High-risk breast cancer:
From biology to personalized therapeutic strategies

Toshima Z. Parris

Department of Oncology
Institute of Clinical Sciences
Sahlgrenska Academy at University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2014

Cover illustration: Toshima Z. Parris

High-risk breast cancer: From biology to personalized therapeutic targets

toshima.parris@oncology.gu.se

ISBN 978-91-628-8841-1
http://hdl.handle.net/2077/34391

Printed in Bohus, Sweden 2014
Ale Tryckteam AB

To my family, for inspiring me
to be the best that I can be

If you can't get round a stump, leap over it, high and dry.
Have nerves of steel, a will of iron. Never mindside aches,
or heartaches, or headaches; dig away without stopping
to breathe, or to notice envy or malice. Set your target
in the clouds and aim at it. If your arrow falls short of the mark,
what of that? Pick it up and go at it again. If you should never
reach it, you'll shoot higher than if you only aimed at a bush.
Fanny Fern (1853)

Abstract

HIGH-RISK BREAST CANCER: FROM BIOLOGY TO PERSONALIZED THERAPEUTIC STRATEGIES

Toshima Z. Parris
Department of Oncology, Institute of Clinical Sciences,
Sahlgrenska Academy at University of Gothenburg, Sweden, 2014

Adjuvant treatment regimens for breast cancer are primarily based on patient- and tumor-related factors, e.g. patient menopausal status, tumor stage and histological grade, and the status of molecular tumor markers (HER2*/neu* and the estrogen receptor). Despite improvements in survival rates, about 20% of patients experience recurrence within five years of initial therapy. There is therefore a need to improve patient risk assessment and to personalize therapy according to a combination of patient-specific clinicopathological features and tumor characteristics. This doctoral thesis is a multidisciplinary effort between molecular biologists, clinicians, and pathologists to identify potential therapeutic targets for high-risk breast carcinoma.

This work exploits common knowledge that the accumulation of deleterious genetic and epigenetic modulators contribute to breast cancer risk for recurrence and death by deregulating key cellular processes within a specific tumor. In the first work, we found that tumors from high-risk breast cancer patients were genetically instable, containing a 2-fold increase in genetic alterations, an overrepresentation of alterations on chromosomes 3, 18, and 20, and the recurrent deregulation of a 13-marker transcriptome signature associated with significantly shorter disease-specific survival rates (*AZGP1*, *CBX2*, *DNALI1*, *LOC389033*, *NME5*, *PIP*, *S100A8*, *SCUBE2*, *SERPINA11*, *STC2*, *STK32B*, *SUSD3,* and *UBE2C*). Second, subsequent validation of the 13-marker signature demonstrated the importance of not only performing external validation in independent breast cancer microarray datasets, but also to assess the biological and clinical relevance of individual markers at the protein level because of frequent poor mRNA-protein correlation. It was shown that breast cancer outcome prediction was improved significantly by combining a four-marker immunohistochemