer Res 3:1-4.

Lakhani SR, Reis-Filho JS, Fulford L, Penault-Llorca F, van der Vijver M, Parry S, Bishop T, Benitez J, Rivas C, Bignon YJ, Chang-Claude J, Hamann U, Cornelisse CJ, Devilee P, Beckmann MW, Nestle-Kramling C, Daly PA, Haites N, Varley J, Lalloo F, Evans G, Maugard C, Meijers-Heijboer H, Klijn JG, Olah E, Gusterson BA, Pilotti S, Radice P, Scherneck S, Sobol H, Jacquemier J, Wagner T, Peto J, Stratton MR, McGuffog L, and Easton DF (2005): Prediction of BRCA1 status in patients with breast cancer using estrogen receptor and basal phenotype. Clin Cancer Res 11:5175-5180.

Li Y, Welm B, Podsypanina K, Huang S, Chamorro M, Zhang X, Rowlands T, Egeblad M, Cowin P, Werb Z, Tan LK, Rosen JM, and Varmus HE (2003): Evidence that transgenes encoding components of the Wnt signaling pathway preferentially induce mammary cancers from progenitor cells. Proc Natl Acad Sci USA 100:15853-15858.

Lin F, Shen T, and Prichard JW (2005): Detection of Her-2/neu Oncogenein Breast Carcinoma by Chromogenic In Situ Hybridization in Cytologic Specimens. Diagn Cytopathol 33:376-380.

Liu S, Dontu G, and Wicha MS (2005): Mammary stem cells, self-renewal pathways, and carcinogenesis. Breast Cancer Res 7:86-95.

Liu S, Dontu G, Mantle ID, Patel S, Ahn N, Jackson KW, Suri P, and Wicha MS (2006): Hedgehog signaling and Bmi-1 regulate self-renewal of normal and malignant human mammary stem cells. Cancer Res 66:6063-6071.

Livasy CA, Karaca G, Nanda R, Tretiakova MS, Olopade OI, Moore DT, and Perou CM (2006): Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. Mod Pathol 19:264-271.

Loman N, Johannsson O, Kristoffersson U, Olsson H, and Borg Å (2001): Family history of breast and ovarian cancers and BRCA1 and BRCA2 mutations in a population-based series of early-onset breast cancer. J Natl Cancer Inst 93:1215-1223.

Loring P, Cumming R, O'Grady A, and Kay EW (2005): HER2 positivity in breast carcinoma: A comparison of chromogenic in situ hybridization with fluorescence in situ hybridization in tissue microarrays, with targeted evaluation of intratumoral heterogeneity by in situ hybridization. Appl Immunohistochem Mol Morphol 13:194-200.

Lundin M, Lundin J, Helin H, and Isola J (2004): A digital atlas of breast histopathology: an application of web based virtual microscopy. J Clin Pathol 57:1288-1291.

Ma X, Wang Z, Ryan PD, Isakoff SJ, Barmettler A, Fuller A, Muir B, Mohapatra G, Salunga R, Tuggle JT, Tran Y, Tran D, Tassin A, Amon P, Wang W, Wang W, Enright E, Stecker K, Estepa-Sabal E, Smith B, Younger J, Balis U, Michaelson J, Bhan A, Habin K, Baer TM, Brugge J, Haber DA, Erlander MG, and Sgroi DC (2004): A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. Cancer Cell 5:607-616.

Makretsov NA, Huntsman DG, Nielsen TO, Yorida E, Peacock M, Cheang MC, Dunn SE, Hayes M, van de Rijn M, Bajdik C, and Gilks CB (2004): Hierarchical clustering analysis of tissue microarray immunostaining data identifies prognostically significant groups of breast carcinoma. Clin Cancer Res 10:6143-6151.

Malzahn K, Mitze M, Thoenes M, and Moll R (1998): Biological and prognostic significance of stratified epithelial cytokeratins in infiltrating ductal breast carcinomas. Virchows Arch 433:119-129.

Mass RD, Press M, Anderson S, Cobleigh MA, Vogel CL, Dybdal N, Leiberman G, and Slamon DJ (2005): Evaluation of clinical outcomes according to HER2 detection by fluorescence in situ hybridization in women with metastatic breast cancer treated with trastuzumab. Clin Breast Cancer 6:240-246.

Matos I, Dufloth R, Alvarenga M, Zeferino LC, and Schmitt F (2005): p63, cytokeratin 5, and P-cadherin: three molecular markers to distinguish basal phenotype in breast carcinomas. Virchows Arch 447:688-694.

Matros E, Wang ZC, Lodeiro G, Miron A, Iglehart JD, and Richardson AL (2005): BRCA1 promoter methylation in sporadic breast tumors: relationship to gene expression profiles. Breast Cancer Res Treat 91:179-186.

McCormick SR, Lillemoe TJ, Beneke J, Schrauth J, and Reinartz J (2002): HER2 assessment by immunohistochemical analysis and fluorescence in situ hybridization: comparison of HercepTest and PathVysion commercial assays. Am J Clin Pathol 117:935-943.

Miettinen M and Lasota J (2005): KIT (CD117): A review on expression in normal and neoplastic tissues, and mutations and their clinicopathologic correlation. Appl Immunohistochem Mol Morphol 13:205-220.

Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, Viale A, Olshen AB, Gerald WL, and Massagué J (2005): Genes that mediate breast cancer metastasis to lung. Nature 436:518-524.

Moll R (1998): Cytokeratins as markers of differentiation in the diagnosis of epithelial tumors. Subcell Biochem 31:205-262. Eds. Herrmann and Harris, Plenum Press, New York.

Moll R, Franke WW, and Schiller DL (1982): The catalog of human cytokeratins: Patterns of expression in normal epithelia, tumors and cultured cells. Cell 31:11-24.

Moll R, Krepler R, and Franke WW (1983): Complex cytokeratin polypeptide patterns observed in certain human carcinomas. Differentiation 23:256-269.

Moyano JV, Evans JR, Chen F, Lu M, Werner ME, Yehiely F, Diaz LK, Turbin D, Karaca G, Wiley E, Nielsen TO, Perou CM, and Cryns VL (2006): αB-Crystallin is a novel

oncoprotein that predicts poor clinical outcome in breast cancer. J Clin Invest 116:261-270.

Nagle RB, Böcker W, Davis JR, Heid HW, Kaufmann M, Lucas DO, and Jarasch E (1986): Characterization of breast carcinomas by two monoclonal antibodies distinguishing myoepithelial from luminal epithelial cells. J Histochem Cytochem 34:869-881.

Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, Hernandez-Boussard T, Livasy C, Cowan D, Dressler L, Akslen LA, Ragaz J, Gown AM, Gilks CB, van de Rijn M, and Perou CM (2004): Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res 10:5367-5374.

Nylander K, Vojtesek B, Nenutil R, Lindgren B, Roos G, Zhanxiang W, Sjöström B, Dahlqvist A, and Coates PJ (2002): Differential expression of p63 isoforms in normal tissues and neoplastic cells. J Pathol 198:417-427.

Omary MB, Ku N-O, Liao J, and Price D (1998): Keratin modification and solubility properties in epithelial cells and in vitro. In Subcellular Biochemistry 31:105-133. Eds. Herrmann and Harris, Plenum Press, New York.

Osborne MP (1996): Breast development and anatomy. In: Diseases of the Breast, pp. 1-14. Eds. Harris JS, Lippmann ME, Morrow M and Hellman S, Lippincott-Raven Publishers, Philadelphia.

Otterbach F, Bànkfalvi À, Berhner S, Decker T, Krech R, and Boecker W (2000): Cytokeratin 5/6 immunohistochemistry assists the differential diagnosis of atypical proliferations of the breast. Histopathology 37:232-240.

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

Palacios J, Benito N, Pizarro A, Suarez A, Espada J, Cano A, and Gamallo C (1995): Anomalous expression of P-cadherin in breast carcinomas. Correlation with E-cadherin expression and pathological features. Am J Pathol 146:605-612.

Palacios J, Honrado E, Osorio A, Cazorla A, Sarrió D, Barroso A, Rodrígues S, Cigudosa JC, Diez O, Alonso C, Lerma E, Dopazo J, Rivas C, and Benítez J (2005): Phenotypic characterization of BRCA1 and BRCA2 tumors based in a tissue microarray study with 37 immunohistochemical markers. Breast Cancer Res Treat 90:5-14.

Park K, Kim J, Lim S, Han S, and Lee JY (2003): Comparing fluorescence in situ hybridization and chromogenic in situ hybridization methods to determine the HER2/neu status in primary breast carcinoma using tissue microarray. Mod Pathol 16:937-943.

Parmar H and Cunha GR (2004): Epithelial-stromal interactions in the mouse and human mammary gland in vivo. Endocrine-Related Cancer 11:437-458.

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

Perou CM, Sørlie 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.

Piccart-Gebhart MJ, Procter M, Leyland-Jones B, Goldhirsch A, Untch M, Smith I, Gianni L, Baselga J, Bell R, Jackisch C, Cameron D, Dowsett M, Barrios CH, Steger G, Huang CS, Andersson M, Inbar M, Lichinitser M, Lang I, Nitz U, Iwata H, Thomssen C, Lohrisch C, Suter TM, Ruschoff J, Suto T, Greatorex V, Ward C, Straehle C, McFadden E, Dolci MS, and Gelber RD; Herceptin Adjuvant (HERA) Trial Study

Team (2005): Trastuzumab after adjuvant chemotherapy in HER-2-positive breast cancer. N Engl J Med 353:1659-1672.

Pinkel D, Straume T, and Gray JW (1986): Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. Proc Natl Acad Sci USA 83:2934-2938.

Polyak K and Hu M (2005): Do myoepithelial cells hold the key for breast tumor progression? J Mammary Gland Biol Neopl 10:231-247.

Polychronis A, Sinnett HD, Hadjiminas D, Singhal H, Mansi JL, Shivapatham D, Shousha S, Jiang J, Peston D, Barrett N, Vigushin D, Morrison K, Beresford E, Ali S, Slade MJ, Coombes RC (2005): Preoperative gefitinib versus gefitinib and anastrozole in postmenopausal patients with oestrogen-receptor positive and epidermal-growth-factor-receptor-positive primary breast cancer: a double-blind placebo-controlled phase II randomised trial. Lancet Oncol 6:383-391.

Ponti D, Costa A, Zaffaroni N, Pratesi G, Petrangolini G, Coradini D, PIlotti S, Pierotti MA, and Daidone MG (2005): In vitro propagation of breast tumorigenic cancer cells with stem/progenitor cell properties. Cancer Res 65:5506-5511.

Ponti D, Zaffaroni N, Capelli C, and Daidone MG (2006): Breast cancer stem cells: An overview. Eur J Cancer 42:1219-1224.

Potemski P, Kusinska R, Watala C, Pluciennik E, Bednarek AK, and Kordek R (2005): Prognostic relevance of basal cytokeratin expression in operable breast cancer. Oncology 69:478-485.

Pukkala E, Sankila R, and Rautalahti M (2003): Syöpä Suomessa 2003. Suomen syöpäyhdistyksen julkaisuja nro 64, Painotalo Miktor Oy, Helsinki, Finland.

Pusztai L, Ayers M, Stec J, Clark E, Hess K, Stivers D, Damokosh A, Sneige N, Buchholz TA, Esteva FJ, Arun B, Cristofanilli M, Booser D, Rosales M, Valero V, Adams C, Hortobagyi GN, and Symmans WF (2003): Gene expression profiles obtained from fine-needle aspiration of breast cancer reliably identify routine prognostic markers and reveal large-scale molecular differences between estrogen-negative and estrogen-positive tumors. Clin Cancer Res 9:2406-2415.

Putti TC, Abd El-Rehim DM, Rakha EA, Paish CE, Lee AHS, Pinder SE, and Ellis IO (2005): Estrogen receptor-negative breast carcinomas: a review of morphology and immunophenotypical analysis. Mod Pathol 18:26-35.

Rakha EA, Putti TC, Abd El-Rehim DM, Paish C, Green AR, Powe DG, Lee AH, Robertson JF, and Ellis IO (2006): Morphological and immunophenotypic analysis of breast carcinomas with basal and myoepithelial differentiation. J Pathol 208:495-506.

Reis-Filho JS, Simpson PT, Martins A, Preto A, Gaertner F, and Schmitt FC (2003): Distribution of p63, cytokeratin 5/6 and cytokeratin 14 in 51 normal and 400 neoplastic human tissue samples using TARP-4 multi-tumor tissue microarray. Virchows Arch 443:122-132.

Reis-Filho JS, Milanezi F, Steele D, Savage K, Simpson PT, Nesland JM, Pereira EM, Lakhani SR, and Schmitt FC (2006a): Metaplastic breast carcinomas are basal-like tumors. Histopathology 49:10-21.

Reis-Filho JS, Savage K, Lambros MBK, James M, Steele D, Jones RL, and Dowsett M (2006b): Cyclin D1 protein overexpression and CCND1 amplification in breast carcinomas: an immunohistochemical and chromogenic in situ hybridization analysis. Mod Pathol 19:999-1009.

Reis-Filho JS, Steele D, Di Palma S, Jones RL, Savage K, James M, Milanezi F, Schmitt FC, and Ashworth A (2006c): Distribution and significance of nerve growth factor receptor (NGFR/p75^{NTR}) in normal, bening and malignant breast tissue. Mod Pathol 19:307-319.

Reya T, Morrison SJ, Clarke MF, and Weissman IL (2001): Stem cells, cancer, and cancer stem cells. Nature 414:105-111.

Ribeiro-Silva A, Zambelli Ramalho LN, Britto Garcia S, and Zucoloto S (2003): The relationship between p63 and p53 expression in normal and neoplastic breast tissue. Arch Pathol Lab Med 127:336-340.

Ribeiro-Silva A, Ramalho LNZ, Garcia SB, Brandão DF, Chahud F, and Zucoloto S (2005): p63 correlates with both BRCA1 and cytokeratin in invasive breast carcinomas: further evidence for the pathogenesis of the basal phenotype of breast cancer. Histopathology 47:458-466.

Ribeiro-Silva A, Ribeiro do Vale F, and Zucoloto S (2006): Vascular endothelial growth factor expression in the basal subtype of breast carcinoma. Am J Clin Pathol 125:512-518.

Rice JC, Ozcelik H, Maxeiner P, Andrulis I, and Futscher BW (2000): Methylation of the BRCA1 promoter is associated with decreased BRCA1 mRNA levels in clinical breast cancer specimens. Carcinogenesis 21:1761-1765.

Richardson AL, Wang ZC, De Nicolo A, Lu X, Brown M, Miron A, Liao X, Iglehart JD, Livingston DM, and Ganesan S (2006): X chromosomal abnormalities in basal-like human breast cancer. Cancer Cell 9:121-132.

Rodríguez-Pinilla SM, Sarrío D, Honrado E, Hardisson D, Calero F, Benítez J, and Palacios J (2006a): Prognostic significance of basal-like phenotype and fascin expression in node-negative invasive breast carcinomas. Clin Cancer Res 12:1533-1539.

Rodríguez-Pinilla SM, Honrado E, Hardisson D, Benítez J, and Palacios J (2006b): Caveolin-1 expression is associated with a basal-like phenotype in sporadic and hereditary breast cancer. Breast Cancer Res Treat 99:85-90.

Romond EH, Perez EA, Bryant J, Suman VJ, Geyer CE Jr, Davidson NE, Tan-Chiu E, Martino S, Paik S, Kaufman PA, Swain SM, Pisansky TM, Fegrenbacher L, Kutter LA, Vogel VG, Visscher DW, Yother G, Jenkins RB, Brown AM, Dakhil SR, Mamounas EP, Lingle WL, Klein PM, Ingle JN, and Wolmark N (2005): Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. N Engl J Med 353:1673-1684.

Ross JS and Harbeck N (2005): Prognostic and Predictive factors overview. In: Molecular oncology of breast cancer, pp. 128-141. Eds. Ross JS and Hortobagyi GN, Jones and Bartlett Publishers, Sudbury, MA.

Ross JS, Fletcher JA, Bloom KJ, Linette GP, Stec J, Clark E, Ayers M, Symmans WF, Pusztai L, and Hortobagyi GN (2003): HER-2/neu testing in breast cancer. Am J Clin Pathol 120:S53-S71.

Ross JS, Fletcher JA, Bloom KJ, Linette GP, Stec J, Symmans WF, Pusztai L, and Hortobagyi GN (2004): Targeted therapy in breast cancer; The HER-2/neu gene and protein. Mol Cell Proteomics 3:379-398.

Rouzier R, Perou CM, Symmans WF, Ibrahim N, Cristofanilli M, Anderson K, Hess KR, Stec J, Ayers M, Wagner P, Morandi P, Fan C, Rabiul I, Ross JS, Hortobagyi GN, and Pusztai L (2005): Breast cancer molecular subtypes respond differently to preoperative chemotherapy. Clin Cancer Res 11:5678-5685.

Rønnov-Jessen L, Lundstrom GP, Keene DR, and Burgeson RE (1996): Cellular changes involved in conversion of normal to malignant breast: the importance of the stromal reaction. Physiol Rev 76:69-125.

Rubin I and Yarden Y (2001): The basic biology of HER2. Ann Oncol 12:S3-S8.

Saal LH, Troein C, Vallon-Christerson J, Gruvberger S, Borg A, and Peterson C (2002): BioArray Software Environment (BASE): a platform for comprehensive management and analysis of microarray data. Genome Biol 3:SOFTWARE0003.

Saez A, Andreu FJ, Segui MA, Bare ML, Fernandez S, Dinares C, and Rey M (2006): HER-2 gene amplification by chromogenic in situ hybridization (CISH) compared with

fluorescence in situ hybridization (FISH) in breast cancer - A study of two hundred cases. Breast 15:519-527.

Sauer T, Wiedswang G, Boudjema G, Christensen H, and Karesen R (2003): Assessment of HER-2/neu overexpression and/or gene amplification in breast carcinomas: should in situ hybridization be the method of choice? APMIS 111:444-450.

Sheridan C, Kishimoto H, Fuchs RK, Mehrotra S, Bhat-Nakshatri P, Turner CH, Goulet R, Badve S, and Nakshatri H (2006): CD44+/CD24- breast cancer cells exhibit enhanced invasive properties, an early step necessary for metastasis. Breast Cancer Res 8:R59 [Epub ahead of print].

Shipley J (2006): Puttin colours into chromogenic in situ hybridization (CISH). J Pathol 210:1-2.

Sidoni A, Ferri I, Cavaliere A, Bellezza G, Scheibel M, and Bucciarelli E (2006): Detection of HER-2/neu (c-erbB-2) overexpression and amplification in breast carcinomas with ambiguous immunohistochemical results. A further contribution to defining the role of fluorescent in situ hybridization. Anticancer Res 26:2333-2338.

Signoretti S, Marcotullio L, Richardson A, Ramaswamy S, Isaac B, Rue M, Monti F, Loda M, and Pagano M (2002): Oncogenic role of the ubiquitin ligase subunit Skp2 in human breast cancer. J Clin Invest 110:633–641.

Simon R, Nocito A, Hübscher T, Bucher C, Torhorst J, Schraml P, Bubendorf L, Mihatsch MM, Moch H, Wilber K, Schötzau A, Kononen J, and Sauter G (2001): Patterns of HER-2/neu amplification and over-expression in primary and metastatic breast cancer. J Natl Cancer Inst 93:1141-1146.

Simpson PT, Gale T, Reis-Filho JS, Jones C, Parry S, Steele D, Cossu A, Budroni M, Palmieri G and Lakhani SR (2004): Distribution and significance of 14-3-3σ, a novel myoepithelial marker, in normal, benign, and malignant breast tissue. J Pathol 202: 274–285.

Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, and McCuire 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, and Shak S (2001): Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med 344:783-792.

Smalley M and Ashworth A (2003): Stem cells and breast cancer: A field in transit. Nature 3:832-844.

Sotiriou C, Neo S-Y, McShane LM, Korn EL, Long PM, Jazaeri A, Martiat P, Fox SB, Harris AL, and Liu ET (2003): Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci USA 100:10393-10398.

Sørlie 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, Eystein-Lonning P, and Borresen-Dale AL (2001): Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci 98:10869-10874.

Sørlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lonning PE, Brown PO, Borresen-Dale AL, and Botstein D (2003): Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci 100:8418-8423.

Sørlie T, Wang Y, Xiao C, Johnsen H, Naume B, Samaha RR, and Børresen-Dale A (2006): Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms. BMC Genomics 7:127-141.

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

Stingl J, Eaves CJ, Kuusk U, and Emerman JT (1998): Phenotypic and functional characterization in vitro of a multipotent epithelial cell present in the normal adult human breast. Differentiation 63:210-213.

Stingl J, Eirew P, Ricketson I, Shackleton M, Vaillant F, Choi D, Li HI, and Eaves CJ (2006): Purification and unique properties of mammary epithelial stem cells. Nature 439:993-997.

Stitziel NO, Mar BG, Liang J, and Westbrook CA (2004): Membrane-associated and secreted genes in breast cancer. Cancer Res 64:8682-8687.

Suzuki R, Atherton AJ, O'Hare MJ, Entwistle A, Lakhani SR, and Clarke C (2000): Proliferation and differentiation in the human breast during pregnancy. Differentiation 66:106-115.

Tacha DE and Miller RT (2004): Use of p63/P504S monoclonal antibody cocktail in immunohistochemical staining of prostate tissue. Appl Immunohistochem Mol Morphol 12:75-78.

Tanner M, Gancberg D, Di Leo A, Larsimont D, Rouas G, Piccard MJ, and Isola J (2000): Chromogenic in situ hybridization: A practical alternative for fluorescence in situ hybridization to detect HER-2/neu oncogene amplification in archival breast cancer samples. Am J Pathol 157:1467-1472.

Tanner M, Järvinen P, and Isola J (2001): Amplification of HER-2/neu and Topoisomerase IIα in primary and metastatic breast cancer. Cancer Res 61:5345-5348.

Tanner M, Kapanen AI, Junttila T, Raheem O, Grenman S, Elo J, Elenius K, and Isola J (2004): Characterization of a novel cell line established from a patient with Herceptin-resistant breast cancer. Mol Cancer Ther 3:1585-1592.

Tanner M, Isola J, Wiklund T, Erikstein B, Kellokumpu-Lehtinen P, Malmström P, Wilking N, Nilsson J, and Bergh J (2006): Topoisomerase IIalpha gene amplification predicts favorable treatment response to tailored and dose-escalated anthracycline-based adjuvant chemotherapy in HER-2/neu amplified breast cancer: Scandinavian Breast Group trial 9401. J Clin Oncol 24:2428-2436.

Taylor-Papadimitriou J, Stampfer M, Bartek J, Lewis A, Boshell M, Lane EB, and Leigh IM (1989): Keratin expression in human mammary epithelial cells cultured from normal and malignant tissue: relation to in vivo phenotypes and influence of medium. J Cell Sci 94: 403-413.

Thomas PA, Kirschmann DA, Cerhan JR, Folberg R, Seftor EA, Sellers TA, and Hendrix MJ (1999): Association between keratin and vimentin expression, malignant phenotype and survival in postmenopausal breast cancer patients. Clin Cancer Res 5:2698-2703.

Tsuda H, Takarabe T, Hasegawa T, Murata T, and Hirohashi S (1999): Myoepithelial differentiation in high-grade invasive ductal carcinoma with large central acellular zones. Hum Pathol 30:1134-1139.

Tsuda H, Takarabe T, Hasegawa F, Fukutomi T, and Hirohashi S (2000): Large, central acellular zones indicating myoepithelial tumor differentiation in high-grade invasive ductal carcinomas as markers of predisposition to lung and brain metastasis. Am J Surg Pathol 24:197-202.

Tsuda H, Morita D, Kimura M, Shinto E, Ohtsuka Y, Matsubara O, Inazawa J, Tamaki K, Mochizuki H, Tamai S, and Hiraide H (2005a): Correlation of KIT and EGFR over expression with invasive ductal breast carcinoma of the solid-tubular subtype, nuclear grade 3, and mesenchymal or myoepithelial differentiation. Cancer Sci 96:48-53.

Tsuda H, Tani Y, Weisenberger J, Kitada S, Hasegawa T, Murata T, Tamai S, Hirohashi S, Matsubara O, and Natori T (2005b): Frequent KIT and epidermal growth factor

receptor overexpressions in undifferentiated-type breast carcinomas with "stem-cell-like" features. Cancer Sci 96:333-339.

Turner NC and Reis-Filho JS (2006): Basal-like breast cancer and BRCA1 phenotype. Oncogene 25:5846-5853.

Turner NC, Reis-Filho JS, Russell AM, Springall RJ, Ryder K, Steele D, Savage K, Gillett CE, Schmitt FC, Ashworth A, and Tutt AN (2006): BRCA1 dysfunction in sporadic basal-like breast cancer. Oncogene [Epub ahead of print].

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.

van de Rijn M, Perou CM, Tibshirani R, Haas P, Kallioniemi O, Kononen J, Torhorst J, Sauter G, Zuber M, Köchli OR, Mross F, Dieterich H, Seitz R, Ross D, Botstein D, and Brown P (2002): Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. Am J Pathol 161:1991-1996.

van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AAM, 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 LT, Dai H, and van de Vijver MJ (2002): Gene expression profiling predicts clinical outcome of breast cancer. Nature 415:530-536.

Vera-Román JM and Rubio-Martínez LA (2004): Comparative assays for the HER-2/neu oncogene status in breast cancer. Arch Pathol Lab Med 128:627-633.

Wang X, Mori, I, Tang W, Nakamura M, Nakamura Y, Sato M, Sakurai T, and Kakudo K (2002): p63 expression in normal, hyperplastic and malignant breast tissues. Breast Cancer 9:216-219.

Wang ZC, Lin M, Wei L, Li C, Miron A, Lodeiro G, Harris L, Ramaswamy S, Tanenbaum DM, Meyerson M, Iglehart JD, and Richardson A (2004): Loss of heterozygosity and its correlation with expression profiles in subclasses of invasive breast cancer. Cancer Res 64:64-71.

Welm BE, Tepera SB, Venezia T, Graubert TA, Rosen JM, and Goodell MA (2002): Sca-1^{pos} cells in the mouse mammary gland represent an enriched progenitor cell population. Dev Biol 245:42-56.

Welm B, Behbod F, Goodell MA, and Rosen JM (2003): Isolation and characterization of functional mammary gland stem cells. Cell Prolif 36:17-32.

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

Wetzels RHW, Kuijpers HJH, Lane EB, Leigh IM, TRoyanovsky SM, HOlland R, van Haelst UJ, and Ramaekers FC (1991): Basal cell-specific and hyperproliferation-related keratins in human breast cancer. Am J Pathol 138:751-763.

Wixom CR, Albers EA, and Weidner N (2004): Her2 amplification; Correlation of chromogenic in situ hybridization with immunohistochemistry and fluorescence in situ hybridization. Appl Immunohistochem Mol Morphol 12:248-251.

Yamauchi H, Stearns V, and Hayes DF (2001): When is a tumor marker ready for prime time? A case study of c-erbB-2 as a predictive factor in breast cancer. J Clin Oncol 19:2334-2356.

Yang YH, Duboit S, Luu P, Lin DM, Peng V, Ngai J, and Speed TP (2002): Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 30:e15.

Yarden Y and Sliwkowski MX (2001): Untangling the ErbB signalling network. Mol Cell Biol 2:127-137.

Yaziji H, Gown AM, and Sneige N (2000): Detection of stromal invasion in breast cancer: the myoepithelial markers. Adv Anat Pathol 7:100-109.

ORIGINAL COMMUNICATIONS

Modern Pathology (2005) 18, 1321–1328

www.modernpathology.org

Cytokeratin 5/14-positive breast cancer: true basal phenotype confined to BRCA1 tumors

Mervi Laakso[1,2], Niklas Loman[3], Åke Borg[3] and Jorma Isola[2]

[1]*Department of Pathology, Seinäjoki Central Hospital, Seinäjoki, Finland;* [2]*Institute of Medical Technology, University of Tampere, Tampere, Finland and* [3]*Department of Oncology, University of Lund, Lund, Sweden*

Breast ducts contain two types of epithelial cells, inner luminal cells and outer basal/myoepithelial cells. These cells can be distinguished by their immunophenotype. Cytokeratins (CKs) 8 and 18 are expressed in the luminal layer, whereas CK5/14 and the transcription factor p63 characterize the basal epithelial layer. We studied a population-based cohort of 288 sporadic ductal invasive cancers and found 9% positive for CK5/14 and 4% positive for p63. Using a highly sensitive polymer-based immunohistochemical staining, all sporadic tumors were positive for the luminal CK8/18, including those positive for CK5/14. Pairs of primary tumors and metastases ($n = 38$) were always concordant for CK5/14 expression. The majority of the CK5/14-positive cases were of histologic grade III ($P = 0.0007$) and steroid hormone receptor negative ($P < 0.0001$). CK5/14 expression was inversely associated with *HER-2* oncogene amplification, but only in the subgroup of estrogen receptor-negative tumors ($P = 0.007$). In a separate set of 42 hereditary breast cancers, the majority (78%) of the *BRCA1*-associated tumors, but only one of 15 *BRCA2*-associated tumors was positive for CK5/14. In contrast to sporadic CK5/14-positive tumors, *BRCA1*-associated tumors displayed less intense CK8/18 staining, including some truly CK5/14-positive CK8/18-negative cases. These results suggest that CK5/14-positive sporadic breast cancers arise from glandularly committed progenitor cells rather than true CK8/18-negative basal cells.
Modern Pathology (2005) **18,** 1321–1328. doi:10.1038/modpathol.3800456; published online 1 July 2005

Keywords: basal phenotype; *BRCA1*; *BRCA2*; cytokeratin; *HER-2*; immunohistochemistry; oncogene; progenitor cell

The degree of differentiation and functional characteristics of epithelial cells giving rise to breast carcinoma have remained unclear. Most investigators have addressed breast carcinoma precursors by analyzing expression of cytokeratins (CKs) as differentiation markers, since their expression is thought to remain stable throughout carcinogenesis.[1] In general, breast cancers are thought to arise from luminally differentiated epithelial cells, as evidenced by strong expression of CK8, CK18 and CK19, similar to the situation in the cells lining the lumen of normal breast ducts.[1–3] A small fraction of breast cancers express CK5 together with its major partners CK14 and CK17,[2,4–7] which are normally found in the basal cell layer of the mammary duct.[1–3,8] For this reason, tumors expressing these CKs have been named 'basal-type' breast cancer. Breast cancers are generally thought to express either luminal (CK8/18/19+) or basal (CK5/14+) cytokeratins.[5,9,10] However, some CK5/14- and CK8/18-coexpressing tumors have also been found.[2,4,6,8] It has been proposed that tumors positive for CK5 originate from multipotent CK5-expressing progenitor cells,[2,8,11,12] located between the basal/suprabasal and luminal cell layers in normal ducts.[8] CK5-positive progenitor epithelial cells can gradually differentiate towards glandular and myoepithelial lineages.[2,8,11,12]

Basal phenotype tumors represent a histologically poorly differentiated estrogen receptor (ER)-negative tumor subtype.[4,6,7,9,10,13–16] The precise prevalence and clinicopathological characteristics of basal and luminal CK-expressing and -coexpressing tumors remain unclear. In particular, it is currently not known whether amplification of the *HER-2* oncogene is characteristic of basal or luminal phenotype tumors, or whether there is no association. The results of some studies have suggested that basal phenotype tumors may express less HER-2 protein.[7,16] These findings are supported by the results of gene expression microarray studies, which exclusively classify HER-2-expressing and basal phenotype tumors as separate entities.[10,17,18] In contrast,

Correspondence: M Laakso, MSc, Department of Pathology, Seinäjoki Central Hospital, Hanneksenrinne 7, Seinäjoki 60220, Finland.
E-mail: mervi.laakso@epshp.fi
Received 17 February 2005; revised 24 May 2005; accepted 25 May 2005; published online 1 July 2005

Birmbaum *et al*[9] have reported that *HER-2* oncogene amplification would be associated with basal phenotype breast cancer. More coherent data is available on hereditary breast cancers in *BRCA1* mutation carriers. These tumors have a high frequency of basal cell CK expression,[18–20] but for *BRCA2*-associated tumors, no immunohistochemical CK expression data have been reported.

In addition to CKs, the nuclear transcription factor p63 is a newly discovered marker of basal and myoepithelial cells in normal breast.[21–24] Its expression has been found in 10–12% of breast tumors,[24–26] and it is associated with high grade, large tumor size, nodal metastasis and ER negativity. However, some investigators have found no p63 in invasive breast carcinomas.[21,23] Thus, current data on p63 expression and its associations is controversial.

In the present work, we studied basal (CK5/14) and luminal (CK8/18/19) CK and p63 expression in a large population-based cohort of sporadic invasive ductal breast cancers as well as in tumors from a separate cohort of *BRCA1* and *BRCA2* germline mutation carriers. Our aim was to define the precise prevalence of basal phenotype breast tumors, and to examine clinicopathological correlations, with special emphasis on association with the *HER-2* oncogene. For this purpose, we analyzed an additional set of 141 *HER-2*-amplified breast tumors.

Materials and methods

Tumor Samples

The study material consisted of a population-based cohort of 288 consecutive sporadic invasive ductal breast cancers derived from the archives of the Department of Pathology at Seinäjoki Central Hospital. Histopathological information was collected for each patient sample (including grade, tumor size, lymph node metastasis, ER, progesterone receptor (PR) and *HER-2*). Separate sets of 27 tumors from *BRCA1* germline mutation carriers, 15 tumors from *BRCA2* germline mutation carriers (both from Department of Oncology, University of Lund, Lund, Sweden), 141 *HER-2*-amplified cancers and 38 pairs of primary and metastatic carcinomas (both from Department of Pathology, Seinäjoki Central Hospital, Seinäjoki, Finland and Institute of Medical Technology, University of Tampere, Tampere, Finland) were also studied. Mutation analyses of *BRCA1* and *BRCA2* have been described previously.[27]

To optimize and validate the basal phenotype immunostaining method, a subset ($n = 101$) of the tumors was prepared as five tissue microarray blocks, each containing cores (1 mm diameter) of carcinoma-containing tumor tissue. All samples were routinely formalin fixed and paraffin embedded. Sections were cut (3 μm) and used for immunohistochemistry.

Immunohistochemistry

The slides were deparaffinized and rehydrated before pretreatment. Antibodies tested were against: CK5, clone XM26 (Novocastra, Newcastle upon Tyne, UK); CK14, clone LL002 (Novocastra); CK5/6, clone D5/16B4 (DakoCytomation, Glostrup, Denmark) and p63, clone 4A4 + Y4A3 (Neomarkers, Fremont, CA, USA). Five different antigen retrieval pretreatments were compared for all antibodies using adjacent tissue microarray sections: protease (from Bacillus Licheniformis, Sigma-Aldrich, St Louis, MO, USA) treatment at 37°C for 3 min 30 s, heat treatment in an autoclave at 103°C for 5 min with 0.05 M Tris-HCl buffer, pH 9.0, containing 0.001 M EDTA and with Antigen Retrieval AR-10 Solution pH 10 (Biogenex, San Ramon, CA, USA), and the same autoclave pretreatment with the two different buffers followed by proteinase K treatment at room temperature for 10 min (ChemMate Proteinase K, DakoCytomation). In subsequent experiments, antigen retrieval for all the antibodies tested and the basal cell antibody cocktail (see Results) was carried out in autoclave at pH 9.0 as described earlier.

For the CK5/CK14/p63 antibody cocktail (clones XM26/LL002/4A4 + Y4A3) dilutions of 1:400, 1:400 and 1:1500 were used, respectively. Luminal CK8/18 were detected with a monoclonal antibody (5D3, 1:400, Novocastra). In control experiments, CK8, CK18, CK19 and CK17 were immunostained separately, using monospecific monoclonal antibodies (TS1, 1:800, NeoMarkers; DC10, 1:300, DakoCytomation; RCK108, 1:200, Euro-Diagnostica (Arnhem, Netherlands) and E3, 1:50, Neomarkers, respectively). ER and PR were immunostained via monoclonal antibodies 6F11 (1:400, Novocastra) and PgR 636 (1:1000, DakoCytomation), respectively. Smooth muscle actin was immunostained via monoclonal antibody 1A4 (DakoCytomation). For ER, PR, smooth muscle actin, CK8/18 and CK17, antigen retrieval was carried out as described for the CK5/CK14/p63 antibody cocktail. For monospecific CK8, CK18 and CK19 antibodies, proteinase K enzymatic pretreatment was carried out after autoclave treatment.

Immunostaining was carried out with a Techmate 500 + autostainer (DakoCytomation). Endogenous peroxidase was blocked with hydrogen peroxide (ChemMate Peroxidase-Blocking Solution, DakoCytomation) for 3 × 2 min 30 s followed by 30 min primary antibody incubation. A Powervision + polymer kit (PowerVision + ™, Immunovision Technologies Co., Brisbane, CA, USA) was used for detection. The reaction was visualized with DAB chromogen (2 × 5 min, Liquid DAB +, DakoCytomation) and enhanced with 0.5% $CuSO_4$ for 5 min. Hematoxylin (Chemmate hematoxylin, DakoCytomation) was used as a counterstain.

Basal cell carcinoma, skin and normal breast tissue were used as positive controls. For negative

controls, we omitted the primary antibodies. The specimens were interpreted as positive for CK5/14 if more than 20% of the neoplastic cells showed cytoplasmic staining and positive for p63 when the staining was nuclear. Two observers (ML and JI) agreed with the interpretation in more than 95% of the cases. Amplification of the *HER-2* oncogene was determined by chromogenic *in situ* hybridization as described previously.[28]

Results

Optimization of Immunohistochemical Staining for Basal Phenotype Breast Cancer

Tissue microarrays (consisting of 97 analyzable primary breast tumors) were first used to optimize detection of basal phenotype breast cancer. The two basal cell CKs (CK5 and CK14) did not identify exactly the same tumors when used alone. Among the 97 tumors, there were seven positive for CK5 and six positive for CK14. Two of the CK5-positive tumors were negative for CK14, and one tumor showed the opposite condition. The transcription factor p63 was expressed in the basal cells of normal breast ducts, and was found in three CK5/14-negative tumors and in one CK5/14-positive tumor. To ensure detection of basal phenotype breast cancers with high sensitivity, a cocktail of the three antibodies was used in subsequent experiments. The CK5/CK14/p63 antibody cocktail was further validated by comparing the results of antibody cocktail with those from tests with single antibodies. Only one tumor of the 97 showed a discordant result. An example of a CK5- and p63-positive case immunostained via CK5/CK14/p63 antibody cocktail and single antibodies is shown in Figure 1. More example micrographs can be seen in our website http://www.webmicroscope.net/supplements/LaaksoM.asp.

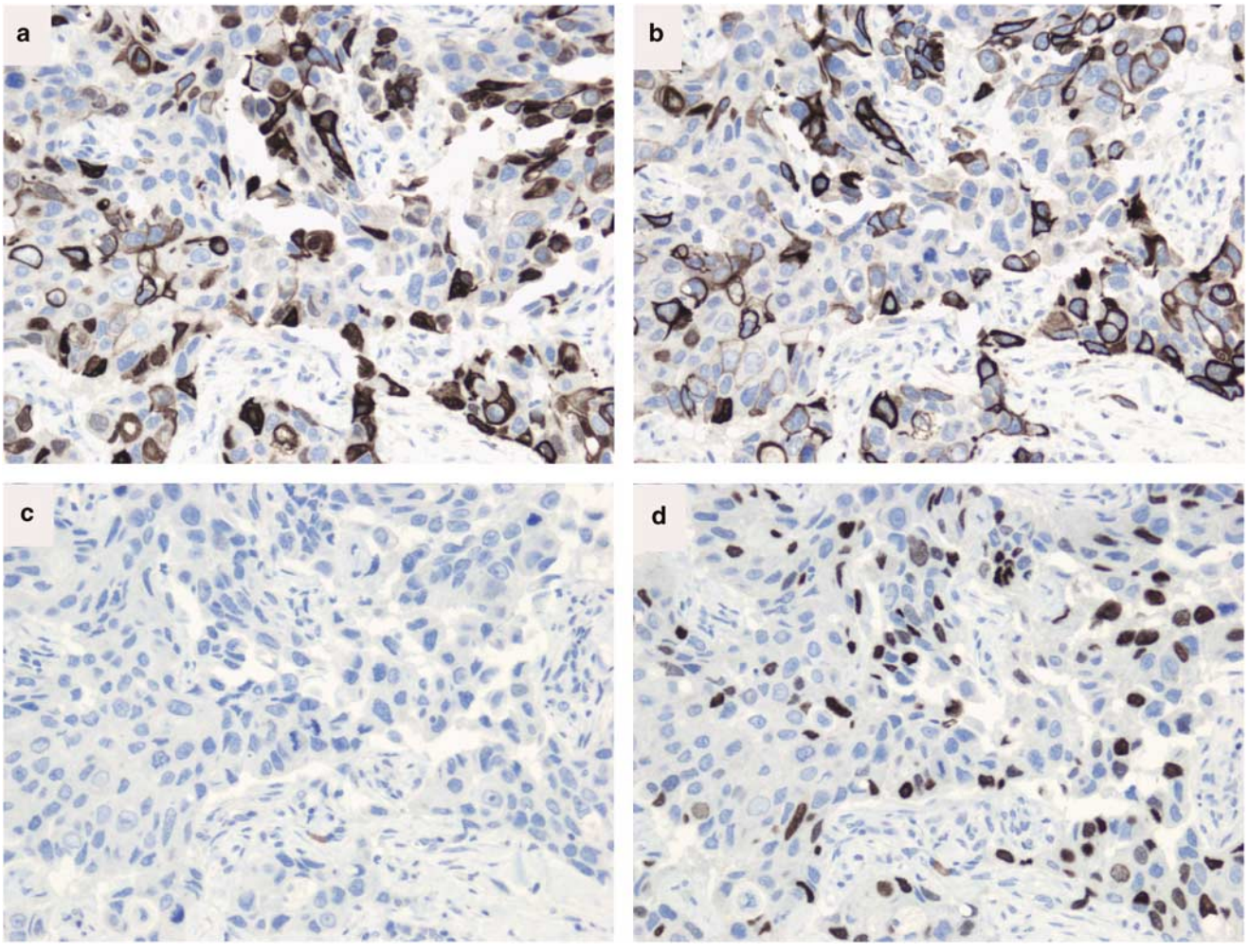

Figure 1 Immunohistochemical staining with CK5/CK14/p63 antibody cocktail (**a**) shows both cytoplasmic CK and nuclear p63 immunoreactivity. The single antibodies CK5 (**b**), CK14 (**c**) and p63 (**d**) show concordant results with antibody cocktail for this CK5/p63-positive tumor. More examples of the immunohistochemistry using the CK5/CK14/p63 antibody cocktail *vs* single antibodies can be seen as an appendix in our website http://www.webmicroscope.net/supplements/LaaksoM.asp.

Prevalence and Characterization of Basal Phenotype Breast Cancer

Of the 288 sporadic invasive ductal breast cancers, 24 (9%) were positive for CK5/14 and 12 (4%) were positive for p63 (Table 1). Matched pairs of primary tumors and metachronous metastases (from 38 patients) were always concordant for CK5/14 expression. There were four CK5/14-positive tumor-metastasis pairs (11%) and the rest were negative (examples are shown in the web appendix; http://www.webmicroscope.net/supplements/LaaksoM.asp).

Of the 288 breast cancers, 207 were immunostained for CK8/18 to investigate the coexpression of basal and luminal CKs. Surprisingly, all tumors in this cohort stained strongly for CK8/18, including all those scored positive for CK5/14 (Table 1). To study luminal CK expression pattern further, we immunostained CK5/14-positive tumors for CK8, CK18 and CK19 separately, using monospecific antibodies. All 20 CK5/14-positive tumors investigated were immunopositive for both CK8 and CK18, although some tumors exhibited relatively weaker staining intensity (Table 2). Staining for CK19 was also positive in all but one of the CK5/14-positive tumors (Table 2). Further characterization of the CK5/14-positive tumors with the myoepithelial differentiation marker smooth muscle actin showed no reactivity in any of the tumors studied.

To ascertain that CK17 does not identify more basal tumors than found with the antibody cocktail CK5/CK14/p63, we stained 110 tumors out of the 288 invasive ductal breast cancers with CK17. These tumors included all the CK5/14-positive tumors and the tissue microarrays used for antibody cocktail staining optimization. There were 11 (10%) CK17-positive tumors out of the 110 and only one of the CK17-positive tumors was CK5/14 negative (data not shown). Half of the CK5/14-positive tumors (10/20) showed CK17 positivity (data not shown).

Of the 27 hereditary *BRCA1* germline mutation tumors, 21 (78%) were positive for CK5/14 and one was also positive for p63 (Table 1, see http://www.webmicroscope.net/supplements/LaaksoM.asp for figures). Out of 20 CK5/14-positive *BRCA1*-associated tumors, 10 were CK8/18 positive with strong intensity (3+), five with moderate intensity (1−2+) and five were totally negative for CK8/18 (data not shown). In contrast to the *BRCA1*-associated tumors, only one of the 15 *BRCA2*-associated tumors showed CK5/14 and CK8/18 coexpression, whereas the rest showed the luminal CK5/14-negative CK8/18-positive phenotype (Table 1).

Clinicopathological Correlations of the CK5/14-Positive Tumors

The presence of CK5/14 immunoreactivity showed no correlation with patient age ($P=0.81$), tumor size ($P=0.42$) or the presence of axillary lymph node metastasis ($P=0.76$, Table 3). The great majority (72%) of the CK5/14-positive tumors were of histological grade 3 ($P=0.0007$) and, *vice versa*, 17% of the grade III tumors were CK5/14-positive. A strong association was found with negative ER and PR status ($P<0.0001$ for both). As many as 92% of the CK5/14-positive tumors were ER and PR negative. Positivity for CK5/14 was seen in 12% of the *HER-2* oncogene-amplified cases (amplification confirmed by means of CISH) and in 8% of the *HER-2*-nonamplified samples ($P=0.59$, Table 3). A total of 24% of the CK5/14-positive tumors showed *HER-2* amplification. The association of CK5/14 and *HER-2* was further studied in the subgroup of ER-negative tumors (Table 4). In this group, there was an inverse association, which was statistically highly significant ($P=0.007$). To confirm further the prevalence of CK5/14 expression and *HER-2* amplification, a separate set of 141 sporadic *HER-2*-amplified breast tumors was studied. There were 19 (13%) CK5/14-positive tumors in this cohort (Table 1). Of these tumors, 116 were stained for CK8/18, including all CK5/14-positive tumors. Confirming the result from

Table 2 Immunoreactivity of luminal CK8, CK18 and CK19 in CK5/14-positive invasive ductal breast cancer

Luminal CK	*Negative*	*Positive* *Intensity of the IHC staining*		
	0	*1+*	*2+*	*3+*
CK8	0%	10%	10%	80%
CK18	0%	15%	0%	85%
CK19	5%	0%	5%	90%

Data from 20 tumors.

Table 1 Proportion of tumors immunohistochemically positive for basal CK5/14, luminal CK8/18 and the transcription factor p63 in sporadic, sporadic *HER-2*-amplified and hereditary breast cancers

Tumor entity	*Positive for CK5/14*	*Positive for CK8/18*	*Positive for p63*
Sporadic ductal tumors (population-based cohort)	9% (25/288)	100% (207/207)	4% (12/288)
Sporadic *HER-2*-amplified tumors (selected cohort)	13% (19/141)	100% (116/116)	8% (11/141)
BRCA1-associated tumors	78% (21/27)	81% (21/26)	4% (1/27)
BRCA2-associated tumors	7% (1/15)	100% (15/15)	0% (0/15)

Table 3 Association of CK5/14 and p63 positivity with clinicopathological features in 288 sporadic invasive ductal breast cancers

Clinicopathological parameter	*CK5/14-positive/total*	P-*value*	*p63 positive/total*	P-*value*
All invasive ductal tumors	25/288 (9%)		12/288 (4%)	
Patient age (years)				
<50	5/47 (11%)	$P=0.81$	3/47 (6%)	$P=0.67$
≥50	20/241 (8%)		9/241 (4%)	
Tumor grade				
I	1/19 (5%)	$P=0.0007$	2/19 (11%)	$P=0.16$
II	6/163 (4%)		4/163 (2%)	
III	18/106 (17%)		6/106 (6%)	
Tumor size (cm)				
<2	13/120 (11%)	$P=0.42$	6/120 (5%)	$P=0.49$
2–5	11/140 (8%)		4/140 (3%)	
≥5	1/28 (4%)		2/28 (7%)	
Axillary lymph node metastasis				
No	17/182 (9%)	$P=0.76$	7/182 (4%)	$P=0.96$
Yes	8/106 (8%)		5/106 (5%)	
ER				
Negative	23/55 (42%)	$P<0.0001$	3/55 (5%)	$P=0.88$
Positive	2/233 (1%)		9/233 (4%)	
PR				
Negative	23/103 (22%)	$P<0.0001$	5/103 (5%)	$P=0.90$
Positive	2/185 (1%)		7/185 (4%)	
HER-2 amplification				
No	19/236 (8%)	$P=0.59$	10/236 (4%)	$P=0.80$
Yes	6/52 (12%)		2/52 (4%)	

Table 4 Association between CK5/14 positivity, *HER-2* amplification and ER status in a population-based cohort of 288 invasive ductal breast cancer

CK5/14 immunoreactivity vs ER and HER-2 status	*CK5/14 negative,* n = *263*	*CK5/14 positive (%),* n = *25*
ER−*HER2*−	12	18 (60%)
ER−*HER2*+	20	5 (20%)
ER+*HER2*−	205	1 (0.5%)
ER+*HER2*+	26	1 (4%)

$P=0.59$ (*HER-2 vs* CK5/14 in ER+ and ER− combined).
$P=0.007$ (*HER-2 vs* CK5/14 in ER− subgroup).

the population-based cohort, all tumors showed strong CK8/18 immunoreactivity (Table 1).

Clinicopathological Correlations Regarding Transcription Factor p63

In all, 4% of the invasive ductal breast cancers were p63 positive, but there was no correlation with patient age, tumor grade, tumor size, steroid hormone receptor status, axillary lymph node metastasis or *HER-2* oncogene amplification (Table 3). In the cohort of 141 *HER-2*-amplified tumors, 8% of the tumors showed p63 expression (Table 1), again with no correlation to ER status.

Discussion

Antibody cocktails have become increasingly popular in immunohistochemical staining of diagnostic tumor markers. When using a single chromogen, they are technically as easy to use as single antibodies, but can offer significantly more diagnostic information, as has been shown with the P504S/p63 antibody cocktail in prostate cancer.[29] In breast cancer, CK14 is a major partner of CK5 and both of these are associated with the basal phenotype.[1–6,8] CK6, which is often used in combination with CK5, using the bi-specific antibody D5/16B4,[2] is not expressed in normal breast tissue,[2,30] speaking against its importance as a predominant marker of basal and progenitor cells of the mammary duct.[2] CK17, also a known partner of CK5, is associated with basal phenotype breast cancer, but it has not been shown to relate to breast progenitor cells.[1–5,14] Further, only one CK17-positive CK5/14-negative tumor was found among 110 invasive ductal breast cancers in this study. Our results showed that CK5 and CK14 are coexpressed in most tumors, but that tumors expressing CK5 only or CK14 only also exist.

Somewhat surprisingly, nuclear transcription factor p63, which is another basal cell marker and has also been associated with breast cancer,[24–26] was expressed only in a minority of CK5/14-positive tumors. However, there were rare p63-positive cases alone without CK5 or CK14. For these reasons, we used a CK5/CK14/p63 antibody cocktail, which recognizes basal phenotype breast cancers specifically, but as widely as possible.

We found that 9% of invasive ductal breast cancers were CK5/14-positive and that they all coexpressed CK8/18. Clarke *et al*[31] suggested that breast progenitor cells positive for CK5 only, originally described by Boecker *et al*,[2] express low amounts of CK8/18, which is detectable only in frozen sections. When using a highly sensitive peroxidase-polymer-based immunostaining method and optimized antigen retrieval, we found that all sporadic invasive ductal tumors, including those positive for CK5/14, immunostained positively for CK8/18. The high sensitivity of CK8/18 immunodetection was also evident among *BRCA1*-associated tumors, previously considered as CK5/14 positive and mainly CK8/18 negative,[19] but which mostly immunostained positively for CK8/18. However, a relatively low staining intensity was found in many cases, which suggest that these tumors might have been scored CK8/18 negative with less sensitive immunohistochemical method. The CK8/18 antibody used (clone 5D3) recognizes CK8 and CK18, and also to some extent CK19.[32] To further characterize luminal CKs in CK5/14-positive tumors, immunostaining using monospecific CK8, CK18 and CK19 antibodies was carried out. All three luminal CKs were expressed in CK5/14-positive tumors, further confirming the coexpression of basal and luminal CKs in this tumor type. Further, all CK5/14-positive tumors in our study lacked expression of smooth muscle actin, indicating that these tumors are not differentiated towards myoepithelial lineage. Our observations indicate that CK5/14-positive invasive ductal tumors originate from glandularly committed progenitor cells of the breast, which have been shown to coexpress CK5/14 and CK8/18 during maturation to fully differentiated luminal cells.[2,8]

Cancers associated with the *BRCA1* germline mutation have been shown to be associated strongly with basal phenotype breast cancer.[18–20] This was also the finding in our material, as 78% of the *BRCA1*-associated tumors expressed CK5/14. In contrast to sporadic cancer, there were some CK5/14-positive CK8/18-negative phenotypes among the *BRCA1*-associated tumors, suggesting that these tumors probably originate from true basal phenotype cells. Foulkes *et al* has hypothesized that the key function of the wild-type BRCA1 is to act as a stem cell regulator and promote the differentiation towards glandular epithelium in the normal breast. In *BRCA1* mutated tumors, this transition has failed or not completed and basal cell phenotype gene expression is retained.[33] This kind of incidents might lead to CK5/14 positive but CK8/18 negative or weakly these luminal CKs expressing *BRCA1* germline-mutated tumors seen in this study. However, *BRCA1*-associated tumors were found to be a more heterogeneous group than previously thought. Although CK5/14 positivity was a common finding, there were several tumors resembling sporadic 'luminal-type' tumors in their CK profile (CK5/14−; CK8/18+). The results from *BRCA2* germline mutation carriers were more uniform and showed that their breast tumors were of luminal epithelium phenotype, similar to the majority of sporadic cancers.

Our results showed that CK5/14-positive breast cancers were exclusively ER- and PR-negative grade III carcinomas, which is in line with the results of earlier studies.[4,7,9,10,16–18] This finding could suggest a link to *HER-2* oncogene amplification, which has also been associated with negative ER status in a large number of studies (Konecny *et al*[34] and references therein). Unexpectedly, our results showed that there was no statistically significant association between CK5/14 positivity and *HER-2* amplification (CISH) when both ER-positive and ER-negative tumors were considered. However, there was a significant inverse association between *HER-2* oncogene amplification and CK5/14 immunopositivity within the subgroup of ER-negative tumors. This indicates that CK5/14-positive breast cancers and tumors with *HER-2* amplification are mainly different tumor entities. Published data on *HER-2* and basal phenotype are limited and somewhat conflicting. The results of microarray studies have suggested that basal phenotype breast cancer is *HER-2* nonamplified.[10,17,18] In contrast, Birmbaum *et al*[9] have reported that basal phenotype breast cancer is associated with *HER-2* amplification. Our results clearly show that *HER-2*-amplified CK5/14-positive breast cancers exist, although it is a minority. A recent *in vitro* study suggests that these tumors may form an interesting subgroup of *HER-2*-positive tumors, for example, with respect to therapeutic sensitivity to the HER-2-inhibiting drug trastuzumab.[35] Both CK5/14 positivity and *HER-2* amplification are each associated with poor clinical outcome,[17,18,26] but currently there are no data available on prognosis or therapeutic response prediction among patients whose tumors are *HER-2* amplified and positive for basal CKs.

In conclusion, CK5/14-positive breast cancers represent about 9% of sporadic invasive ductal breast cancers and 78% of *BRCA1*-associated tumors. These tumors are mostly aggressive grade III steroid hormone receptor-negative breast cancers, and they are inversely associated with *HER-2* oncogene amplification in the subgroup of ER-negative tumors. These tumors express basal and luminal CKs concomitantly and therefore may originate from luminally committed progenitor cells of the breast.

Acknowledgements

We thank Ms Ritva Kujala, Ms Helvi Salmela, Ms Pirjo Pekkala, Ms Päivi Kärki, Ms Sari Toivola and Ms Eeva Riikonen for skillful technical assistance. This study was financially supported by grants from the Pirkanmaa Hospital District Research Foundation, Medical Research Fund of Seinäjoki Central Hospital, Sigrid Juselius Foundation, the Finnish Cancer Foundation and Algol-Award.

References

1 Moll R, Franke WW, Schiller DL. The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells. Cell 1982;31:11–24.
2 Boecker W, Moll R, Poremba C, *et al.* Common adult stem cells in the human breast give rise to glandular and myoepithelial cell lineages: a new cell biological concept. Lab Invest 2002;82:737–745.
3 Chu PG, Weiss LM. Keratin expression in human tissues and neoplasms. Histopathology 2002;40: 403–439.
4 Malzahn K, Mitze M, Thoenes M, *et al.* Biological and prognostic significance of stratified epithelial cytokeratins in infiltrating ductal breast carcinomas. Virchows Arch 1998;433:119–129.
5 Wetzels RHW, Kuijpers HJH, Lane EB, *et al.* Basal cell-specific and hyperproliferation-related keratins in human breast cancer. Am J Pathol 1991;138:751–763.
6 Abd El-Rehim DM, Pinder SE, paish CE, *et al.* Expression of luminal and basal cytokeratins in human breast cancer. J Pathol 2004;203:661–671.
7 Korsching E, Packeisen J, Agelopoulos K, *et al.* Cytogenetic alterations and cytokeratin expression patterns in breast cancer: integrating a new model of breast differentiation into cytogenetic pathways of breast carcinogenesis. Lab Invest 2002;82: 1525–1533.
8 Boecker W, Buerger H. Evidence of progenitor cells of glandular and myoepithelial cell lineages in the human adult female breast epithelium: a new progenitor (adult stem) cell concept. Cell Prolif 2003;36:73–84.
9 Birmbaum D, Bertucci F, Ginestier C, *et al.* Basal and luminal breast cancer: basic or luminous? Int J Oncol 2004;25:249–258.
10 Perou CM, Sorlie T, Eisen MB, *et al.* Molecular portraits of human breast tumours. Nature 2000; 406:747–752.
11 Bánkfalvi A, Ludwig A, de-Hesselle B, *et al.* Different proliferative activity of the glandular and myoepithelial lineages in benign proliferative and early malignant breast diseases. Mod Pathol 2004;17:1051–1061.
12 Boecker W, Moll R, Dervan P, *et al.* Usual ductal hyperplasia of the breast is a committed stem (progenitor) cell lesion distinct from atypical ductal hyperplasia and ductal carsinoma *in situ*. J Pathol 2002;198:458–467.
13 Otterbach F, Bànkfalvi À, Berhner S, *et al.* Cytokeratin 5/6 immunohistochemistry assists the differential diagnosis of atypical proliferations of the breast. Histopathology 2000;37:232–240.
14 van de Rijn M, Perou CM, Tibshirani R, *et al.* Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. Am J Pathol 2002;161:1991–1996.
15 Gordon LA, Mulligan KT, Maxwell-Jones H, *et al.* Breast cell invasive potential relates to the myoepithelial phenotype. Int J Cancer 2003;106:8–16.
16 Nielsen TO, Hsu FD, Jensen K, *et al.* Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res 2004;10:5367–5374.
17 Sorlie T, Perou CM, Tibshirani R. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci USA 2001;98:10869–10874.
18 Sorlie T, Tibshirani R, Parker J, *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci USA 2003;100:8418–8423.
19 Foulkes WD, Brunet J-S, Stefansson IM, *et al.* The prognostic implication of the basal-like (Cyclin E^{high}/$p27^{low}$/$p53^{+}$/glomeruloid-microvascular-proliferation^{+}) phenotype of BRCA1-related breast cancer. Cancer Res 2004;64:830–835.
20 Foulkes WD, Stefansson IM, Chappuis PO, *et al.* Germline *BRCA1* mutations and a basal epithelial phenotype in breast cancer. J Natl Cancer Inst 2003;95:1482–1485.
21 Barbareschi M, Pecciarini L, Cangi MG, *et al.* p63, a p53 homologue, is a selective nuclear marker of myoepithelial cells of the human breast. Am J Surg Pathol 2001;25:1054–1060.
22 Nylander K, Vojtesek B, Nenutil R, *et al.* Differential expression of p63 isoforms in normal tissues and neoplastic cells. J Pathol 2002;198:417–427.
23 Wang X, Mori I, Tang W, *et al.* p63 expression in normal, hyperplastic and malignant breast tissues. Breast Cancer 2002;9:216–219.
24 Ribeiro-Silva A, Zambelli Ramalho LN, Britto Garcia S, *et al.* The relationship between p63 and p53 expression in normal and neoplastic breast tissue. Arch Pathol Lab Med 2003;127:336–340.
25 Reis-Filho JS, Simpson PT, Martins A, *et al.* Distribution of p63, cytokeratin 5/6 and cytokeratin 14 in 51 normal and 400 neoplastic human tissue samples using TARP-4 multi-tumor tissue microarray. Virchows Arch 2003;443:122–132.
26 Makretsov NA, Huntsman DG, Nielsen TO, *et al.* Hierarchical clustering analysis of tissue microarray immunostaining data identifies prognostically significant groups of breast carcinoma. Clin Cancer Res 2004;10:6143–6151.
27 Loman N, Johannsson O, Kristoffersson U, *et al.* Family history of breast and ovarian cancers and BRCA1 and BRCA2 mutations in a population-based series of early-onset breast cancer. J Natl Cancer Inst 2001;93:1215–1223.
28 Isola J, Tanner M, Forsyth A, *et al.* Interlaboratory comparison of *HER-2* oncogene amplification as detected by chromogenic and fluorescence *in situ* hybridization. Clin Cancer Res 2004;10: 4793–4798.
29 Tacha DE, Miller RT. Use of p63/P504S monoclonal antibody cocktail in immunohistochemical staining of prostate tissue. Appl Immunohistochem Mol Morphol 2004;12:75–78.
30 Moll R. Cytokeratins as markers of differentiation in the diagnosis of epithelial tumors. Subcell Biochem 1998;31:205–262.

31 Clarke CL, Sandle J, Parry SC, *et al.* Cytokeratin 5/6 in normal human breast: lack of evidence for a stem cell phenotype. J Pathol 2004;204:147–152.
32 Angus B, Purvis J, Stock D, *et al.* NCL-5D3: a new monoclonal antibody recognizing low molecular weight cytokeratins effective for immunohistochemistry using fixed paraffin-embedded tissue. J Pathol 1987;153:377–384.
33 Foulkes WD. BRCA1 functions as a breast stem cell regulator. J Med Genet 2003;41:1–5.
34 Konecny G, Pauletti G, Pegram M, *et al.* Quantitative association between HER-2/neu and steroid hormone receptor in hormone receptor positive primary breast cancer. J Natl Cancer Inst 2003;95: 142–153.
35 Tanner M, Kapanen AI, Junttila T, *et al.* Characterization of a novel cell line established from a patient with Herceptin-resistant breast cancer. Mol Cancer Ther 2004;3:1585–1592.

Basal-like phenotype is not associated with patient survival in estrogen receptor negative breast cancers

Mervi Jumppanen[1,2]*, Sofia Gruvberger-Saal[3]*, Päivikki Kauraniemi[2], Minna Tanner[2,4], Pär-Ola Bendahl[3], Mikael Lundin[5], Morten Krogh[6], Pasi Kataja[2], Åke Borg[3], Mårten Fernö[3] and Jorma Isola[2]

[1]Department of Pathology, Seinäjoki Central Hospital, Hanneksenrinne 7, FIN-60220 Seinäjoki, Finland
[2]Institute of Medical Technology, University and University Hospital of Tampere, Biokatu 6, FIN-33014 Tampere, Finland
[3]Department of Oncology, Clinical Sciences, University of Lund, SE-221 85 Lund, Sweden
[4]Department of Oncology, Tampere University Hospital, FIN-33014 Tampere, Finland
[5]Biomedical Informatics group, Department of Oncology, University of Helsinki, FIN-00290 Helsinki, Finland
[6]Department of Theoretical Physics, Lund University, SE-221 85 Lund, Sweden

*These authors contributed equally.

ABSTRACT

Introduction Basal phenotype or basal-like breast cancers are characterized by basal epithelium cytokeratin (CK5/14/17) expression, negative estrogen receptor (ER) status and distinct gene expression signature. We studied the clinical and biological features of the basal phenotype tumors determined by immunohistochemistry and cDNA microarrays especially within the ER-negative subgroup.

Methods Immunohistochemistry (IHC) was used to evaluate the CK5/14 status of 445 stage II breast cancers. Gene expression signature of the CK5/14 immunopositive tumors was investigated within a subset (100) of the breast tumors (including 50 ER-negatives) using cDNA microarray. Survival for basal phenotype tumors as determined by CK5/14 IHC and gene expression signature was assessed.

Results From the 375 analyzable tumor specimens 48 (13%) were immunohistochemically positive for CK5/14. We found adverse distant disease-free survival for the CK5/14-positive tumors during the first years (three years HR=2.23, 95% CI=1.17-4.24, p=0.01 and five years HR=1.80, 95% CI=1.02-3.15, p=0.04) but the significance was lost at the end of the follow-up period (10 years HR=1.43, 95% CI=0.84-2.43, p=0.19). Gene expression profiles of immunohistochemically determined CK5/14-positive tumors within the ER-negative tumor group implicated 1713 differently expressed genes (p-value <0.05). Hierarchical clustering analysis using top 500 of these genes formed one basal-like and a non-basal like cluster also within the ER-negative tumor entity. A highly concordant classification could be constructed using a published gene set (Sorlie's intrinsic gene set, concordance 90%). Both gene sets identified a basal-like cluster that included most of the CK5/14-positive tumors, but also immunohistochemically CK5/14-negative tumors. Within the ER-negative tumor entity there was no survival difference between the non-basal and basal-like tumors as identified by immunohistochemical or gene expression based classification.

Conclusions Basal cytokeratin positive tumors have a biologically distinct gene expression signature from other estrogen receptor negative tumors. Even if basal cytokeratin expression predicts early relapse among non-selected tumors, the clinical outcome of basal tumors is similar to non-basal ER-negative tumors. Immunohistochemically basal cytokeratin positive tumors belong almost always to the basal-like gene expression profile, but this cluster includes also few basal cytokeratin negative tumors.

INTRODUCTION

cDNA microarray studies have shown that the most powerful denominator in determining the gene expression profiles and prognostic groups of breast cancer is estrogen receptor (ER) and ER-related genes [1-5]. Breast cancers have been separated by gene expression profiles to luminal, basal-like, ERBB2 and normal breast-like subgroups [6-9]. Basal-like tumors express many of the genes characteristics of breast basal epithelial cells [6] and the most typical feature of the basal-like breast cancers is the lack of expression of estrogen receptor (ER) and genes usually co-expressed with ER [6-9].

In addition to the gene expression microarray studies, basal phenotype breast tumors have long been identified by using basal cytokeratin immunohistochemistry (IHC) [10-20]. Basal cytokeratin (CK5/14/17) positive tumors represent ~10% of sporadic breast carcinomas and are almost exclusively ER-negative, poorly differentiated, and have been associated to EGFR, p53, vimentin and c-kit immunopositivity and Bcl-2 negativity [11,12,14-16,19-21]. Even though gene expression studies separate the basal-like tumors from the ERBB2 tumor subgroup [6-9], there are some immunohistochemically basal cytokeratin expressing tumors, which show *HER-2* oncogene amplification [12,17,22]. The relationship between immunohistochemical and microarray based classification of basal phenotype breast cancer has not been established.

Apart from hypothesis-generating scientific research, a breast tumor classification should correlate with clinical outcome of patients or predict efficacy to therapy. Negative ER status, which is the most prominent feature of basal phenotype tumors, is a well-established prognostic and predictive factor in breast cancer. The microarray studies that have shown that basal-like tumors show poor prognosis when compared to estrogen receptor positive luminal tumor groups but not when compared to ERBB2 tumor cluster [7,8]. Immunohistochemical studies using basal cytokeratin IHC for the basal breast cancer phenotype classification have almost exclusively addressed that basal phenotype tumors have poor prognosis, but they have also made the comparison in cohorts not selected by matching ER status (ER-negative) [10,11,16,17,20,23-25]. In this study we defined the gene expression profile of basal cytokeratin immunopositive tumors and studied the clinical outcome especially within ER-negative tumor entity.

MATERIALS AND METHODS

Tumor samples

The tumor cohort comprised 445 primary stage II breast cancers collected from the South Sweden Health Care Region between 1985-1994 with approval from the Lund University Hospital ethics committee and was described earlier in more detail in Chebil et al. [26]. In this study, patients treated with 20 mg daily tamoxifen for two years with complete follow-up data and uniform immunohistochemical method for hormone receptor analysis were included. Radical mastectomy or breast-conserving surgery was used with axillary lymph node dissection. Radiotherapy was introduced to all patient treated with breast-conserving surgery and to patients with lymph node positive disease. The patients were not treated with adjuvant chemotherapy. Median follow-up time for distant disease-free survival was six years.

Immunohistochemistry (IHC)

The formalin-fixed paraffin-embedded sample material was provided as eight tissue microarrays (TMA) containing three core samples (diameter 0.6 mm) for each primary tumor. The immunohistochemical stainings with CK5/CK14/p63 antibody cocktail (XM26, 1:400, Novocastra, Newcastle upon Tyne, UK; LL002, 1:400, Novocastra; 4A4+Y4A3, 1:1500, Neomarkers, Fremont, CA, respectively) and with p53 antibody (DO-7, 1:500, Novocastra) were done as described earlier [12,22 respectively]. Hormone receptors (ER and PR) have been conducted earlier by IHC from the original tissue blocks as described in Chebil et al. [26].

Analysis of the *HER-2* oncogene amplification was done by using chromogenic in situ

hybridization (CISH) method as described earlier [27]. Histological type of the tumors was determined according to WHO classification as described in Chebil et al. [26].

Sample scoring

Immunohistochemically stained TMA samples for CK5/CK14/p63 and p53 as well as *HER-2* CISH stainings were scanned using a virtual microscopy technique as described earlier [28]. Immunostaining for CK5/CK14/p63 was considered CK5/14-positive if at least 20% of the tumor cells showed cytoplasmic staining and positive for p63 when the staining was nuclear. p53 was regarded as positive when at least 20% of the tumor cells were stained. *HER-2* oncogene was considered amplified if six or more gene copies were found per cell in at least 10% of the tumor cells.

Statistical analysis

Fisher's exact test and Chi-square test were used to test the significance of the cross-tabulated data (using Stata 9.2 [Stata Corporation, College Station, TX] and MedCalc [MedCalc Software, Mariakerke, Belgium] statistical softwares). Survival analyses were calculated using Kaplan-Maier life table curves, log-rank test and univariate Cox-model. Distant disease-free survival was calculated from the primary diagnosis to the date of an event (distant recurrence or death) or for event-free patients to the date of the most recent follow-up. All reported p-values are two-sided.

Gene expression microarrays

cDNA microrrays were manufactured in the SWEGENE Microarray Facility, Department of Oncology, Lund University. The gene set consisted of 24,301 sequence-verified IMAGE clones (Research Genetics, Huntsville, AL) and 1,296 internally-generated clones, together representing ≈16,000 Unigene clusters (build 180) and ≈1,200 unclustered ESTs. The clones were PCR amplified using vector-specific primers essentially as previously described [29].

A selected subset (n=100, of which 50 were ER-negative) from the total cohort was analyzed with microarrays. Nineteen of these tumors showed positive CK5/14 staining and the rest were negative. Only one of the CK5/14-positive tumors was ER-positive. Total RNA was extracted from grossly dissected frozen tissue samples (approximately 100 mg) by subsequent Trizol (Invitrogen, Carlsbad, CA) and the RNeasy kit (Qiagen, Hilden, Germany). For each hybridization, 15 μg of Universal Human Reference RNA (Stratagene, La Jolla, CA) was used to synthesize reference Cy5-labeled targets and 25 μg of sample total RNA for Cy3-labeled targets using anchored oligo(dT) primers and the CyScribe indirect amino-allyl cDNA synthesis and labeling protocol and GFX purification columns (Amersham Biosciences, Buckinghamshire, UK). Together with blocking agents (12 μg poly-d(A), 6 μg yeast tRNA, and 20 μg Cot-1 DNA), targets were hybridized to the microarrays for 18 hours under a glass coverslip using humidified Corning hybridization chambers at 42°C and the Pronto Universal Hybridization System (Corning Inc., Corning, NY). Slides were scanned at 10 μm resolution in an Agilent DNA Microarray Scanner (Agilent Technologies, Palo Alto, CA) and the images were analyzed using GenePix Pro software (Axon Instruments, Union City, CA).

Microarray data analysis

The data was analyzed using the BASE software [30]. In brief, background-corrected intensities for sample and reference channels were calculated by subtracting the median local background signal from the median foreground signal for each spot. Filters were applied to remove all spots flagged during image analysis. Data within individual arrays were then normalized using an implementation of the lowess algorithm [31]. Poorly-measured/expressed spots with a signal-to-noise ratio ≤3 in either the Cy3 or Cy5 channel were removed, and genes with missing data in more than 20 percent of all arrays or genes with a variation across arrays of ≤0.45 standard deviations of the $\log_2$(ratio) were filtered, leaving 10,479 informative genes. The expression ratios

for each gene were then median-centered across all tumors.

To generate a gene list for the basal phenotype tumors, correlation scores were calculated between gene expression (log_2(ratio)) for all reporters and the CK5/14 immunopositive tumors [32]. To evaluate the significance of the expression signatures between the two annotation classes (CK5/14-positive and CK5/14-negative), 1,000 permutations were run where the samples were randomly given an annotation label and the p-value for a score was calculated as the average number of reporters exceeding the score in the permutation test, divided by the total number of reporters in the gene list. The false discovery rate (FDR), *i.e.* the estimated number of genes per a given set of scored genes that could receive an equal or better score by chance, was calculated by random permutations and used as an indicator of the robustness of the gene expression profile. An FDR of 0 percent indicates no false positives whereas an FDR of 100 percent indicates a completely random signal. Gene expression profiles were analyzed with hierarchical clustering using centered Pearson correlation and average linkage clustering [33].

The ranked gene list was subject to gene ontology annotation analysis using EASE (Expression Analysis Systematic Explorer; http://david.niaid.nih.gov/david/ease.htm) whereby only biological process ontology categories were included and the enrichment of categories in the gene list was evaluated by comparison to the total list of genes used for the microarray analysis. An EASE score of $p \leq 0.05$ was considered to be significant. The UniGene clusters representing the top 200 genes were annotated with subcellular location by cross-reference to two published microarray datasets [33,34] and to Swiss-Prot. The Swiss-Prot Subcellular Locations annotations were downloaded from the DRAGON database [36]. A gene was considered to be membrane associated or secreted if the Swiss-Prot annotation contained one of the words "membrane", "vesicle" or "secreted", or if the membrane to cytosolic ratio in the polysome fraction study exceeded 2 or 1.08 in the studies by Diehn et al. or Stitziel et al., respectively [34,35]. Primary expression data is available from the NCBI Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/).

RESULTS

Immunohistochemical detection of basal phenotype tumors

Immunohistochemical analysis was performed on TMAs containing 445 tumors, of which 375 (84%) were analyzable for CK5/CK14/p63 antibody cocktail. There were 48 (13%) CK5/14-positive and 13 (3.5%) p63-positive tumors. Although CK5/14 and p63 are co-expressed in normal cells of breast ducts, there was no association in malignant epithelial cells (p=0.22). The CK5/14 immunopositivity was significantly correlated to negative ER status ($p<0.0001$, data not shown). There were 13 ER-positive basal cytokeratin expressing tumors. Association with negative PR status ($p<0.0001$), with negative lymph node status (p=0.0005), and with p53 immunopositivity (p=0.003) were also seen but there was no association with *HER-2* oncogene amplification (p=0.80, data not shown). Among the 95 ER-negative tumors, 35 (37%) showed positive staining for CK5/14 (Table 1). When CK5/14 positivity was correlated with clinicopathological characteristics within the ER-negative tumor subgroup, association to negative lymph node status and positive p53 status were not seen (p=0.14 and p=0.65, respectively), but significant association between CK5/14 immunopositivity and negative *HER-2* status emerged (p=0.01, Table 1). The majority of the basal cytokeratin positive tumors were of the ductal histotype (80%) and the rest were of the medullary or atypical medullary histotypes (20%, Table 1). Over half (7/12) of the medullary histotype tumors (medullary or atypical medullary) were in fact CK5/14-positive.

Table 1. Clinicopathological characteristics of estrogen receptor negative breast tumors according to basal cytokeratin (CK5/14) status.

Clinicopathological parameter	*CK5/14-negative (%)*	*CK5/14-positive (%)*	*p-value**
All estrogen receptor negative tumors	60 (63%)	35 (37%)	
Axillary lymph node status			
negative	23 (38%)	19 (54%)	0.14
positive	37 (62%)	16 (46%)	
HER-2 status			
non-amplified	35 (58%)	30 (86%)	0.01
amplified	18 (30%)	3 (9%)	
data missing	7 (12%)	2 (6%)	
p63			
negative	57 (95%)	33 (94%)	1.00
positive	3 (5%)	2 (6%)	
p53			
negative	26 (43%)	13 (37%)	0.65
positive	26 (43%)	17 (49%)	
data missing	8 (13%)	5 (14%)	
Histologic type			
Invasive ductal or mixed type	49 (82%)	28 (80%)	0.10
Invasive lobular	5 (8%)	0 (0%)	
Medullary or atypical medullary	5 (8%)	7 (20%)	
Other types	1 (2%)	0 (0%)	

* The p-values have been calculated without the data missing values with Fisher's exact test except p-value of histological type with Chi-square test.

Gene expression profile of basal phenotype tumors

A clear difference was seen in gene expression profiles between the basal cytokeratin (CK5/14) immunopositive and negative subgroups in the whole data set (False Discovery Rate= 0.03% per 100 genes and 0.3% per top 500 using the Golub algorithm) including both ER-positive and ER-negative tumors. However, since the basal phenotype determined by IHC strongly correlated with negative ER status (only one of the 50 ER-positive tumors stained positive for CK5/14), and since ER status has been shown to have a strong influence on the gene expression signature of breast tumors [2,4,6], we performed analysis in the subset of ER-negative tumors (n=50) separately. In this subset CK5/14-positive and CK5/14-negative tumors were also associated with two distinct gene expression signatures (False Discovery Rate 6.7% per top 100 genes and 16.1% per top 500 genes). Hierarchical clustering analysis of the ER-negative tumors using top 500 basal discriminatory genes generated within the ER-negative tumor group identified two separate clusters (Figure 1., see Supplemental figure 1 for the heat map): one cluster holding a large number of CK5/14-positive tumors (17/24) in addition to seven CK5/14-negative tumors, and another cluster where all but one of the tumors (25/26) were immunohistochemically CK5/14-negative and were frequently *HER-2* oncogene amplified (18/26). Although the signal for basal phenotype among ER-negative tumors was weaker than in the whole data set, where the classification may have been highly influenced by the strong estrogen receptor related signal, it was statistically highly significant (1713 genes were identified with $p<0.05$; see Supplemental Table 1 for the top 200 genes).

We next explored how the so called "intrinsic" gene set generated by Perou and co-workers [6-8] would perform in our data set. Mapping of their intrinsic gene list [8] to our data using Unigene Cluster ID as an identifier produced a list of 522 clones. These clones were used to cluster the whole data set, which gave expected results separating basal/ER-, luminal/ER+ and ERBB2+/ER- tumor groups from each other similarly to the original study (data not shown) [6,8]. Hierarchical clustering of the ER-negative tumor group

separately, using the intrinsic gene set, generated a dendrogram with two major subgroups very similar to the hierarchical clustering analysis using our top 500 ranked basal genes (concordance 90%, p=0.0001, Figure 2). The basal-like cluster included the majority of the CK5/14-positive tumors and nine additional CK5/14 negative tumors. The tumors in the non-basal subgroup showed frequent *HER-2* amplification (17/27) and predominantly a CK5/14-negative immunophenotype (23/27, Figure 2, see Supplemental figure 2 for the heat map). The basal phenotype classification by Sorlie's intrinsic gene set correlated strongly with basal cytokeratin immunohistochemistry (concordance 76%, p=0.0011). Interestingly, seven of the nine misclassified CK5/14-negative tumors by Sorlie´s intrinsic gene set were also found to belong to the basal-like cluster when our top 500 CK5/14-associated genes were used in hierarchical clustering analysis.

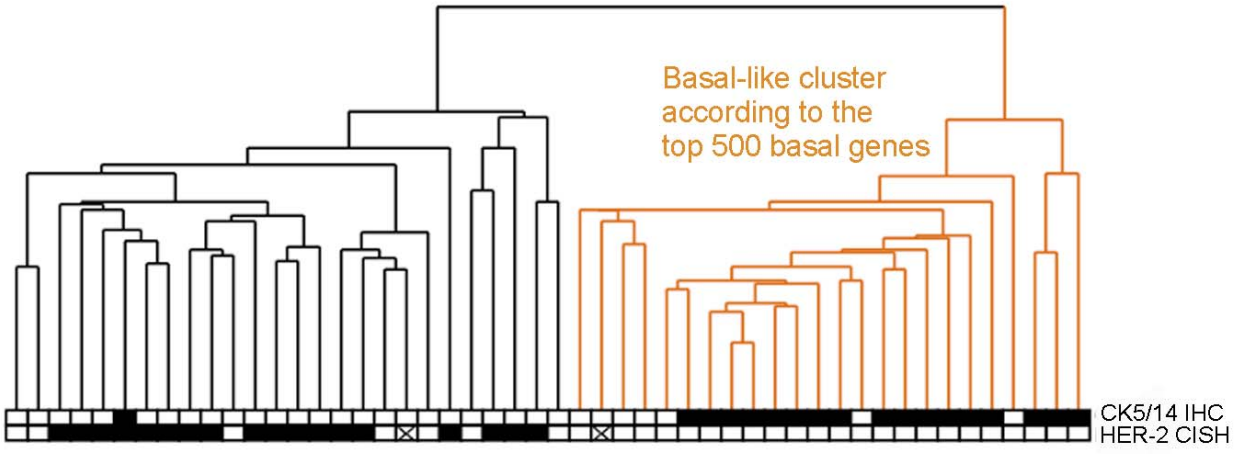

Figure 1. Hierarchical clustering of 50 ER-negative breast cancers based on the top 500 gene set generated for the CK5/14-positive tumors. Yellow indicates the basal-like cluster and black non-basal like cluster. The black boxes below indicate the immunohistochemically CK5/14-positive tumors and the *HER-2* oncogene amplified tumors (solid box=positive, open box=negative, crossed box=data missing).

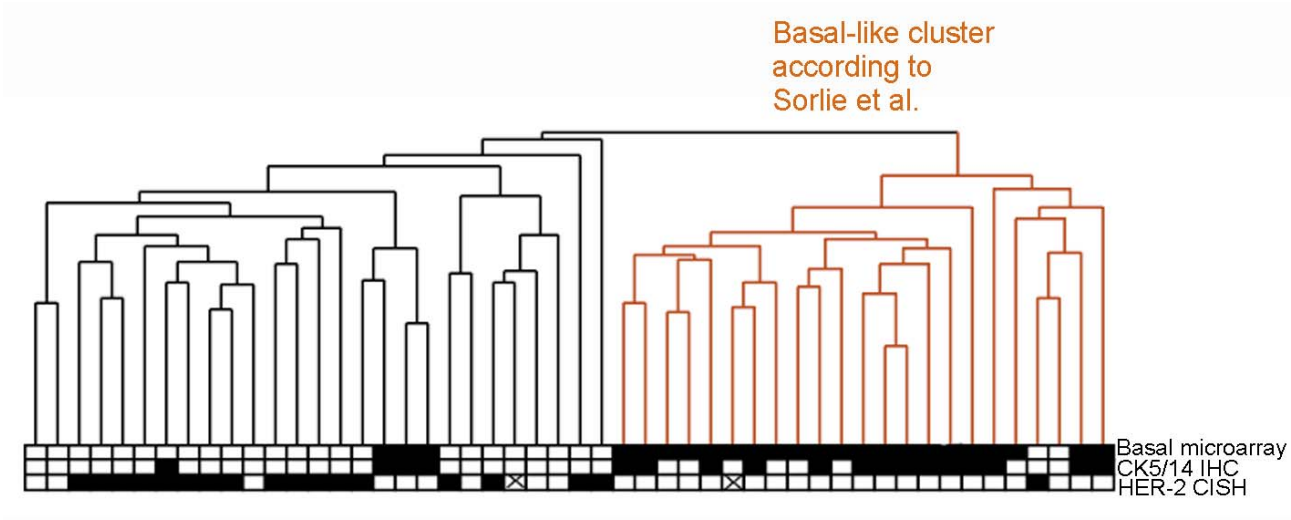

Figure 2. Hierarchical clustering of 50 ER-negative tumors based on the intrinsic gene set by Sorlie et al. Yellow indicates the basal-like cluster and black non-basal like cluster. The black boxes below indicate the basal-like cluster by the top 500 basal genes, immunohistochemically CK5/14-positive tumors, and *HER-2* amplified tumors (solid box=positive, open box=negative, crossed box=data missing).

The gene list generated for the basal cytokeratin immunopositive tumors within the ER-negative tumor entity (Supplemental Table 1) included genes associated with ER status like *TTF1* (rank 13) and *XBP1* (rank 16) and other genes previously associated with the basal-like tumor subtype such as *CRYAB* (rank 26), *TRIM29* (rank 51), *ERBB2* (rank 55), *ANXA8* (rank 134) and *EGFR* (rank 193) [6-9]. Twelve of the genes with a high expression in basal-like tumors (within the top 200 genes) were annotated as having a membrane bound cellular localization, but not to the mitochondria or the Golgi apparatus (Supplemental table 1).

Distant disease-free survival of basal phenotype tumors

Association of the basal status with patient prognosis was evaluated first in the immunohistochemically defined basal (CK5/14-positive) and non-basal (CK5/14-negative) tumor subgroups. In the whole tumor material, the distant disease-free survival was significantly shorter for the CK5/14-positive tumors during the first years of follow-up (three years HR=2.23, 95% CI=1.17-4.24, p=0.01 and five years HR=1.80, 95% CI=1.02-3.15, p=0.04), but this difference was lost at the end of the follow-up period (10 years HR=1.43, 95% CI=0.84-2.43, p=0.19, Figure 3). Next we studied clinical outcome within the ER-negative entity. The survival of immunohistochemically CK5/14-positive and CK5/14-negative tumor groups were identical, as demonstrated by the superimposed Kaplan-Meier curves and log-rank test (p=0.93, Figure 4A). The same result was obtained when the basal-like classification was based on gene expression microarrays (p=0.42 and p=0.55 for classifications based on our gene list and Sorlie's gene list, Figure 4B and C, respectively).

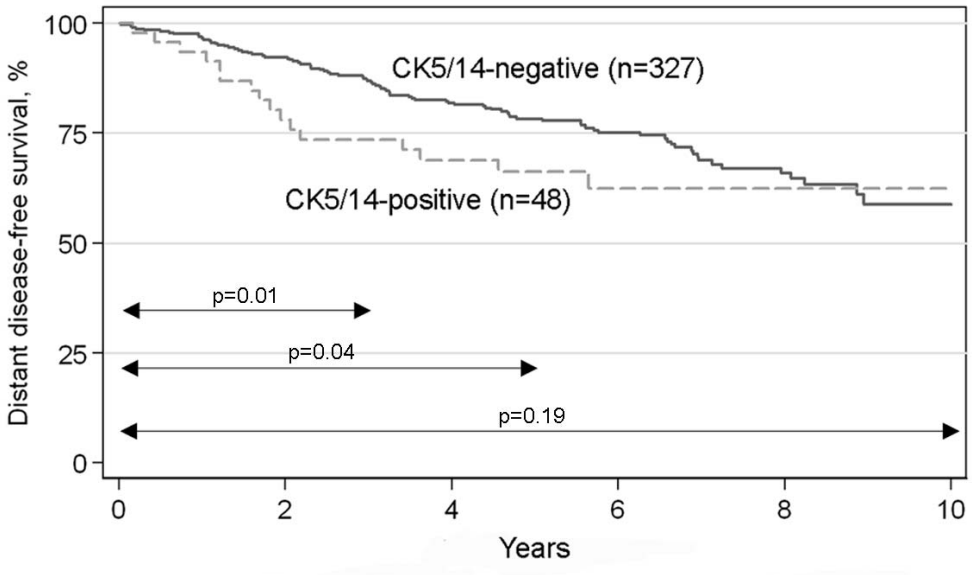

Figure 3. Distant disease-free survival of immunohistochemically CK5/14-negative and CK5/14-positive tumors in the whole data set. The basal cytokeratin positive tumors show significantly shorter survival during the first years of the follow-up, but this difference is lost in time.

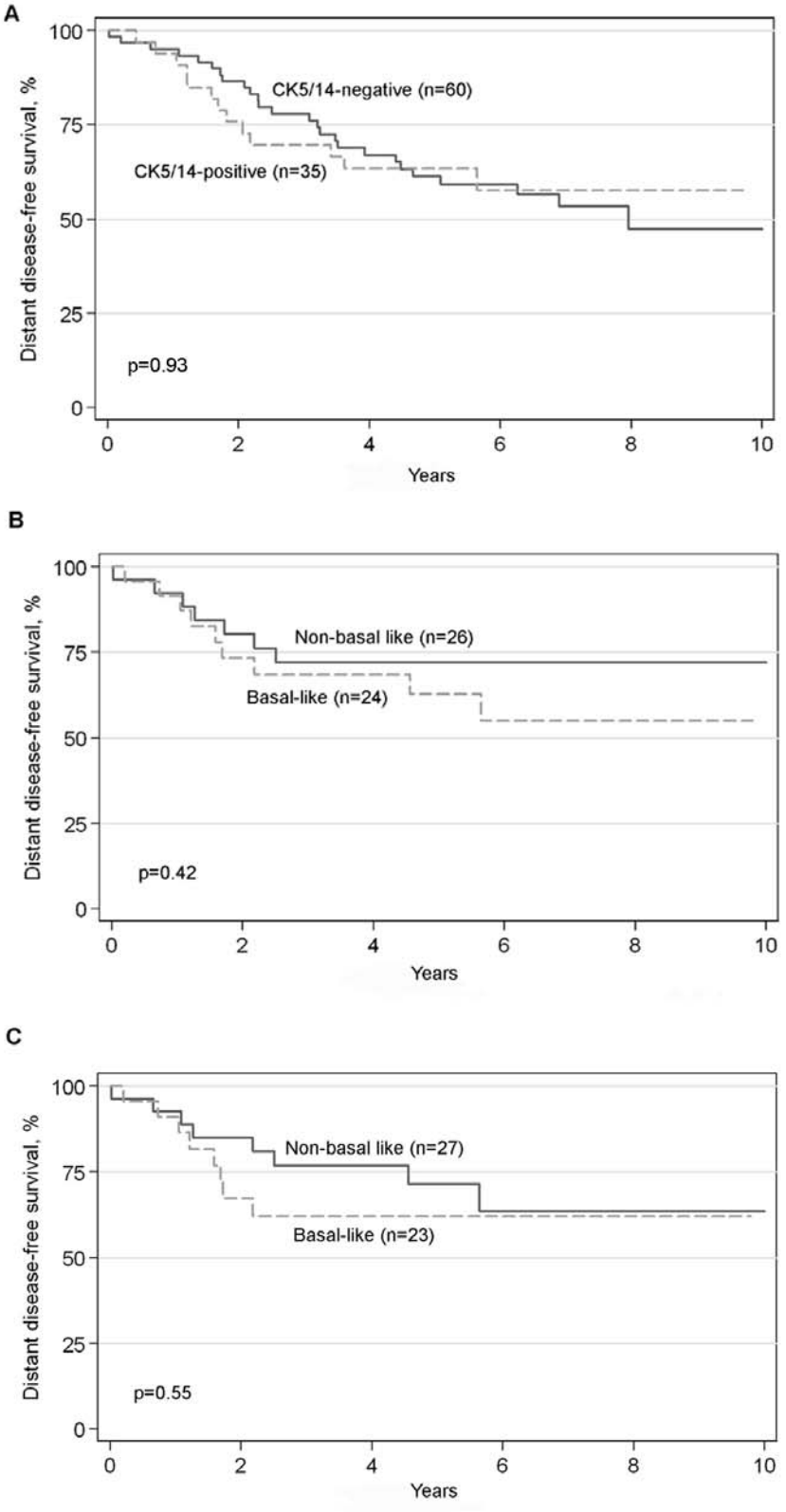

Figure 4. Distant disease-free survival of basal-like and non-basal like tumors within the ER-negative tumor entity. The basal phenotype was defined by using immunohistochemistry (A), cDNA microarray and the top 500 gene set for the basal cytokeratin immunopositive tumors (B) or cDNA microarray and the intrinsic gene set by Sorlie et al (C). There is no difference in survival between basal-like and non-basal like tumors within the ER-negative tumor subgroup.

Functional analysis of genes aberrantly expressed in basal phenotype tumors

We next performed a gene ontology annotation analysis of the top 1,000 genes on our basal gene list (within ER-negative tumors), and found that 823 genes were associated with a functional gene annotation category. Three hundred and eighty-three of these genes were up-regulated in the CK5/14-positive tumors and 440 were down-regulated (Supplemental table 2). Genes up-regulated in basal-like tumors (with EASE score ≤0.05) belonged to the annotation categories: epidermal differentiation (GO:0008544) & ectoderm development (GO:0007398), protein biosynthesis (GO:0006412), nuclear division (GO:0000280), development (GO:0007275), biosynthesis (GO:0009058), histogenesis (GO:0009888), macromolecule biosynthesis (GO:0009059), and M phase (GO:0000279). Basal cytokeratins 14 and 17 were present in the gene category of epidermal differentiation & ectoderm development, which was the most significantly up-regulated biological process in the basal phenotype tumors. Genes down-regulated in basal phenotype tumors were characterized to have functions in cell surface receptor linked signal transduction (GO:0007166), enzyme linked receptor protein signaling pathway (GO:0007167), transmembrane receptor protein tyrosine kinase signaling pathway (GO:0007169), and regulation of G-protein coupled receptor protein signaling pathway (GO:0008277).

DISCUSSION

Basal-like breast cancer has been associated with poor prognosis in several immunohistochemical [10,11,15-18,20,22-25] and gene expression microarray-based studies [7-9]. Still, there are conflicting results between studies about the independent prognostic significance of the basal phenotype [11,15,18,20]. Adjuvant chemotherapy could be recognized as one possible confounding factor, since it has been postulated that basal-like and non-basal tumors would respond differently to chemotherapy [37]. Our results showed that when using immunohistochemistry to identify basal-like tumors, a survival difference was seen in the entire patient population during the first years of the follow-up. This suggests that basal cytokeratin expression predicts early relapse when compared to non-basal tumors including both ER-positive and ER-negative breast cancers. This is in agreement with earlier results [11,15-18,20,22-25]. Furthermore, our tumor series represents early-stage disease not treated with chemotherapy. Thus it presents a more coherent data of breast cancer natural biology than when studying chemotherapy treated patients. It must still be noted that in this study all the patients were treated with tamoxifen for two years, which most likely has affected the natural history of the ER-positive tumors.

Even if we saw survival difference between basal and non-basal tumors when studying the whole population, this was not true within the estrogen receptor negative tumor subgroup. This thus suggests that basal cytokeratin expression is not an independent prognostic factor. Our results support the findings of Potemski et al. [18] and Malzahn et al. [15], who did not find any difference between basal and non-basal tumor survival within ER-negative tumor entity. However, Abd El-Rehim et al. [11] and Rakha et al. [20] have suggested that adjustment to steroid hormone receptor expression would not alter the adverse survival impact of basal phenotype in breast cancer. In our study the lack of prognostic association was not due to the method of tumor classification, since the same result was obtained within ER-negative subgroup when basal-like tumors were identified either by IHC or by two different microarray based classifications. These results are in agreement with the earlier microarray based prognostic studies, which indicate that tumors with basal-like gene expression signature have similar prognosis as the ERBB2 cluster [7-9]. It is concluded that all ER-negative tumors can be classified as having a relatively poor prognosis, irrespective of the cytokeratin composition or gene expression signature.

Studies of basal-like breast cancer are likely to be influenced by the estrogen receptor status, which is a central factor determining both prognosis and

the gene expression patterns [1,2,5,6]. In order to study the basal phenotype breast cancer more specifically without the influence of ER status, we performed a gene expression microarray study for ER-negative breast cancers. This enabled us to look more specifically gene expression profile and clinical behavior of the basal phenotype tumors when the impact of information already included in the ER status is excluded. We were able to separate two tumor clusters, the basal-like and the non-basal like, by using a gene set generated for the basal cytokeratin immunopositive tumors. The unique gene expression profile found for the CK5/14 immunopositive tumors within the ER-negative tumor entity implicates that the basal-like expression profile differed significantly from the rest of the ER-negative tumors and that this tumor subgroup is biologically distinct not only in the general breast cancer population but also within ER-negative tumor entity.

Our CK5/14-associated gene signature identified basal-like tumors within the ER-negative tumor entity very similarly to the clustering with the intrinsic gene set by Sorlie et al. [7]. Whereas all except one of the CK5/14-positive tumors were classified to the basal-like cluster with our CK5/14-associated genes, four tumors with CK5/14-positive immunophenotype were found in the non-basal like cluster with Sorlie's intrinsic gene set. This indicates that our top 500 ranked basal genes were better classifiers for CK5/14 IHC status than Sorlie's intrinsic gene set. This is not surprising given that our basal gene list was generated for this given purpose and from this very material. Interestingly, all seven CK5/14-negative tumors categorized into the basal-like cluster by our basal-associated genes were also found in the basal-like tumor subgroup when performing the analysis using the intrinsic gene set as defined by Sorlie et al. Hence, for these seven cases the two microarray analysis-based classifiers agreed on the basal-like status but disagreed with the CK5/14 immunostaining. To verify that these tumors had not been misclassified with regard to basal-like status when using TMAs, we immunostained the entire tumor sections of five out of these tumors. Two of these tumors were scored as CK5/14 positive in entire sections indicating that the TMA sampling technique (using tissue cores with 0.6 mm diameter) leads to misclassification of some basal-like tumors in immunohistochemistry. Expression of basal cytokeratins often shows a high degree of intratumoral heterogeneity [22], which is likely to explain differences obtained between TMAs and entire tissue sections. However, even when performed on entire tumor sections, CK5/14 immunohistochemistry may not recognize all basal-like subtype of breast cancer as defined by gene expression profiles. Despite the fact that our gene expression signature was generated to be specifically associated with CK5/14 positivity, it clearly also recognizes a distinct set of CK5/14 negative tumors. It has been suggested previously that the basal-like tumor type cluster is most optimally identified by immunohistochemistry when using a combination of positive CK5/6 and/or EGFR, and negative ER and HER-2 staining results as classification criteria [23,38]. Also vimentin and c-kit, which have been shown to associate to basal cytokeratin immunopositivity along with EGFR [22,39], have been recognized as good discriminators for basal-like expression profile [23,38]. The basal cytokeratin negative tumors that clustered with the basal-like cluster in this study could be EGFR, vimentin, and/or c-kit expressing tumors with similar gene expression signature as basal cytokeratin immunopositive breast cancers. It is concluded that immunohistochemically basal cytokeratin positive tumors belong almost always to the basal-like gene expression profile, but this cluster includes also basal cytokeratin negative tumors. Neither immunohistochemical nor microarray based classification of breast cancers into basal or non-basal subgroup is currently not considered justified in the clinics, since direct predictive or prognostic implications are lacking. This could change in the future if differential treatment responsiveness can be confirmed or if treatments targeting specifically basal-like tumors are developed.

In addition to prognostic assessments,

the microarray-based gene data may be more relevant for revealing biological basis of the basal-like tumor classification. For example, the first genes in the gene list generated for the immunohistochemically predefined CK5/14-positive and ER-negative tumors included some genes like *XBP1* and *TTF1* that are known to positively associate with ER status [1,2,6]. These genes had significantly lower expression in the basal-like than in the non-basal like tumors within ER-negative tumor subgroup. It is thus possible that there are some differences in the hormone-independence of the basal-like and non-basal like tumors within ER-negative tumor subgroup. In addition to ER-negativity and poor response to hormone treatment most basal-like tumors are *HER-2* non-amplified. Thus, there are currently no targeted treatment options available for basal-like breast cancers. Our finding that top signature genes such as *EVA1* (rank 11 and 36), *SLC2A1* (rank 42 and 179), *CEACAM1* (rank 148), which are highly expressed in basal-like tumors and are localized to the cell membrane could serve as interesting targets for new drug developments, similar to the HER-2 oncoprotein in tumors with *ERBB2* gene amplification.

To study the biology of basal-like tumors in more detail and to evaluate the function of the genes found associated to this tumor subtype we next found out which biological processes are enriched in basal-like tumor and used EASE for this purpose. We found that the signature for basal-like tumors was most significantly enriched for genes associated with epidermal differentiation and included CK14 and CK17 genes. Both of these cytokeratins are close partners of CK5 [40] and have been shown to be expressed in basal phenotype tumors both by immunohistochemistry [11,12,17,20] and gene expression microarrays [6,7]. We did not use CK17 in the immunohistochemical determination of basal cytokeratin expression as we have earlier shown there is only very few tumors showing CK17 expression in the absence of CK5 and/or CK14 [12]. The biological process of epidermal differentiation may reflect the basal phenotype tumor origin. It has been suggested that a CK5/14-positive breast progenitor cell able to differentiate to both luminal and myoepithelial cells of normal breast would be the transformed cell in basal phenotype breast cancer [41,42]. If these cells represent the so called cancer stem cell for basal phenotype breast cancer, the tumor cells may have the same ability to differentiate as the cell of origin. The biological process of development was fourth in the ranking list and included *EVA1* gene, which was earlier recognized in the basal gene list (rank 11 and 36) as a membrane protein. Other gene ontology terms enriched in the basal-like gene signature, such as protein and macromolecular biosynthesis, nuclear division, and M phase, were indicative of a high proliferation rate. Previous studies have also associated the basal-like subgroup with a high expression of genes involved in proliferation [14,22] and our results suggest that this is true even when compared to the other subgroups, such as the *HER-2* amplified, within the ER-negative entity.

CONCLUSION

Basal cytokeratin immunopositivity predicts early breast cancer relapse and these tumors differ from other ER-negative breast cancer biologically since they have a distinct gene expression profile. Despite of this, the basal cytokeratin expressing tumors show similar prognosis as non-basal ER-negative tumors. Regarding classification, immunohistochemically basal cytokeratin positive tumors show almost always basal-like gene expression signature. We were able to identify several immunohistochemically basal cytokeratin negative tumors, which have a similar gene expression profile as the basal cytokeratin immunopositive breast cancers.

ABBREVIATIONS

BASE, BioArray Software Environment; CISH, chromogenic in situ hybridization; CK, cytokeratin; EASE, expression analysis systematic explorer; FDR, false discovery rate; ER, estrogen receptor; IHC, immunohistochemistry; TMA, tissue microarray

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

MJ performed and analyzed IHC and CISH stainings from the TMAs and drafted and finalized the manuscript. SG performed and analyzed the microarrays and helped drafting the manuscript. PK (Päivikki Kauraniemi) helped with the interpretation of the results and manuscript drafting. MT helped with the manuscript finalization. PB performed the statistics for the tables and figures. MK did the analysis of the membrane association of the genes. PK (Pasi Kataja) performed the scanning of the slides for virtual microscopy and ML (Mikael Lundin) prepared the final virtual slides to the internet. ÅB and MF coordinated the study on their behalf. JI coordinated the study and helped to draft and finalize the manuscript.

ACKNOWLEDGEMENTS

We are grateful of the South Sweden Breast Cancer Group for providing us with the clinical follow-up data and to the participating departments for providing us the samples. The authors want to thank Ms. Sari Toivola, Ms. Eeva Riikonen, Ms. Ritva Kujala, Ms. Helvi Salmela, Ms. Pirjo Pekkala, and Ms. Päivi Kärki for skillful technical assistance. This study was financially supported by grants from the Pirkanmaa Hospital District Research Foundation, Medical Research Fund of Seinäjoki Central Hospital, Swedish Cancer Society, Swedish Research Council, Sigrid Juselius Foundation, Algol-Award, Oy Eli Lilly Finland Ab, and the Finnish Cancer Foundation.

REFERENCES

1. Gruvberger S, Ringnér M, Chen Y, Panavally S, Saal LH, Borg Å, Fernö M, Peterson C, Meltzer PS. **Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression pattern.** Cancer Res 2001,**61**:5979-5984.

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

3. van't Veer LT, Dai H, van de Vijver MJ. **Gene expression profiling predicts clinical outcome of breast cancer.** Nature 2002,**415**:530-536.

4. Pusztai L, Ayers M, Stec J, Clark E, Hess K, Stivers D, Damokosh A, Sneige N, Buchholz TA, Esteva FJ et al. **Gene expression profiles obtained from fine-needle aspiration of breast cancer reliably identify routine prognostic markers and reveal large-scale molecular differences between estrogen-negative and estrogen-positive tumors.** Clin Cancer Res 2003,**9**:2406-2415.

5. Gruvberger SK, Ringnér M, Edén P, Borg Å, Fernö M, Peterson C, Meltzer PS. **Expression profiling to predict outcome in breast cancer: the influence of sample selection.** Breast Cancer Res 2002,**5**:23-26.

6. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA et al. **Molecular portraits of human breast tumours.** Nature 2000,**406**:747-752.

7. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS et al. **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.** Proc Natl Acad Sci USA 2001,**98**:10869-10874.

8. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S et al. **Repeated observation of breast tumor subtypes in independent gene expression data sets.** Proc Natl Acad Sci USA 2003,**100**:8418-8423.

9. Sotiriou C, Neo S-Y, McShane LM, Korn EL, Long PM, Jazaeri A, Martiat P, Fox SB, Harris AL, Liu ET. **Breast cancer classification and prognosis based on gene expression profiles from a population-based study.** Proc Natl Acad Sci USA 2003,**100**:10393-10398.

10. Dairkee SH, Ljung BM. Smith H, Hackett A. **Immunolocalization of a human basal epithelium specific keratin in benign and malignant breast disease.** Breast Cancer Research and Treatment 1987,**10**:11-20.

11. Abd El-Rehim DM, Pinder SE, Paish CE, Bell J,

Blamey RW, Robertson JFR, Nicholson RI, Ellis IO. **Expression of luminal and basal cytokeratins in human breast cancer.** J Pathol 2004,**203**:661-671.

12. Laakso M, Loman N, Borg Å, Isola J. **CK5/14-positive breast cancer: true basal phenotype confined to BRCA1 tumors.** Mod Pathol 2005,**18**:1321-1328.

13. Wetzels RHW, Kuijpers HJH, Lane EB, Leigh IM, Troyanovsky SM, Holland R, van Haelst UJGM, Ramaekers FCS. **Basal cell-specific and hyperproliferation-related keratins in human breast cancer.** Am J Pathol 1991,**138**:751-763.

14. Korsching E, Packeisen J, Agelopoulos K, Eisenacher M, Voss R, Isola J, van Diest PJ, Brandt B, Boecker W, Buerger H. **Cytogenetic alterations and cytokeratin expression patterns in breast cancer: Integrating a new model of breast differentiation into cytogenetic pathways of breast carcinogenesis.** Lab Invest 2002,**82**:1525-1533.

15. Malzahn K, Mitze M, Thoenes M, Moll R. **Biological and prognostic significance of stratified epithelial cytokeratins in infiltrating ductal breast carcinomas.** Virchows Arch 1998,**433**:119-129.

16. Banerjee S, Reis-Filho JS, Ashley S, Steele D, Ashworth A, Lakhani SR, Smith IE. **Basal-like breast carcinomas: clinical outcome and response to chemotherapy.** J Clin Pathology 2006,**59**:729-735.

17. van de Rijn M, Perou CM, Tibshirani R, Haas P, Kallioniemi O, Kononen J, Torhorst J, Sauter G, Zuber M, Köchli OR et al. **Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome.** Am J Pathol 2002,**161**:1991-1996.

18. Potemski P, Kusinska R, Watala C, Pluciennik E, Bednarek AK, Kordek R. **Prognostic relevance of basal cytokeratin expression in operable breast cancer.** Oncology 2005,**69**:478-485.

19. Reis-Filho JS, Simpson PT, Martins A, Preto A, Gaertner F, Schmitt FC. **Distribution of p63, cytokeratin 5/6 and cytokeratin 14 in 51 normal and 400 neoplastic human tissue samples using TARP-4 multi-tumor tissue microarray.** Virchows Arch 2003,**443**:122-132.

20. Rakha EA, Putti TC, Abd El-Rehim DM, Paish C, Green AR, Powe DG, Lee AH, Robertson JF, Ellis IO. **Morphological and immunophenotypical analysis of breast carcinomas with basal and myoepithelial differentiation.** J Pathol 2006,**208**:495-506.

21. Birmbaum D, Bertucci F, Ginestier C, Tagett R, Jacquemier J, Charafe-Jauffret E. **Basal and luminal breast cancer: Basic or luminous?** Int J Oncol 2004,**25**:249-258.

22. Laakso M, Tanner M, Nilsson J, Viklund T, Erikstein B, Kellokumpu-Lehtinen P, Bergh J, Isola J. **Basoluminal carcinoma: a new biologically and prognostically distinct entity between basal and luminal breast cancer.** Clin Cancer Res 2006,**12**:4185-4191.

23. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, Hernandez-Boussard T, Livasy C, Cowan D, Dressler L et al. **Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma.** Clin Cancer Res 2004,**10**:5367-5374.

24. Makretsov NA, Huntsman DG, Nielsen TO, Yorida E, Peacock M, Cheang MCU, Dunn SE, Hayes M, van de Rijn M, Bajdik C, Gilks CB. **Hierarchical clustering analysis of tissue microarray immunostaining data identifies prognostically significant groups of breast carcinoma.** Clin Cancer Res 2004,**10**:6143-6151.

25. Rodríquez-Pinilla SM, Sarrío D, Honrado E, Hardisson D, Calero F, Benitez J, Palacios J. **Prognostic significance of basal-like phenotype and fascin expression in node-negative invasive carcinomas.** Clin Cancer Res 2006,**12**:1533-1539.

26. Chebil G, Bendahl P-O, Idvall I, Fernö M. **Comparison of immunohistochemical and biochemical assay of steroid receptors in primary breast cancer.** Acta Oncologica 2003,**42**:719-725.

27. Isola J, Tanner M, Forsyth A, Cooke TG, Watters AD, Bartlett JM. **Interlaboratory comparison of HER-2 oncogene amplification as detected by chromogenic and fluorescence in situ hybridization.** Clin Cancer Res 2004,**10**:4793-4798.

28. Lundin M, Lundin J, Helin H, Isola J. **A digital atlas of breast histopathology: an application of web based virtual microscopy.** J Clin Pathol 2004,**57**:1288-1291.

29. Khan J, Saal LH, Bittner ML, Jiang Y, Gooden GC, Glatfelter AA, Meltzer PS. **Gene expression profiling in cancer using cDNA microarrays.** Methods Mol Med 2002,**68**:205-222.

30. Saal LH, Troein C, Vallon-Christerson J, Gruvberger S, Borg A, Peterson C. **BioArray Software Environment (BASE): a platform for comprehensive management and analysis of microarray data.** Genome Biol 2002,**3**:SOFTWARE0003.

31. Yang YH, Duboit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP. **Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation.** Nucleic Acids Res 2002,**30**:e15.

32. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA et al. **Molecular classification of cancer: class discovery and class prediction by gene expression monitoring.** Science

1999,**286**:531-537.

33. Eisen MB, Spellman PT, Brown PO, Botstein D. **Cluster analysis and display of genome-wide expression patterns.** Proc Natl Acad Sci USA 1998,**351**:1451-1467.

34. Diehn M, Eisen MB, Botstein D, Brown PO. **Large-scale identification of secreted and membrane associated gene products using DNA microarrays.** Nature genetics 2000,**25**:58-62.

35. Stitziel NO, Mar BG, Liang J, Westbrook CA. **Membrane-associated and secreted genes in breast cancer.** Cancer Res 2004,**64**:8682-8687.

36. Bouton CM, Pevsner J. **DRAGON: Database referencing of array genes online.** Bioinformatics 2000,**16**:1038-1039.

37. Rouzier R, Perou CM, Symmans WF, Ibrahim N, Cristofanilli M, Anderson K, Hess KR, Stec J, Ayers M, Wagner P et al. **Breast cancer molecular subtypes respond differently to preoperative chemotherapy.** Clin Cancer Res 2005, **11**:5678-5685.

38. Livasy CA, Karaca G, Nanda R, Tretiakova MS, Olopade OI, Moore DT, Perou CM. **Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma.** Mod Pathol 2006,**19**:264-271.

39. Korsching E, Packeisen J, Liedtke C, Hungermann D, Wulfing P, van Diest PJ, Brandt P, Boecker W, Buerger H. **The origin of vimentin expression in invasive breast cancer: epithelial-mesenchymal transition, myoepithelial histogenesis or histogenesis from progenitor cells with bilinear differentiation potential?** J Pathol 2005,**206**:451-457.

40. Moll R, Franke WW, Schiller DL. **The catalog of human cytokeratins: Patterns of expression in normal epithelia, tumors and cultured cells.** Cell 1982,**31**:11-24.

41. Boecker W, Moll R, Poremba C, Holland R, van Diest PJ, Dervan P, Buerger H, Wai D, Diallo RI, Brandt B et al. **Common adult stem cells in the human breast give rise to glandular and myoepithelial cell lineages: A new cell biological concept.** Lab Invest 2002,**82**:737-745.

42. Boecker W, Buerger H. **Evidence of progenitor cells of glandular and myoepithelial cell lineages in the human adult female breast epithelium: a new progenitor (adult stem) cell concept.** Cell Prolif 2003,**36(suppl 1)**:73-84.

ADDITIONAL FILES

Supplemental figure 1

Heat map of 50 ER-negative tumors based on the top 500 gene set generated for the CK5/14-positive tumors. Yellow indicates the basal-like cluster and black non-basal like cluster.

Supplemental figure 2

Heat map of 50 ER-negative tumors based on the intrinsic gene set by Sorlie et al. Yellow indicates the basal-like cluster and black non-basal like cluster.

Supplemental table 1

The top 200 genes list generated for the immunohistochemically CK5/14-positive ER-negative breast cancers. The membrane association is defined.

Supplemental table 2

The results of gene ontology annotation analysis of the top 1,000 basal genes.

The Supplemental figures 1 and 2 are available on website since their size does not allow printing.

http://acta.uta.fi/pdf/kuvat/JumppanenSupplementalfigure_1.pdf http://acta.uta.fi/pdf/kuvat/JumppanenSupplementalfigure_2.pdf

Supplemental table 1

Rank Golub	Sign in Basal positive tumors	Golub score	Gene Name	Gene Symbol	Unigene ClusterID	IMAGE cloneID	Cytoband	Sequence	OMIM	Stitziel	Diehn	SwissProt
1	–	0,777952716	EST	N/A	N/A	1473914	N/A	N/A	N/A	N/A	N/A	N/A
2	–	0,76947476	CDNA FLJ31660 fis, clone NT2RI2004410	N/A	Hs.4749	845037	N/A	AK124576	N/A	No	No	No
3	+	0,763709901	Adenosylmethionine decarboxylase 1	AMD1	Hs.159118	149013	6q21-q22	NM_001634	180980	No	No	No
4	–	0,752787604	ATP-binding cassette, sub-family C (CFTR/MRP), member 6	ABCC6	Hs.442182	453988	16p13.1	NM_001171	N/A	No	No	1>Integral membrane protein (By similarity).
5	+	0,74592012	Jerky homolog-like (mouse)	JRKL	Hs.105940	1608898	11q21	NM_003772	603211	No	No	No
6	–	0,742142015	Melanophilin	MLPH	Hs.102406	1558642	2q37.3	NM_024101	606526	No	No	No
7	–	0,741949388	T-box 3 (ulnar mammary syndrome)	TBX3	Hs.129895	397488	12q24.1	NM_016569	601621	No	No	No
8	–	0,737282995	K+ channel tetramerization protein	GMRP-1	Hs.332382	79632	11p15.2	NM_032320	N/A	No	No	No
9	–	0,731379373	KIAA0310	KIAA0310	Hs.522500	809944	9q34.3	AB002308	N/A	No	No	No
10	+	0,721946171	Adenosylmethionine decarboxylase 1	AMD1	Hs.159118	149013	6q21-q22	NM_001634	180980	No	No	No
11	+	0,712813859	Epithelial V-like antigen 1	EVA1	Hs.116651	233464	11q24	NM_005797	604873	No	No	1>Type I membrane protein (Probable).
12	–	0,712660506	Peroxisome proliferative activated receptor, alpha	LOC150383	Hs.103110	81050	22q13.31	NM_207327	N/A	No	No	No
13	–	0,710981612	Trefoil factor 1 (breast cancer, estrogen-inducible sequence expressed in)	TFF1	Hs.162807	2504927	21q22.3	NM_003225	113710	No	No	1>Secreted.
14	–	0,707832083	EST	N/A	N/A	1616181	N/A	N/A	N/A	N/A	N/A	N/A
15	–	0,701241421	Bone morphogenetic protein 4	BMP4	Hs.68879	797048	14q22-q23	NM_001202	112262	No	No	No
16	–	0,700709265	X-box binding protein 1	XBP1	Hs.437638	417867	22q12.1	NM_005080	194355	No	No	No
17	+	0,700155803	EST	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
18	+	0,698750434	Homo sapiens, clone IMAGE:4337652, mRNA	N/A	Hs.381985	773204	N/A	BC018676	N/A	No	No	No
19	+	0,697654774	XM_011453 p53-induced protein PIGPC1 (PIGPC1)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

20	–	0,689095751	Xeroderma pigmentosum, complementation group A	XPA	Hs.288867	788141	9q22.3	NM_000380	278700	No	No	No
21	–	0,688924871	RAB17, member RAS oncogene family	RAB17	Hs.44278	294537	2q37.3	NM_022449	N/A	No	No	No
22	+	0,687912323	Myosin, light polypeptide 4, alkali; atrial, embryonic	MYL4	Hs.463300	454341	17q21-qter	NM_002476	160770	No	No	No
23	+	0,6861361	Cirrhosis, autosomal recessive 1A (cirhin)	CIRH1A	Hs.461113	377253	16q22.1	NM_032830	607456	No	No	No
24	–	0,683023554	EST	N/A	N/A	86035	N/A	N/A	N/A	N/A	N/A	N/A
25	+	0,682467596	AL023582 Human DNA sequence from clone 496H19 on chromosome 6q24	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
26	+	0,682072424	Crystallin, alpha B	CRYAB	Hs.408767	839736	11q22.3-q23.	NM_001885	123590	No	No	No
27	+	0,678555177	Comparative gene identification transcript 37	CGI-37	Hs.501513	1603408	16q22.1	NM_016101	N/A	No	No	No
28	–	0,675413191	Solute carrier family 13 (sodium-dependent citrate transporter), member 5	SLC13A5	Hs.399496	84079	17p13.1	NM_177550	608305	No	No	No
29	–	0,674498454	EST	N/A	N/A	417393	N/A	N/A	N/A	N/A	N/A	N/A
30	–	0,672095135	Sel-1 suppressor of lin-12-like (C. elegans)	SEL1L	Hs.181300	81578	14q24.3-q31	NM_005065	602329	Yes	No	1>Integral membrane protein (Potential).
31	–	0,670999593	AL021391 RP1-102D24	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
32	–	0,664222374	Adenosine deaminase, RNA-specific	ADAR	Hs.12341	23114	1q21.1-q21.2	NM_001111	601059	No	No	No
33	+	0,663736207	Kallikrein 6 (neurosin, zyme)	KLK6	Hs.79361	809784	19q13.3	NM_002774	602652	No	No	1>Secreted.
34	+	0,663498018	COX4 neighbor	NOC4	Hs.173162	625683	16q24	NM_006067	604886	No	No	No
35	+	0,662889068	Phosphofructokinase, platelet	PFKP	Hs.26010	950682	10p15.3-p15.1	NM_002627	171840	Yes	Yes	No
36	+	0,661724598	Epithelial V-like antigen 1	EVA1	Hs.116651	453112	11q24	NM_005797	604873	No	No	1>Type I membrane protein (Probable).
37	+	0,66092549	Chromosome 1 open reading frame 163	FLJ12439	Hs.349905	2107440	1p32.3	NM_023077	N/A	No	No	No
38	+	0,659300768	Comparative gene identification transcript 37	CGI-37	Hs.501513	1461604	16q22.1	NM_016101	N/A	No	No	No
39	+	0,655870958	BC006481 tubulin alpha 1(microtubule)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
40	–	0,653004258	Chromosome 17 open reading frame 28	C17orf28	Hs.11067	79726	17q25.1	NM_030630	605752	No	No	No
41	+	0,650637776	Small fragment nuclease	DKFZP566E1	Hs.7527	810284	11q23.1-q23.2	NM_015523	607149	No	No	No
42	+	0,645601037	Solute carrier family 2 (facilitated glucose transporter), member 1	SLC2A1	Hs.473721	207358	1p35-p31.3	NM_006516	138140	No	No	1>Integral membrane protein. Localizes primarily at the cell surface (By similarity).
43	–	0,644890532	EST	N/A	N/A	950615	N/A	N/A	N/A	N/A	N/A	N/A

44	–	0,639509676	KIAA0882 protein	KIAA0882	Hs.480819	2284924	4q31.21	NM_015130	N/A	No	No	No
45	+	0,637652935	RIO kinase 1 (yeast)	RIOK1	Hs.437474	452863	6p24.3	NM_031480	N/A	No	No	No
46	+	0,637568948	NM_002799 proteasome (prosome macropain) subunit beta type 7(protein degradation)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
47	–	0,630284181	Tumor protein p53 inducible nuclear protein 1	TP53INP1	Hs.492261	814528	8q22	NM_033285	606185	No	No	No
48	+	0,630258695	PAK1 interacting protein 1	PAK1IP1	Hs.310231	1569107	6p24.1	NM_017906	607811	No	No	No
49	+	0,629578817	Neuropilin (NRP) and tolloid (TLL)-like 2	NETO2	Hs.444046	815556	16q11	NM_018092	607974	No	No	No
50	+	0,627452486	Chromosome 4 open reading frame 14	C4orf14	Hs.8715	814443	4q12	NM_032313	N/A	No	No	No
51	+	0,626319884	Tripartite motif-containing 29	TRIM29	Hs.504115	377275	11q22-q23	NM_012101	N/A	No	No	No
52	–	0,625395395	EST	N/A	N/A	399463	N/A	N/A	N/A	N/A	N/A	N/A
53	+	0,625128751	MRNA; cDNA DKFZp761L1121 (from clone DKFZp761L1121)	N/A	Hs.171939	323322	N/A	BQ719686	N/A	No	No	No
54	–	0,624725632	Neural proliferation, differentiation and control, 1	NPDC1	Hs.105547	2017597	9q34.3	NM_015392	605798	Yes	No	No
55	–	0,62350762	V-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)	ERBB2	Hs.446352	783729	17q11.2-q12	NM_004448	164870	No	No	1>Type I membrane protein.
56	+	0,623021721	Tubulin, alpha 1 (testis specific)	TUBA1	Hs.75318	612274	2q35	NM_006000	N/A	No	No	No
57	–	0,619281702	LIM domain binding 3	LDB3	Hs.49998	2067500	10q22.3-q23.2	NM_007078	605906	No	No	No
58	–	0,618812903	Sorting nexin 9	SNX9	Hs.191213	46360	6q25.1-q26	NM_016224	605952	No	No	No
59	–	0,617917103	Xeroderma pigmentosum, complementation group A	XPA	Hs.288867	668262	9q22.3	NM_000380	278700	No	No	No
60	–	0,616601536	Transcribed locus, weakly similar to XP_517454.1 PREDICTED: similar to hypothetical protein MGC45438 [Pan troglodytes]	N/A	Hs.453381	796531	N/A	BU078529	N/A	No	No	No
61	+	0,616492406	ATP-binding cassette, sub-family B (MDR/TAP), member 10	ABCB10	Hs.17614	193990	1q42	NM_012089	605454	No	No	1>Integral membrane protein. Mitochondrial inner membrane.
62	–	0,616403443	Solute carrier family 16 (monocarboxylic acid transporters), member 2	SLC16A2	Hs.75317	773344	Xq13.2	NM_006517	300095	No	No	1>Integral membrane protein. Plasma membrane (Probable).
63	+	0,615240422	AC003984 Human PAC clone RP4-566F10 from 7q21	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
64	–	0,613520733	Zinc finger protein 511	ZNF511	Hs.422113	1569463	10q26.3	NM_145806	N/A	No	No	No
65	+	0,613385325	Huntingtin interacting protein 2	HIP2	Hs.50308	233581	4p14	NM_005339	602846	No	No	No
66	–	0,613042706	Rabaptin, RAB GTPase binding effector protein 1	RABEP1	Hs.551518	2545220	17p13.2	NM_004703	603616	No	No	No
67	+	0,611916998	ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1	ATP5C1	Hs.271135	2448636	10p15.1	NM_005174	108729	No	No	No

68	–	0,611170632	A disintegrin-like and metalloprotease (reprolysin type) with thrombospondin type 1 motif, 7	ADAMTS7	Hs.16441	1609975	15q24.2	NM_014272	605009	No	No	1>Secreted. Associated with the extracellular matrix (By similarity).
69	–	0,611059886	LAG1 longevity assurance homolog 6 (S. cerevisiae)	LASS6	Hs.506829	2312454	2q24.3	NM_203463	N/A	No	Yes	1>Integral membrane protein. Nuclear (Potential).
70	–	0,609968846	Solute carrier family 2 (facilitated glucose transporter), member 10	SLC2A10	Hs.305971	758347	20q13.1	NM_030777	606145	No	No	1>Integral membrane protein.
71	–	0,606882	Pleckstrin homology domain containing, family F (with FYVE domain) member 2	PLEKHF2	Hs.29724	150003	8q22.1	NM_024613	N/A	No	No	No
72	–	0,606839159	Golgi phosphoprotein 2	GOLPH2	Hs.494337	811582	9q21.33	NM_177937	606804	No	No	No
73	–	0,605354109	Casein kinase 1, delta	CSNK1D	Hs.477070	1610490	17q25	NM_001893	600864	No	No	No
74	+	0,602996134	Chromosome 10 open reading frame 7	C10orf7	Hs.412842	784830	10p13	NM_006023	N/A	No	No	No
75	+	0,602794981	P300/CBP-associated factor	PCAF	Hs.533055	296476	3p24	NM_003884	602303	No	No	No
76	+	0,601001872	Histone 1, H4b	HIST1H4B	Hs.143080	1842170	6p21.3	NM_003544	602829	No	No	No
77	+	0,600702194	Zinc finger, C3HC-type containing 1|Centromere protein F, 350/400ka (mitosin)	ZC3HC1|CENI	MULTIPLE: Hs.194	115443	7q32.2|1q32-q	N/A	N/A	No	No	No
78	–	0,600253636	IQ motif and Sec7 domain 1	IQSEC1	Hs.475506	767819	3p25.2	NM_014869	N/A	No	No	No
79	+	0,599983241	EST	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
80	+	0,599059563	Gamma-aminobutyric acid (GABA) A receptor, pi	GABRP	Hs.26225	563598	5q33-q34	NM_014211	602729	No	No	1>Integral membrane protein.
81	–	0,597954298	SH3 domain binding glutamic acid-rich protein like	SH3BGRL	Hs.108029	1603583	Xq13.3	NM_003022	300190	No	No	No
82	–	0,597371902	Ectonucleotide pyrophosphatase/phosphodiesterase 1	ENPP1	Hs.527295	82991	6q22-q23	NM_006208	173335	No	No	1>Type II membrane protein.
83	+	0,596786952	Chromosome 6 open reading frame 66	C6orf66	Hs.512144	2106756	6q16.1	NM_014165	N/A	No	No	No
84	–	0,593336672	Target of myb1-like 2 (chicken)	TOM1L2	Hs.462379	46278	17p11.2	NM_144678	N/A	No	No	No
85	–	0,592959556	EST	N/A	N/A	704261	N/A	N/A	N/A	N/A	N/A	N/A
86	+	0,592250557	D4, zinc and double PHD fingers family 1	DPF1	Hs.466651	2324147	19q13.13-q13	NM_004647	601670	No	No	No
87	+	0,591940759	Plexin domain containing 2	PLXDC2	Hs.549220	812277	10p12.31	NM_032812	606827	No	No	No
88	+	0,591728166	Acetyl-Coenzyme A acyltransferase 2 (mitochondrial 3-oxoacyl-Coenzyme A thiolase)	ACAA2	Hs.200136	45376	18q21.1	NM_006111	604770	No	No	No
89	+	0,59075616	Mitogen-activated protein kinase 14	MAPK14	Hs.485233	430709	6p21.3-p21.2	NM_001315	600289	No	No	No
90	+	0,590261416	ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1	ATP5C1	Hs.271135	845519	10p15.1	NM_005174	108729	No	No	No
91	+	0,589579788	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 13 (GalNAc-T13)	GALNT13	Hs.470277	51842	2q23.3-q24.1	NM_052917	608369	No	No	1>Type II membrane protein. Golgi (By similarity).

92	+	0,588567476	Desmoplakin	DSP	Hs.519873	240961	6p24	NM_004415	125647	No	No	No
93	–	0,584860362	Hypothetical protein FLJ14299	FLJ14299	Hs.288042	788087	8p12	NM_025069	N/A	No	No	No
94	–	0,583742112	Solute carrier family 16 (monocarboxylic acid transporters), member 6	SLC16A6	Hs.42645	266389	17q24.2	NM_004694	603880	No	No	No
95	–	0,582857873	Tuberous sclerosis 1	TSC1	Hs.370854	812042	9q34	NM_000368	605284	No	No	1>Cytoplasmic. At steady state found in association with membranes.
96	+	0,581764584	Schwannomin interacting protein 1	SCHIP1	Hs.134665	506143	3q25.32-q25.:	NM_014575	N/A	No	No	No
97	–	0,580933952	Dual specificity phosphatase 5	DUSP5	Hs.2128	342378	10q25	NM_004419	603069	No	No	No
98	–	0,580463822	Anterior gradient 2 homolog (Xenopus laevis)	AGR2	Hs.530009	2321113	7p21.3	NM_006408	606358	No	No	No
99	–	0,579831398	EST	N/A	N/A	2108257	N/A	N/A	N/A	N/A	N/A	N/A
100	+	0,578333433	AF141348 alpha-tubulin(cytoskeleton)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
101	+	0,577839376	Histone 1, H4c	HIST1H4C	Hs.46423	1461138	6p21.3	NM_003542	602827	No	No	No
102	–	0,577520322	Wolfram syndrome 1 (wolframin)	WFS1	Hs.518602	138189	4p16	NM_006005	606201	No	No	1>Integral membrane protein. Endoplasmic reticulum.
103	+	0,577471721	Chromatin modifying protein 1B	CHMP1.5	Hs.551551	120684	18p11.21	NM_020412	606486	No	No	No
104	+	0,577087178	SMAD, mothers against DPP homolog 4 (Drosophila)	SMAD4	Hs.75862	788421	18q21.1	NM_005359	600993	No	No	No
105	–	0,576924783	Chromosome X open reading frame 45	CXorf45	Hs.443061	113298	Xq23	NM_024810	N/A	No	No	No
106	+	0,574796794	Retinal outer segment membrane protein 1	ROM1	Hs.281564	223098	11q13	NM_000327	180721	No	No	1>Integral membrane protein.
107	+	0,574786598	Solute carrier family 9 (sodium/hydrogen exchanger), isoform 6	SLC9A6	Hs.62185	20115	Xq26.3	NM_006359	300231	Yes	Yes	1>Integral membrane protein. Mitochondrial.
108	+	0,574475141	BC005145 GDP dissociation inhibitor 2(cellular transportation)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
109	+	0,574245299	KIAA1586	KIAA1586	Hs.180663	1034483	6p12.1	NM_020931	N/A	No	No	No
110	–	0,57400275	CAMP responsive element binding protein 1	CREB1	Hs.516646	148444	2q34	NM_004379	123810	No	No	No
111	+	0,573457478	UTP11-like, U3 small nucleolar ribonucleoprotein, (yeast)	CGI-94	Hs.472038	713205	1p34.3	NM_016037	N/A	No	No	No
112	–	0,572838029	EST	N/A	N/A	854593	N/A	N/A	N/A	N/A	N/A	N/A
113	+	0,572275075	Nucleoside phosphorylase	NP	Hs.75514	769890	14q13.1	NM_000270	164050	No	No	No
114	+	0,570714702	GIPC PDZ domain containing family, member 1|DEAD (Asp-Glu-Ala-Asp) box polypeptide 21	RGS19IP1|DD:	MULTIPLE: Hs.645	586895	19p13.1|10q2	N/A	N/A	No	No	No
115	+	0,570681755	BC006481 tubulin alpha 1(microtubule)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

116	–	0,570538407	Similar to CG4502-PA	N/A	Hs.126856	50582	5p15.31	AK096612	N/A	No	No	No
117	–	0,569621114	GATA binding protein 2	GATA2	Hs.367725	135688	3q21.3	NM_032638	137295	No	No	No
118	+	0,569279999	Pellino homolog 1 (Drosophila)	PELI1	Hs.7886	232842	2p13.3	NM_020651	N/A	No	No	No
119	–	0,569074309	AF131785 clone 24994	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
120	–	0,568769372	SMAD, mothers against DPP homolog 3 (Drosophila)	SMAD3	Hs.555881	345935	15q22.33	NM_005902	N/A	No	No	No
121	+	0,568700835	Hypothetical protein LOC51059	LOC51059	Hs.13497	50329	8q24.23	NM_015912	N/A	No	No	No
122	–	0,567384349	EST	N/A	N/A	448117	N/A	N/A	N/A	N/A	N/A	N/A
123	+	0,567305348	EST	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
124	–	0,566822325	Hypothetical protein LOC201895	LOC201895	Hs.205952	289505	4p14	NM_174921	N/A	No	No	No
125	+	0,565768273	BF57544	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
126	+	0,565736422	SEH1-like (S. cerevisiae)|S-adenosylhomocysteine hydrolase	SEH1L|AHCY	MULTIPLE: Hs.301	757206	18p11.21|20c	N/A	N/A	No	No	No
127	+	0,564806623	DnaJ (Hsp40) homolog, subfamily B, member 6	DNAJB6	Hs.490745	757147	7q36.3	NM_005494	N/A	No	No	No
128	–	0,56411147	Chromosome 9 open reading frame 91	C9orf91	Hs.522357	23579	9q32	NM_153045	N/A	No	No	No
129	+	0,56321745	Z79996 clone SC22CB-33F2 on chromosome 22	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
130	+	0,563059069	PAK1 interacting protein 1	PAK1IP1	Hs.310231	132111	6p24.1	NM_017906	607811	No	No	No
131	+	0,561831749	ElaC homolog 1 (E. coli)	ELAC1	Hs.47572	244654	18q21	NM_018696	608079	No	No	No
132	–	0,561576004	FLJ22794 protein	FLJ22794	Hs.150651	814224	11q12.1	NM_198847	N/A	No	No	No
133	+	0,560259663	ATPase, Na+/K+ transporting, beta 3 polypeptide	ATP1B3	Hs.477789	842894	3q23	NM_001679	601867	No	No	1>Type II membrane protein.
134	+	0,560241001	Annexin A8	ANXA8	Hs.524293	666879	10q11.2	BC033221	N/A	No	No	No
135	–	0,560001512	Ubiquitin specific protease 20	USP20	Hs.5452	462431	9q34.11	NM_006676	N/A	No	No	No
136	+	0,559171513	Abhydrolase domain containing 5	ABHD5	Hs.19385	399513	3p21	NM_016006	604780	No	No	No
137	+	0,558931103	Scm-like with four mbt domains 1	SFMBT1	Hs.343679	743382	3p21.1	NM_016329	607319	No	No	No
138	–	0,55765595	Phosphoinositide-3-kinase, class 2, beta polypeptide	PIK3C2B	Hs.497487	1536172	1q32	NM_002646	602838	No	No	1>Found mostly in the microsome, but also in the plasma membrane and cytosol.
139	+	0,556603404	Malate dehydrogenase 1, NAD (soluble)	MDH1	Hs.526521	725188	2p13.3	NM_005917	154200	No	No	No

140	+	0,556272961	Transmembrane 4 L six family member 1	TM4SF1	Hs.351316	840567	3q21-q25	NM_014220	191155	No	No	No
141	–	0,554272596	Acetyl-Coenzyme A synthetase 2 (AMP forming)-like	ACAS2L	Hs.529353	291426	20p11.23-p11	NM_032501	N/A	No	No	No
142	+	0,554227561	AL050139 DKFZp586M141 hypothetical protein FLJ13910	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
143	–	0,553072146	Golgi autoantigen, golgin subfamily a, 1	GOLGA1	Hs.133469	34102	9q33.3	NM_002077	602502	No	No	1>Peripheral membrane protein associated with Golgi stacks.
144	–	0,552521077	Homeo box B2	HOXB2	Hs.514289	1526826	17q21-q22	NM_002145	142967	No	No	No
145	+	0,552479003	NM_006082 K-ALPHA-1 tubulin(cytoskeleton)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
146	+	0,55220833	EST	N/A	N/A	258033	N/A	N/A	N/A	N/A	N/A	N/A
147	+	0,55213385	BC010494 tubulin alpha 1(cytoskeleton)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
148	+	0,551846804	Carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)	CEACAM1	Hs.512682	2384179	19q13.2	NM_001712	109770	No	No	1>Type I membrane protein (isoforms A, B and C). Secreted (isoforms G, H and I).
149	–	0,551829203	Transcribed locus	N/A	Hs.444858	1896335	N/A	BQ942770	N/A	No	No	No
150	+	0,55165916	Golgi associated PDZ and coiled-coil motif containing	GOPC	Hs.191539	814309	6q21	NM_020399	606845	No	No	No
151	–	0,551006286	Hypothetical gene supported by AK075564; BC060873	LOC400451	Hs.27373	2271170	15q26.1	NM_207446	N/A	No	No	No
152	–	0,550839641	Proteasome (prosome, macropain) subunit, beta type, 7	PSMB7	Hs.213470	824150	9q34.11-q34.	NM_002799	604030	No	No	No
153	+	0,550700587	CDNA FLJ34896 fis, clone NT2NE2018180	N/A	Hs.507978	51078	N/A	AK092215	N/A	No	No	No
154	+	0,550605206	AF081484 alpha-tubulin isoform 1(cytoskeleton)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
155	+	0,550515644	Acidic (leucine-rich) nuclear phosphoprotein 32 family, member E	ANP32E	Hs.385913	856388	1q21.2	NM_030920	N/A	No	No	No
156	–	0,548973575	APG12 autophagy 12-like (S. cerevisiae)	APG12L	Hs.264482	2443399	5q21-q22	NM_004707	N/A	No	No	No
157	–	0,54839699	Chromosome 20 open reading frame 18	C20orf18	Hs.247280	200592	20p13	NM_006462	N/A	No	No	No
158	+	0,547665133	Cyclin A2	CCNA2	Hs.85137	950690	4q25-q31	NM_001237	123835	No	No	No
159	+	0,547078106	NM_001494 GDP dissociation inhibitor 2(cellular transportation)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
160	+	0,54627051	THAP domain containing 11	THAP11	Hs.513681	826166	16q22.1	NM_020457	609119	No	No	No
161	+	0,545548576	Inositol(myo)-1(or 4)-monophosphatase 2	IMPA2	Hs.367992	32299	18p11.2	NM_014214	605922	No	No	No
162	+	0,544999173	Alpha-tubulin isotype H2-alpha	H2-ALPHA	Hs.503749	745138	2q21.1	NM_080386	N/A	No	No	No
163	+	0,543835861	Sin3-associated polypeptide, 30kDa	SAP30	Hs.413835	502142	4q34.1	NM_003864	603378	No	No	No

164	–	0,543131806	Ankyrin repeat and SOCS box-containing 6	ASB6	Hs.125037	1621805	N/A	NM_177999	N/A	No	No	No
165	+	0,543115157	EST	N/A	N/A	2470840	N/A	N/A	N/A	N/A	N/A	N/A
166	+	0,542734	Vaccinia related kinase 2	VRK2	Hs.468623	824117	2p16-p15	NM_006296	602169	No	No	No
167	–	0,541835752	CAMP responsive element binding protein 3-like 4	CREB3L4	Hs.372924	814145	1q21.3	NM_130898	607495	No	No	No
168	+	0,540910002	Tumor suppressor candidate 3	TUSC3	Hs.426324	293859	8p22	NM_178234	601385	Yes	No	No
169	+	0,540148318	EST	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
170	–	0,537278593	IQ motif containing E	IQCE	Hs.520627	272100	7p22.3	NM_152558	N/A	No	No	No
171	–	0,537218925	SATB family member 2	SATB2	Hs.516617	1642145	2q33	NM_015265	608148	No	No	No
172	–	0,53643313	Myelin expression factor 2	MYEF2	Hs.6638	251875	15q21.1	NM_016132	N/A	No	No	No
173	–	0,536268137	Bradykinin receptor B2	BDKRB2	Hs.525572	665674	14q32.1-q32.2	NM_000623	113503	No	No	No
174	–	0,536221937	Solute carrier family 16 (monocarboxylic acid transporters), member 2	SLC16A2	Hs.75317	773344	Xq13.2	NM_006517	300095	No	No	1>Integral membrane protein. Plasma membrane (Probable).
175	–	0,534813241	Neuro-oncological ventral antigen 1	NOVA1	Hs.211225	2015354	14q	NM_002515	602157	No	No	No
176	+	0,534595027	NM_001494 GDP dissociation inhibitor 2(cellular transportation)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
177	+	0,534492537	NM_006082 K-ALPHA-1 tubulin(cytoskeleton)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
178	+	0,534206214	Tigger transposable element derived 2	TIGD2	Hs.58924	416361	4q22.1	NM_145715	N/A	No	No	No
179	+	0,534116315	Solute carrier family 2 (facilitated glucose transporter), member 1	SLC2A1	Hs.473721	25389	1p35-p31.3	NM_006516	138140	No	No	1>Integral membrane protein. Localizes primarily at the cell surface (By similarity).
180	+	0,533049266	NM_006082 K-ALPHA-1 tubulin(cytoskeleton)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
181	–	0,532824529	Ankyrin repeat domain 40	MGC15396	Hs.463426	244277	17q21.33	NM_052855	N/A	No	No	No
182	+	0,532407316	GDP dissociation inhibitor 2	GDI2	Hs.299055	197176	10p15	NM_001494	600767	No	No	No
183	–	0,53230559	Trefoil factor 1 (breast cancer, estrogen-inducible sequence expressed in)	TFF1	Hs.162807	2223790	21q22.3	NM_003225	113710	No	No	1>Secreted.
184	+	0,531413153	LSM6 homolog, U6 small nuclear RNA associated (S. cerevisiae)	LSM6	Hs.190520	397638	4q31.22	NM_007080	607286	No	No	No
185	–	0,530988835	WW domain containing E3 ubiquitin protein ligase 1	WWP1	Hs.533440	141959	8q21	NM_007013	602307	No	No	No
186	+	0,530750964	Uncharacterized bone marrow protein BM039	BM039	Hs.283532	448036	16q23.2	NM_018455	N/A	No	No	No
187	–	0,530536358	Ral GEF with PH domain and SH3 binding motif 2	RALGPS2	Hs.496222	731376	1q25.2	NM_018037	603874	No	No	No

188	+	0,52863254	Nurim (nuclear envelope membrane protein)	NRM	Hs.519993	813318	6p21.33	NM_007243	N/A	No	No	No
189	–	0,527686791	Regulator of G-protein signalling 11	RGS11	Hs.65756	1500542	16p13.3	NM_183337	603895	No	No	No
190	+	0,527531058	EST	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
191	+	0,526359391	Histone 1, H4c	HIST1H4C	Hs.46423	1461138	6p21.3	NM_003542	602827	No	No	No
192	–	0,525972628	KIAA1533	KIAA1533	Hs.515351	843054	19q13.11	NM_020895	N/A	No	No	No
193	+	0,525615861	Epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)	EGFR	Hs.488293	324861	7p12	NM_005228	131550	No	No	1>Type I membrane protein. Isoform 2 is secreted.
194	–	0,52541344	EST	N/A	N/A	2072955	N/A	N/A	N/A	N/A	N/A	N/A
195	–	0,525306453	Nuclear receptor subfamily 2, group F, member 1	NR2F1	Hs.519445	2014382	5q14	NM_005654	132890	No	No	No
196	–	0,525295531	Activin A receptor, type IB	ACVR1B	Hs.438918	730768	12q13	NM_004302	601300	No	No	1>Type I membrane protein.
197	+	0,525160344	PAP associated domain containing 1	PAPD1	Hs.173946	129345	10p11.23	NM_018109	N/A	No	No	No
198	–	0,524972277	Cytoplasmic polyadenylation element binding protein 4	CPEB4	Hs.127126	280985	5q21	NM_030627	N/A	No	No	No
199	+	0,523903044	Eukaryotic translation initiation factor 3, subunit 6 48kDa	EIF3S6	Hs.405590	2568696	8q22-q23	NM_001568	602210	No	No	No
200	–	0,523601249	Egl nine homolog 2 (C. elegans)	EGLN2	Hs.515417	824419	19q13.2	NM_017555	606424	No	No	No

Supplemental table 2

	Gene Category	EASE score	Gene Rank	Gene Name	Gene Symbol	Unigene ID
Up-regulated in CK5/14+	Epidermal differentiation & Ectoderm development	0.007	329	Keratin 17	KRT17	HS.2785
	Biological Process GO:0008544 & GO:0007398		868	Keratin 14 (epidermolysis bullosa simplex, Dowling-Meara, Koebne	KRT14	HS.355214
			828	Fatty acid binding protein 5 (psoriasis-associated)	FABP5	HS.408061
			92	Desmoplakin	DSP	HS.519873
			664	Procollagen-lysine 1, 2-oxoglutarate 5-dioxygenase (lysine hydroxy	PLOD	HS.75093
	Protein biosynthesis	0.027	276	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 5	B3GNT5	HS.257222
	Biological Process GO:0006412		870	Alanyl-tRNA synthetase	AARS	HS.315137
			778	Ribosomal protein S5	RPS5	HS.378103
			865	Eukaryotic translation initiation factor (eIF) 2A	eIF2A	HS.378808
			333	Mitochondrial ribosomal protein S10	MRPS10	HS.380887
			442	Eukaryotic translation initiation factor 3, subunit 6 48kDa	EIF3S6	HS.405590
			934	Ribosomal protein S10	RPS10	HS.406620
			615	Prostaglandin F2 receptor negative regulator	PTGFRN	HS.418093
			433	KIAA0056 protein	KIAA0056	HS.438550
			630	Chromosome 1 open reading frame 33	C1orf33	HS.463797
			131	ElaC homolog 1 (E. coli)	ELAC1	HS.47572
			408	Mitochondrial ribosomal protein L2	MRPL2	HS.55041
			522	Eukaryotic translation elongation factor 1 epsilon 1	EEF1E1	HS.88977
			470	Ribonuclease P/MRP 38kDa subunit	RPP38	HS.94986
	Nuclear division	0.031	930	Cell division cycle 25A	CDC25A	HS.1634
	Biological Process GO:0000280		791	Cyclin B2	CCNB2	HS.194698
			633	Cyclin B1	CCNB1	HS.23960
			646	CHK1 checkpoint homolog (S. pombe)	CHEK1	HS.24529
			939	SKB1 homolog (S. pombe)	SKB1	HS.367854
			577	RAD51 homolog (RecA homolog, E. coli) (S. cerevisiae)	RAD51	HS.446554
			285	Kinesin family member 2C	KIF2C	HS.69360
			158	Cyclin A2	CCNA2	HS.85137
	Development	0.033	5	Jerky homolog-like (mouse)	JRKL	HS.105940
	Biological Process GO:0007275		11	Epithelial V-like antigen 1	EVA1	HS.116651
			504	zinc finger protein 74 (Cos52)	ZNF74	HS.127476
			34	Neighbor of COX4	NOC4	HS.173162
			197	PAP associated domain containing 1	PAPD1	HS.173946
			710	Dynamin 1-like	DNM1L	HS.180628
			824	High-mobility group box 3	HMGB3	HS.19114
			574	Bridging integrator 1	BIN1	HS.193163
			755	SRY (sex determining region Y)-box 9 (campomelic dysplasia, auto	SOX9	HS.2316
			646	CHK1 checkpoint homolog (S. pombe)	CHEK1	HS.24529
			276	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 5	B3GNT5	HS.257222
			165	UDP glycosyltransferase 8 (UDP-galactose ceramide galactosyltran	UGT8	HS.274293
			329	Keratin 17	KRT17	HS.2785
			390	Alpha-2-HS-glycoprotein	AHSG	HS.324746
			305	Forkhead box C1	FOXC1	HS.348883
			868	Keratin 14 (epidermolysis bullosa simplex, Dowling-Meara, Koebne	KRT14	HS.355214
			828	Fatty acid binding protein 5 (psoriasis-associated)	FABP5	HS.408061
			26	Crystallin, alpha B	CRYAB	HS.408767
			835	kinesin family member 1B	KIF1B	HS.444757
			550	Glutaredoxin 2	GLRX2	HS.458283
			92	Desmoplakin	DSP	HS.519873
			513	Chaperonin containing TCP1, subunit 6B (zeta 2)	CCT6B	HS.73072
			637	T-cell acute lymphocytic leukemia 1	TAL1	HS.73828
			664	Procollagen-lysine 1, 2-oxoglutarate 5-dioxygenase (lysine hydroxy	PLOD	HS.75093
			465	Carboxypeptidase Z	CPZ	HS.78068
			33	Kallikrein 6 (neurosin, zyme)	KLK6	HS.79361
			50	Chromosome 4 open reading frame 14	C4orf14	HS.8715
	Biosynthesis	0.035	748	Hypothetical protein CL640	CL640	HS.144304
	Biological Process GO:0009058		3	Adenosylmethionine decarboxylase 1	AMD1	HS.159118
			276	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 5	B3GNT5	HS.257222
			165	UDP glycosyltransferase 8 (UDP-galactose ceramide galactosyltran	UGT8	HS.274293
			204	Cytidine monophosphate N-acetylneuraminic acid synthetase	CMAS	HS.311346
			870	Alanyl-tRNA synthetase	AARS	HS.315137
			778	Ribosomal protein S5	RPS5	HS.378103
			865	Eukaryotic translation initiation factor (eIF) 2A	eIF2A	HS.378808
			281	stearoyl-CoA desaturase 4	SCD4	HS.379191
			333	Mitochondrial ribosomal protein S10	MRPS10	HS.380887
			442	Eukaryotic translation initiation factor 3, subunit 6 48kDa	EIF3S6	HS.405590
			934	Ribosomal protein S10	RPS10	HS.406620
			615	Prostaglandin F2 receptor negative regulator	PTGFRN	HS.418093
			887	Methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1, n	MTHFD1	HS.435974
			433	KIAA0056 protein	KIAA0056	HS.438550
			417	Uridine-cytidine kinase 2	UCK2	HS.458360
			630	Chromosome 1 open reading frame 33	C1orf33	HS.463797
			131	ElaC homolog 1 (E. coli)	ELAC1	HS.47572
			408	Mitochondrial ribosomal protein L2	MRPL2	HS.55041
			394	Adenylosuccinate lyase	ADSL	HS.75527
			522	Eukaryotic translation elongation factor 1 epsilon 1	EEF1E1	HS.88977
			470	Ribonuclease P/MRP 38kDa subunit	RPP38	HS.94986
	Histogenesis	0.037	461	CD59 antigen p18-20 (antigen identified by monoclonal antibodies '	CD59	HS.2785
	Biological Process GO:0009888		868	Keratin 14 (epidermolysis bullosa simplex, Dowling-Meara, Koebne	KRT14	HS.355214
			828	Fatty acid binding protein 5 (psoriasis-associated)	FABP5	HS.408061
			92	Desmoplakin	DSP	HS.519873
			664	Procollagen-lysine 1, 2-oxoglutarate 5-dioxygenase (lysine hydroxy	PLOD	HS.75093
	Macromolecule biosynthesis	0.041	276	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 5	B3GNT5	HS.257222
	Biological Process GO:0009059		165	UDP glycosyltransferase 8 (UDP-galactose ceramide galactosyltran	UGT8	HS.274293
			204	Cytidine monophosphate N-acetylneuraminic acid synthetase	CMAS	HS.311346
			870	Alanyl-tRNA synthetase	AARS	HS.315137
			778	Ribosomal protein S5	RPS5	HS.378103
			865	Eukaryotic translation initiation factor (eIF) 2A	eIF2A	HS.378808
			281	stearoyl-CoA desaturase 4	SCD4	HS.379191
			333	Mitochondrial ribosomal protein S10	MRPS10	HS.380887
			442	Eukaryotic translation initiation factor 3, subunit 6 48kDa	EIF3S6	HS.405590
			934	Ribosomal protein S10	RPS10	HS.406620
			615	Prostaglandin F2 receptor negative regulator	PTGFRN	HS.418093
			887	Methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1, n	MTHFD1	HS.435974
			433	KIAA0056 protein	KIAA0056	HS.438550
			630	Chromosome 1 open reading frame 33	C1orf33	HS.463797
			131	ElaC homolog 1 (E. coli)	ELAC1	HS.47572
			408	Mitochondrial ribosomal protein L2	MRPL2	HS.55041
			394	Adenylosuccinate lyase	ADSL	HS.75527
			522	Eukaryotic translation elongation factor 1 epsilon 1	EEF1E1	HS.88977
			470	Ribonuclease P/MRP 38kDa subunit	RPP38	HS.94986
	M phase	0.044	470	Ribonuclease P/MRP 38kDa subunit	RPP38	HS.1634
	Biological Process GO:0000279		791	Cyclin B2	CCNB2	HS.194698
			633	Cyclin B1	CCNB1	HS.23960
			646	CHK1 checkpoint homolog (S. pombe)	CHEK1	HS.24529
			939	SKB1 homolog (S. pombe)	SKB1	HS.367854
			577	RAD51 homolog (RecA homolog, E. coli) (S. cerevisiae)	RAD51	HS.446554
			285	Kinesin family member 2C	KIF2C	HS.69360
			158	Cyclin A2	CCNA2	HS.85137
Down-regulated in CK5/14+	Cell surface receptor linked signal transduction	<0.001	452	Ral guanine nucleotide dissociation stimulator	RALGDS	HS.106185
	Biological Process GO:0007166		463	Transforming growth factor, beta 1 (Camurati-Engelmann disease)	TGFB1	HS.1103
			845	Calcium and integrin binding 1 (calmyrin)	CIB1	HS.135471
			443	Fibroblast growth factor receptor 4	FGFR4	HS.165950
			291	Interferon-stimulated transcription factor 3, gamma 48kDa	ISGF3G	HS.1706
			512	Integrin, alpha L	ITGAL	HS.174103
			428	Adenosine A2a receptor	ADORA2A	HS.197029
			438	Regulator of G-protein signalling 5	RGS5	HS.24950
			801	Transducin-like enhancer of split 3 (E(sp1) homolog, Drosophila)	TLE3	HS.287362
			396	Regulator of G-protein signalling 16	RGS16	HS.413297
			568	Sprouty homolog 1, antagonist of FGF signaling (Drosophila)	SPRY1	HS.436944
			55	V-erb-b2 erythroblastic leukemia viral oncogene homolog 2	ERBB2	HS.446352
			189	Regulator of G-protein signalling 11	RGS11	HS.65756
			15	Bone morphogenetic protein 4	BMP4	HS.68879
			923	GRB2-associated binding protein 1	GAB1	HS.80720
			679	MAP-kinase activating death domain	MADD	HS.82548
			309	Growth factor receptor-bound protein 7	GRB7	HS.86859
	Enzyme linked receptor protein signaling pathway	0.002	463	Transforming growth factor, beta 1 (Camurati-Engelmann disease)	TGFB1	HS.1103
	Biological Process GO:0007167		443	Fibroblast growth factor receptor 4	FGFR4	HS.165950
			568	Sprouty homolog 1, antagonist of FGF signaling (Drosophila)	SPRY1	HS.436944
			55	V-erb-b2 erythroblastic leukemia viral oncogene homolog 2	ERBB2	HS.446352
			15	Bone morphogenetic protein 4	BMP4	HS.68879
			923	GRB2-associated binding protein 1	GAB1	HS.80720
			309	Growth factor receptor-bound protein 7	GRB7	HS.86859
	Transmembrane receptor protein tyrosine kinase signaling pathway	0.018	443	Fibroblast growth factor receptor 4	FGFR4	HS.165950
	Biological Process GO:0007169		568	Sprouty homolog 1, antagonist of FGF signaling (Drosophila)	SPRY1	HS.436944
			55	V-erb-b2 erythroblastic leukemia viral oncogene homolog 2	ERBB2	HS.446352
			923	GRB2-associated binding protein 1	GAB1	HS.80720
			309	Growth factor receptor-bound protein 7	GRB7	HS.86859
	Regulation of G-protein coupled receptor protein signaling pathway	0.029	438	Regulator of G-protein signalling 5	RGS5	HS.24950
	Biological Process GO:0008277		396	Regulator of G-protein signalling 16	RGS16	HS.413297
			189	Regulator of G-protein signalling 11	RGS11	HS.65756

Journal of Pathology
J Pathol 2006; **210**: 3–9
Published online 6 July 2006 in Wiley InterScience
(www.interscience.wiley.com) **DOI:** 10.1002/path.2022

Original Paper

Dual-colour chromogenic *in situ* hybridization for testing of *HER-2* oncogene amplification in archival breast tumours

M Laakso,[1,2]* M Tanner[2,3] and J Isola[2]

[1] *Department of Pathology, Seinäjoki Central Hospital, Seinäjoki, Finland*
[2] *Institute of Medical Technology, University and University Hospital of Tampere, Tampere, Finland*
[3] *Department of Oncology, Tampere University Hospital, Tampere, Finland*

**Correspondence to:*
M Laakso, Department of Pathology, Seinäjoki Central Hospital, Hanneksenrinne 7, 60220 Seinäjoki, Finland.
E-mail: mervi.laakso@epshp.fi

Received: 9 March 2006
Revised: 24 April 2006
Accepted: 6 May 2006

Abstract

Chromogenic *in situ* hybridization (CISH), which uses an enzymatic reaction to detect the hybridized DNA probe, is a new alternative to fluorescence *in situ* hybridization (FISH) for the assessment of *HER-2* oncogene amplification status in breast cancer. The main advantage of CISH over FISH is the use of bright-field microscopy, which is rapid and allows the histopathological evaluation of tumour tissue sections. The main disadvantage of CISH has been the use of a single probe, thereby making it necessary to hybridize the control probe (chromosome 17 centromere) on an adjacent tissue section. The present paper presents an efficient protocol for dual-colour CISH (dc-CISH) based on the co-hybridization of probes to the *HER-2* oncogene and chromosome 17 centromere. The probes were detected sequentially with antibodies to digoxigenin and biotin and with secondary antibody polymers labelled with horseradish peroxidase and alkaline phosphatase. The peroxidase reaction was visualized with tetramethyl benzidine (green reaction product) and the alkaline phosphatase reaction with New Fuchsin (red reaction product). The accuracy of the method was verified by comparing the results for four cell lines and 40 tumour samples with those obtained using FISH (Vysis Inc.). The results of FISH and dc-CISH showed high concordance (91%, Kappa coefficient = 0.82). It is concluded that dual-colour CISH, which is a new alternative to FISH enables the assessment of copy number ratio (*HER-2*/17 centromere) in conjunction with proper histopathological evaluation and the ease of bright-field microscopy.

Keywords: breast cancer; chromogenic *in situ* hybridization (CISH); chromosome 17 centromere; dual-colour CISH (dc-CISH); fluorescence *in situ* hybridization (FISH); *HER-2*

Introduction

Fluorescence *in situ* hybridization (FISH) is widely used in clinical tumour diagnostics to identify oncogene amplification and chromosomal translocation in the cellular genome. Since the initial reports on the use of FISH [1,2], its principle has remained essentially unchanged. Probes specific to genomic DNA sequences are hybridized onto cells after denaturing the genomic target DNA into a single-stranded form. The probes are labelled either directly with fluorescent reporter molecules (fluorescein, rhodamine, etc) or indirectly with haptens (digoxigenin or biotin) that are then detected using fluorescence-labelled antibodies. The FISH method was translated into clinical practice in pathology laboratories when the analysis of formalin-fixed and paraffin-embedded tumour tissues became possible [3,4].

As the use of monoclonal antibody-based therapy targeting the *HER-2* oncogene (trastuzumab, Herceptin) is now broadening to the adjuvant setting [5–7], pathology laboratories are currently requested to perform *HER-2* assays on all newly diagnosed breast cancers. It is well established that the response to trastuzumab therapy is best predicted by determining *HER-2* oncogene amplification status using FISH [8]. Immunohistochemistry (IHC), which detects HER-2 protein on the cell surface, is commonly used, but its accuracy has been questioned in several studies [8–12]. It is generally accepted that an IHC score of 2+ must be confirmed with an *in situ* hybridization test [8]. However, occasionally, *HER-2* oncogene amplification can also be found in tumours with an IHC score of 0/1+ [9,10,13] and up to 10–15% of cases that are immunohistochemically strongly positive (3+) show normal *HER-2* copy number [7–10,13]. Thus, there is high pressure that all newly diagnosed breast cancers should be tested for *HER-2* oncogene amplification with an *in situ* hybridization test.

For pathologists, the main difficulty encountered in adopting FISH is fluorescence microscopy, since it is

time-consuming and requires special equipment and familiarity with the method. As an alternative to fluorescent dyes, enzymatic detection of the probe can also be used in DNA *in situ* hybridization. This method, termed chromogenic *in situ* hybridization (CISH), is based on peroxidase- or alkaline phosphatase-labelled reporter antibodies that are detected using an enzymatic reaction [14]. The main advantage of CISH is that it can be viewed with a standard bright-field microscope along with a haematoxylin counterstain.

In FISH, *HER-2* and chromosome 17 centromere probes are most often co-hybridized when determining *HER-2* status [8]. The copy number of the reference probe aids in distinguishing true oncogene amplification from chromosomal aneuploidy [2,8,13,15–17]. Current CISH methods are based on the single-colour detection of a digoxigenin-labelled *HER-2* probe [8,14,15]. Here, we present an optimized protocol for dual-colour CISH (dc-CISH) that provides information regarding the copy number of the *HER-2* oncogene and the chromosome 17 centromere on a single sample slide in a manner similar to FISH. The dc-CISH method was found to be highly concordant with FISH and it now enables reliable analysis of gene amplification even in equivocal copy number cases along with the ease of bright-field microscopy.

Materials and methods

Samples

Routine formalin-fixed and paraffin-embedded tissue samples ($n = 40$) from newly diagnosed invasive breast carcinomas were used for the study. The samples were selected to represent *HER-2* non-amplified, amplified, and equivocal borderline copy number cases according to *HER-2* results obtained in routine diagnostics by using CISH. Four breast cancer cell lines with normal (MCF-7), borderline (MDA-453), moderately amplified (JIMT-1), and highly amplified (BT-474) *HER-2* copy number were included in the study. The cultured cells were formalin-fixed and embedded in paraffin wax as cell pellets using the thrombin clotting technique. This study was performed in accordance with local ethical guidelines.

Dual-colour chromogenic *in situ* hybridization (dc-CISH)

The slides were dewaxed and incubated in 0.01 M Tris–HCl with 0.001 M EDTA at pH 9.0 in an immunostaining pretreatment incubator (PT-Module, LabVision, Fremont, CA, USA) at 98–99 °C for 15 min, followed by a cooling period of 20 min. After a wash with phosphate-buffered saline (PBS), enzymatic digestion was carried out by applying 100 µl of digestion enzyme (Digest-All III solution; Zymed Inc, South San Francisco, CA, USA) to the slides for 1–2 min at room temperature. The slides were then washed with PBS and dehydrated with graded ethanols. For cases that were under-digested with this short pretreatment protocol, alternative pretreatment according to the FISH pretreatment protocol was carried out (see below).

A bacterial artificial chromosome (BAC) clone specific to *HER-2* DNA (Invitrogen, Carlsbad, CA, USA) was labelled with dUTP-digoxigenin (Roche Biochemicals, Mannheim, Germany) and the chromosome 17 centromere probe (p17H8) was labelled with dUTP-biotin (Roche Biochemicals) using nick translation [3]. The probe mixture [1.5 µl of nick-translated *HER-2* probe and 0.5 µl of nick-translated chromosome 17 centromere probe (both 250 ng/µl), 1.0 µl of placental DNA (1 µg/µl, Sigma-Aldrich), 0.5 µl of human COT1 DNA (1 µg/µl, Roche Biochemicals), and 6.5 µl of hybridization buffer containing 15% w/v dextran sulphate and 70% formamide in standard saline citrate (SSC) at pH 7.0] was applied to the slides and they were then covered with 18 × 18 mm coverslips and sealed with rubber cement. The sections were denatured on a thermal plate (3 min at 94 °C) and hybridized overnight at 42 °C. After hybridization, the cover slips were removed and the slides were washed with 0.5 × SSC (3 min at 72 °C), followed by another wash (1 min) at room temperature.

The protocol for dual-colour detection of the *HER-2* and chromosome 17 centromere probes is described in detail in Table 1. The probe for the chromosome 17 centromere was detected first by means of sequential incubations with mouse anti-biotin antibody (Zymed Inc.) and reagents from the Powervision + alkaline phosphatase polymer kit using New Fuchsin as a chromogen (Immunovision Technologies Co, Brisbane, CA, USA). After the enzymatic reaction, the slides were washed with distilled water and the detection of the digoxigenin-labelled *HER-2* probe was conducted as follows. The slides were incubated with anti-digoxigenin antibody (Roche Biochemicals) and reagents from the Powervision + horseradish peroxidase polymer kit (Immunovision Technologies Co). A ready-to-use tetramethyl benzidine (TMB) solution was used as a chromogen (cat. no. RDI-TMBSH; Research Diagnostics, Inc (RDI) Division of Fitzgerald Industries International Inc, Concord, MA, USA). The tissue sections were counterstained with haematoxylin, cleared and dehydrated with graded ethanol and xylene, and then embedded in paraffin wax.

Fluorescence *in situ* hybridization (FISH)

FISH was performed using a commercially available probe mixture of *HER-2* (SpectrumGreen), chromosome 17 centromere (SpectrumAqua), and *Topoisomerase II-alpha* (SpectrumOrange) based on the recommended protocol (cat. no. 32-191095; Vysis Inc, Abbott Laboratories, Illinois, USA). The slides were pretreated with 0.2 M HCl for 20 min and subsequently with a 0.01 M citric acid buffer with 0.05% citraconic

Table 1. Dual-colour CISH protocol for the detection of digoxigenin- and biotin-labelled genomic probes for the *HER-2* oncogene and chromosome 17 centromere

Reagent	Dilution	Source	Incubation time (min; RT)
1. Mouse monoclonal anti-biotin, Z021	1 : 300	Zymed Inc	30
2. Powervision + alkaline phosphatase kit		Immunovision Technologies Co	
2a. Post-blocking reagent	RTU*		20
2b. Alkaline phosphatase polymer	RTU		30
2c. New Fuchsin (NF)	1 drop of NF solution + 1 drop of NF activator + 0.9 ml of naphthol phosphate		10
3. Wash dH_2O	—	—	5
4. Anti-digoxigenin, 1.71.256	1 : 300	Roche Biochemicals	30
5. Powervision + HRP kit		Immunovision Technologies Co	
5a. Post-blocking reagent	RTU		20
5b. HRP polymer	RTU		30
6. Tetramethyl benzidine	RTU	RDI Division of Fitzgerald Industries International Inc	5
7. Wash dH_2O	—	—	10
8. Haematoxylin counterstain	1 : 4	DakoCytomation	1
9. Dehydration with ethanol	70%, 94%, and 99.5%	—	2 in each
10. Clearing with xylene	—	—	2 × 5
11. Mounting with xylene-based mountant	—	—	—
			Total time 3 h 30 min

* RTU = a ready-to-use reagent from the provider; RT = room temperature. Steps 1–8 were performed in a programmable immunostaining robot (LabVision Autostainer).

anhydride at 98 °C for 15 min. Enzymatic digestion was conducted at 37 °C for 20–25 min with Digest-All III solution. The slides were post-fixed with 10% formaldehyde for 10 min and dehydrated with graded ethanols.

Scoring of dc-CISH and FISH

The dc-CISH hybridizations were evaluated with an Olympus BX61 microscope using a 40× objective. For dc-CISH, three different tumour areas were visually chosen and at least 100 tumour cells were scored. The counting was performed using a CCD camera live image and the TouchCount mode of the AnalySIS imaging system (Soft Imaging Systems GmbH, Münster, Germany). In the TouchCount mode, the copy number is assessed by clicking on each gene copy with the mouse: this counts and marks each object on a live camera image with an overlay graphic symbol. The scoring of FISH was conducted in the same manner except that a 60× oil-immersion objective was used. SpectrumGreen and SpectrumAqua fluorescence (for *HER-2* and chromosome 17 centromere, respectively) was inspected with appropriate single-band pass filters. SpectrumOrange fluorescence (for *Topoisomerase II-alpha*, included in the probe) was ignored in this study. A minimum of 20 cells was counted for each FISH sample. All dc-CISH hybridizations were evaluated by an observer who was unaware of the results of the FISH assays.

Results

The dc-CISH method resulted in clearly distinguishable signal colours for *HER-2* (green) and the chromosome 17 centromere (red), and allowed separate signals to be counted without difficulty. This was true for both breast cancer cell lines and paraffin-embedded tumour samples showing a normal, borderline, or amplified *HER-2* copy number. Example micrographs of the cell lines and tumour samples are shown in Figure 1.

Dual-colour CISH versus FISH

Detailed results of FISH and dc-CISH are shown in Table 2. The mean *HER-2* and chromosome 17 centromere copy numbers and their ratios were highly concordant in the specimens representing normal, borderline, and amplified *HER-2* copy numbers. Neither the mean copy number of *HER-2* or chromosome 17 centromere nor their ratio displayed a systematic shift in either direction in dc-CISH when compared with FISH. The correlation of *HER-2*/17 centromere copy number ratios when using dc-CISH or FISH is shown in Figure 2 ($r = 0.89$, from all samples that were enumerable with both methods). Two tumours showed a low number of chromosome 17 centromere copies and only one copy of the *HER*-2 oncogene (tumours 1 and 2 in Table 2). This condition was also reliably recognized with dc-CISH, as the results were concordant with those of FISH.

When the generally accepted cut-off for *HER-2*/17 centromere copy number ratio (ratio 2) was applied

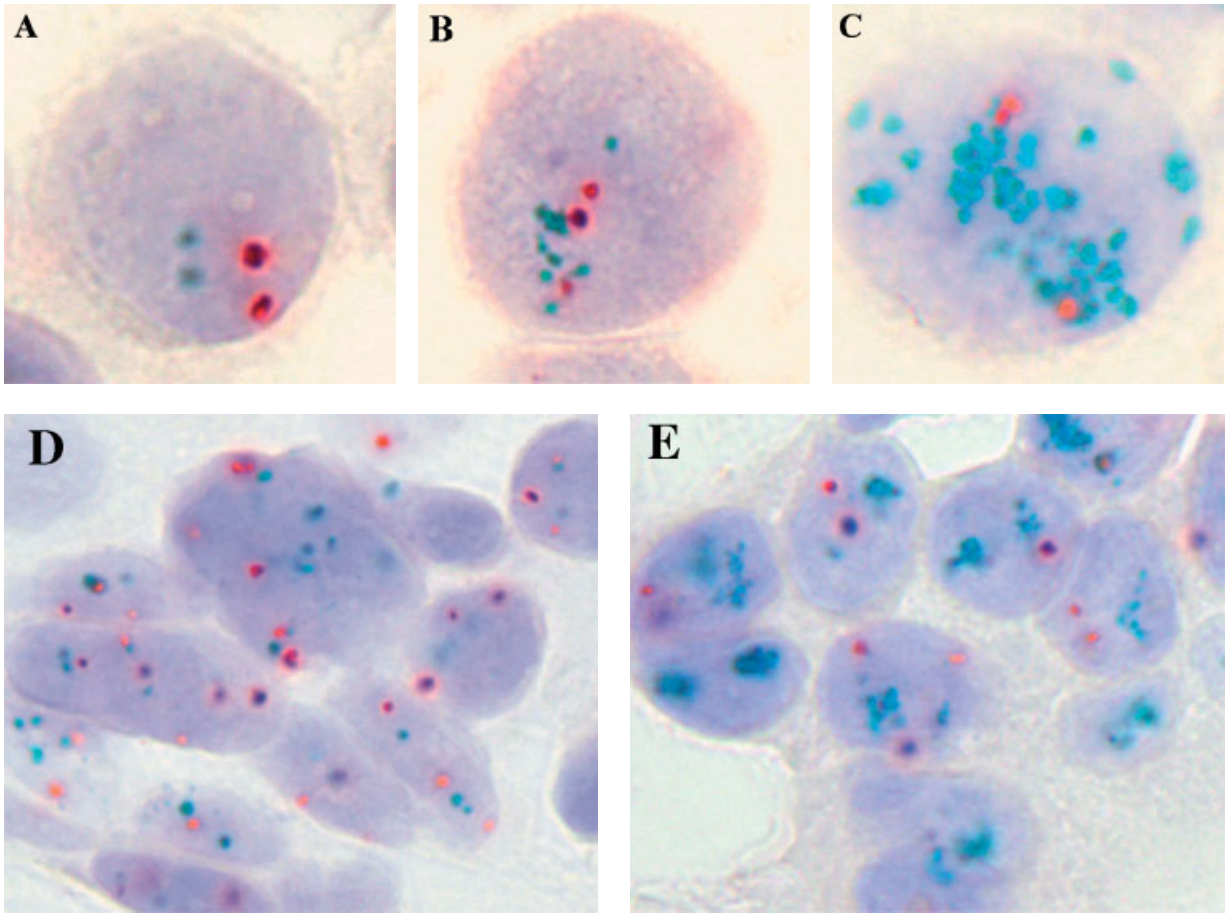

Figure 1. Micrographs demonstrating dual-colour CISH of the *HER-2* oncogene (green signal) and chromosome 17 centromere (red signal) in control cell lines MCF-7 (A), MDA-453 (B), BT-474 (C), and paraffin-embedded tumours with equivocal copy number (D) and gene amplification showing a gene copy cluster (E). Haematoxylin was used as a counterstain

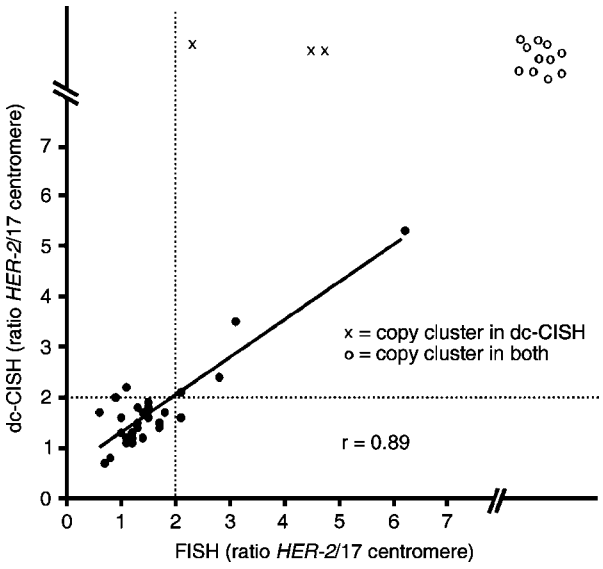

Figure 2. Correlation of *HER-2*/17 centromere ratio obtained by using the dc-CISH or FISH method in four cell lines and 40 breast tumour samples

here, there was 91% agreement between the two methods (Kappa coefficient 0.82, Table 3). One cell line (MDA-453) showed *HER-2* amplification by FISH but not by dc-CISH (Table 2), and three breast cancers were identified as having a *HER-2*/17 centromere copy number ratio of 2 or more by dc-CISH but not by FISH (tumours 5, 8, and 19 in Table 2). Of the 18 tumours with highly amplified *HER-2* oncogene, 14 showed a high number of gene copies that were clustered together, making enumeration by dc-CISH unreliable (Figure 1E). When critically scoring the copy numbers on the FISH slides, we considered *HER-2* copy number enumeration impossible in 11 of these 18 tumours.

Discussion

Our dual-colour modification of the CISH method proved to be successful, reliable, and highly reproducible, and it showed good concordance with FISH. Although most cases with *HER-2* amplification (or a normal copy number) are easy to distinguish both by FISH and by CISH, cases with an equivocal gene copy number (four to ten gene copies per cell) require careful examination and, often, enumeration of chromosome 17 centromeres. Although a set of samples selected to represent equivocal borderline copy number tumours was used in our series, dc-CISH and FISH yielded almost identical results when the

Table 2. Mean copy numbers and copy number ratios of the *HER-2* oncogene and chromosome 17 centromere by FISH and dual-colour CISH (dc-CISH) of four cell lines and 40 breast tumour samples

	HER-2		Chromosome 17		Ratio*	
Specimen	**FISH**	**dc-CISH**	**FISH**	**dc-CISH**	**FISH**	**dc-CISH**
MCF-7	1.7	1.8	2.3	2.6	0.7	0.7
MDA-453	7.0	5.9	3.3	3.6	2.1	1.6
JIMT-1	Amplified†	Amplified	1.6	1.9	Amplified	Amplified
BT-474	Amplified	Amplified	4.1	4.2	Amplified	Amplified
Tumour 1	0.8	1.1	1.1	1.7	0.7	0.7
Tumour 2	1.0	1.1	1.3	1.3	0.8	0.8
Tumour 3	1.4	2.1	1.3	1.8	1.1	1.1
Tumour 4	1.6	2.3	1.5	1.8	1.1	1.2
Tumour 5	1.7	3.4	1.6	1.6	1.1	2.2
Tumour 6	1.8	2.0	1.5	1.5	1.2	1.3
Tumour 7	1.8	2.6	1.8	1.6	1.0	1.6
Tumour 8	1.9	3.9	2.0	2.0	0.9	2.0
Tumour 9	1.9	3.0	0.7	1.2	2.8	2.4
Tumour 10	2.1	4.3	3.8	2.5	0.6	1.7
Tumour 11	2.2	3.8	1.5	2.3	1.5	1.6
Tumour 12	2.5	2.4	2.1	2.2	1.2	1.1
Tumour 13	2.7	3.6	2.1	1.9	1.3	1.8
Tumour 14	2.7	3.2	2.1	2.1	1.3	1.5
Tumour 15	2.7	2.9	1.6	1.9	1.7	1.5
Tumour 16	2.8	1.8	1.6	1.3	1.7	1.4
Tumour 17	3.0	3.0	2.4	2.2	1.3	1.4
Tumour 18	3.0	2.5	2.9	1.9	1.0	1.3
Tumour 19	3.0	3.6	3.3	1.7	0.9	2.0
Tumour 20	3.1	4.0	2.5	3.5	1.2	1.2
Tumour 21	3.1	2.3	2.2	1.3	1.4	1.7
Tumour 22	3.3	3.1	2.3	2.6	1.4	1.2
Tumour 23	3.3	2.4	1.8	1.4	1.8	1.7
Tumour 24	3.3	3.2	1.6	1.5	2.1	2.1
Tumour 25	3.5	3.9	2.3	2.0	1.5	1.9
Tumour 26	5.4	5.2	1.8	1.5	3.1	3.5
Tumour 27	8.2	7.5	5.3	4.0	1.5	1.8
Tumour 28	9.3	Amplified	4.2	2.3	2.2	Amplified
Tumour 29	12.5	11.0	2.0	2.1	6.2	5.3
Tumour 30	13.1	Amplified	2.8	2.3	4.6	Amplified
Tumour 31	16.1	Amplified	3.5	2.5	4.6	Amplified
Tumour 32	Amplified	Amplified	1.5	1.8	Amplified	Amplified
Tumour 33	Amplified	Amplified	1.6	1.8	Amplified	Amplified
Tumour 34	Amplified	Amplified	2.4	2.1	Amplified	Amplified
Tumour 35	Amplified	Amplified	3.0	2.0	Amplified	Amplified
Tumour 36	Amplified	Amplified	4.4	2.7	Amplified	Amplified
Tumour 37	Amplified	Amplified	Amplified	Amplified	Amplified	Amplified
Tumour 38	Amplified	Amplified	2.2	1.7	Amplified	Amplified
Tumour 39	Amplified	Amplified	2.4	2.0	Amplified	Amplified
Tumour 40	Amplified	Amplified	2.7	3.1	Amplified	Amplified

* A ratio of arithmetic means of *HER-2* and chromosome 17 centromere copy numbers.
† A high number of gene copies clustered together, making enumeration unreliable.

HER-2/17 centromere copy number ratio of 2 was used as a discriminator of amplification (91% concordance). In the cell line MDA-453, copy number ratios were similar to each other but on opposite sides of the given cut-off. In our opinion, this reflects the arbitrary nature of the cut-off, which was defined in the first FISH study of the *HER-2* oncogene [2]. The cut-off value has not been optimized by the most important biological correlate, ie the responsiveness to trastuzumab therapy.

Although some studies have reported very high concordance between FISH and CISH by detecting only *HER-2* in CISH [18], the highest concordance has been achieved when the chromosome 17 centromere copy number was assessed in equivocal cases

Table 3. The concordance between dual-colour CISH and FISH

	dc-CISH		Concordance
FISH	**Non-amplified***	**Amplified†**	**(%)**
Non-amplified*	22	3	91%
Amplified†	1	18	

* *HER-2*/17 centromere ratio < 2.
† *HER-2*/17 centromere ratio ≥ 2
Kappa coefficient = 0.82 (95% CI 0.65–0.99).

by a separate CISH hybridization [13,15,17,19–22]. The proportion of cases requiring a second CISH hybridization of the chromosome 17 centromere has been estimated to vary between 7% and 21% [13,15]. Our new dc-CISH method allows the simultaneous viewing of both the *HER-2* oncogene and the chromosome 17 centromere, and it closely resembles the US Food and Drug Administration (FDA)-approved dual-colour FISH technique (PathVysion, Vysis Inc.).

When optimizing the protocol for dc-CISH, the main difficulty encountered was to achieve a clearly visible separation of the hybridization signals in different colours. Diamino-benzidine (DAB) produces a dark brown signal, which was difficult to distinguish from the dark red reaction product of the chromosome 17 centromere probe detection. We therefore combined the red chromogen (New Fuchsin) with a new formulation of tetramethyl benzidine (TMB), which under these reaction conditions yields a bright green colour. The laboratory procedures for dc-CISH and single-colour CISH are principally the same, but dc-CISH requires two rounds of detection steps (Table 1). Nevertheless, the entire detection protocol can easily be programmed in any automated immunostainer that can perform the dual-colour detection protocol without additional manpower.

In general, the need for diagnostic *HER-2* oncogene tests is rapidly increasing. Trastuzumab therapy (Herceptin, Roche) has been used for the treatment of metastatic breast cancer for some years. Recently, it has been indicated as an adjuvant therapy in the primary stage of breast cancer [5–7]. Due to the development of adjuvant trastuzumab therapy, pathology laboratories are currently requested to perform *HER-2* assays for all newly diagnosed breast cancers. The assessment of HER-2 status by immunohistochemistry is a technically simple laboratory procedure, but it has shown an inferior association with trastuzumab treatment response when compared with FISH [8–11]. A confirmatory FISH or CISH test is considered necessary for cases showing a 2+ IHC result [8]. Furthermore, in several studies, up to 10–15% of the samples scored as 3+ by IHC are non-amplified [7–10,13], leading to costly, but most probably ineffective, trastuzumab treatment. In this scenario, it has been estimated that performing an *in situ* hybridization test for all newly diagnosed primary breast cancers is more cost-effective than the use of IHC and *in situ* hybridization in combination [23]. For many pathology laboratories, this translates to tens of *HER-2* FISH assays per week. Although the laboratory protocol for FISH is relatively simple, it has one drawback — namely, slide evaluation using fluorescence microscopy. This typically takes 5–15 min per sample, thereby creating a significant extra workload for pathologists.

CISH was developed to provide a more user-friendly alternative for pathologists who are accustomed to inspecting peroxidase-based immunohistochemical antibody staining [14]. In addition, CISH also enables proper assessment of tumour histopathology and the long-term storage of hybridized sample slides. According to several studies, CISH shows 90–100% concordance with FISH [13,15,17–22]. In a recent clinical trial, CISH was found to identify patients who showed excellent response to adjuvant trastuzumab therapy [7]. The results obtained in the present study demonstrate that dc-CISH combines the advantages of both FISH and CISH and allows the assessment of copy number ratio (*HER-2*/17 centromere) along with proper histopathological evaluation and the ease of bright-field microscopy. This is likely to facilitate more accurate patient selection for trastuzumab (Herceptin) therapy without the need for another hybridization of the chromosome 17 centromere on a separate sample slide as performed previously with single-colour CISH.

In addition to the determination of HER-2 status, the dc-CISH method presented here can be applied to other *in situ* hybridization-based applications. The detection of gene amplifications and deletions in human cancers is a growing field in pathology laboratories and the dc-CISH method now provides a competitive approach to these questions.

Acknowledgements

We wish to thank Ms Sari Toivola, Ms Eeva Riikonen, Ms Ritva Kujala, Ms Helvi Salmela, Ms Pirjo Pekkala, and Ms Päivi Kärki for skilful technical assistance. This study was financially supported by grants from the Pirkanmaa Hospital District Research Foundation, the Medical Research Fund of Seinäjoki Central Hospital, the Sigrid Juselius Foundation, Algol-Award, Oy Eli Lilly Finland Ab, the Finnish Medical Foundation, and the Finnish Cancer Foundation.

References

1. Pinkel D, Straume T, Gray JW. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc Natl Acad Sci U S A* 1986;**83:**2934–2938.
2. Kallioniemi O-P, Kallioniemi A, Kurisu W, Thor A, Chen L-C, Smith HS, *et al*. ERBB2 amplification in breast cancer analyzed by fluorescence *in situ* hybridization. *Proc Natl Acad Sci U S A* 1992;**89:**5321–5325.
3. Hyytinen E, Visakorpi T, Kallioniemi A, Kallioniemi OP, Isola JJ. Improved technique for analysis of formalin-fixed, paraffin-embedded tumors by fluorescence *in situ* hybridization. *Cytometry* 1994;**16:**93–99.
4. Pauletti G, Godolphin W, Press MF, Slamon DJ. Detection and quantification of HER-2/neu gene amplification in human breast cancer archival material using fluorescence *in situ* hybridization. *Oncogene* 1996;**13:**63–72.
5. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, Goldhirsch A, Untch M, Smith I, *et al*. Herceptin Adjuvant (HERA) Trial Study Team. Trastuzumab after adjuvant chemotherapy in HER-2-positive breast cancer. *N Engl J Med* 2005;**353:**1659–1672.
6. Romond EH, Perez EA, Bryant J, Suman VJ, Geyer CE Jr, Davidson NE, *et al*. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med* 2005;**353:**1673–1684.
7. Joensuu H, Kellokumpu-Lehtinen P-L, Bono P, Alanko T, Kataja V, Asola R, *et al*. Adjuvant Docetaxel and virorelbine with or without trastuzumab for breast cancer. *N Engl J Med* 2006;**354:**809–820.

8. Ross JS, Fletcher JA, Bloom KJ, Linette GP, Stec J, Symmans WF, *et al*. Targeted therapy in breast cancer; the HER-2/neu gene and protein. *Mol Cell Proteomics* 2004;**3:**379–398.
9. Bartlett JMS, Going JJ, Mallon EA, Watters AD, Reeves JR, Stanton P, *et al*. Evaluating HER2 amplification and overexpression in breast cancer. *J Pathol* 2001;**195:**422–428.
10. Sauer T, Wiedswang G, Boudjema G, Christensen H, Karesen R. Assessment of HER-2/neu overexpression and/or gene amplification in breast carcinomas: should *in situ* hybridization be the method of choice? *APMIS* 2003;**111:**444–450.
11. Mass RD, Press M, Anderson S, Cobleigh MA, Vogel CL, Dybdal N, *et al*. Evaluation of clinical outcomes according to HER2 detection by fluorescence *in situ* hybridization in women with metastatic breast cancer treated with trastuzumab. *Clin Breast Cancer* 2005;**6:**240–246.
12. Wixom CR, Albers EA, Weidner N. Her2 amplification; correlation of chromogenic *in situ* hybridization with immunohistochemistry and fluorescence *in situ* hybridization. *Appl Immunohistochem Mol Morphol* 2004;**12:**248–251.
13. Saez A, Andreu FJ, Segui MA, Bare ML, Fernandez S, Dinares C, *et al*. HER-2 gene amplification by chromogenic *in situ* hybridization (CISH) compared with fluorescence *in situ* hybridization (FISH) in breast cancer — a study of two hundred cases. *Breast* 2005;[Epub ahead of print].
14. Tanner M, Gancberg D, Di Leo A, Larsimont D, Rouas G, Piccard MJ, *et al*. Chromogenic *in situ* hybridization: a practical alternative for fluorescence *in situ* hybridization to detect HER-2/neu oncogene amplification in archival breast cancer samples. *Am J Pathol* 2000;**157:**1467–1472.
15. Isola J, Tanner M, Forsyth A, Cooke TG, Watters AD, Bartlett JMS. Interlaboratory comparison of HER-2 oncogene amplification by chromogenic and fluorescence *in situ* hybridization. *Clin Cancer Res* 2004;**10:**4793–4798.
16. McCormick SR, Lillemoe TJ, Beneke J, Schrauth J, Reinartz J. HER2 assessment by immunohistochemical analysis and fluorescence *in situ* hybridization: comparison of HercepTest and PathVysion commercial assays. *Am J Clin Pathol* 2002; **117:**935–943.
17. Loring P, Cumming R, O'Grady A, Kay EW. HER2 positivity in breast carcinoma: a comparison of chromogenic *in situ* hybridization with fluorescence *in situ* hybridization in tissue microarrays, with targeted evaluation of intratumoral heterogeneity by *in situ* hybridization. *Appl Immunohistochem Mol Morphol* 2005;**13:**194–200.
18. Park K, Kim J, Lim S, Han S, Lee JY. Comparing fluorescence *in situ* hybridization and chromogenic *in situ* hybridization methods to determine the HER2/neu status in primary breast carcinoma using tissue microarray. *Mod Pathol* 2003;**16:**937–943.
19. Vera-Román JM, Rubio-Martínez LA. Comparative assays for the HER-2/neu oncogene status in breast cancer. *Arch Pathol Lab Med* 2004;**128:**627–633.
20. Bhargava R, Lal P, Chen B. Chromogenic *in situ* hybridization for the detection of HER-2/neu gene amplification in breast cancer with an emphasis on tumors with borderline and low-level amplification: does it measure up to fluorescence *in situ* hybridization? *Am J Clin Pathol* 2005;**123:**237–243.
21. Gong Y, Gilcrease M, Sneige N. Reliability of chromogenic *in situ* hybridization for detecting HER-2 gene status in breast cancer: comparison with fluorescence *in situ* hybridization and assessment of interobserver reproducibility. *Mod Pathol* 2005;**18:**1015–1021.
22. Hauser-Kronberger C, Dandachi N. Comparison of chromogenic *in situ* hybridization with other methodologies for HER2 status assessment in breast cancer. *J Mol Histol* 2004;**35:**647–653.
23. Elkin EB, Weinstein MC, Winer EP, Kuntz KM, Schnitt SJ, Weeks JC. HER-2 testing and trastuzumab therapy for metastatic breast cancer: A cost-effectiveness analysis. *J Clin Oncol* 2006;**22:**854–863.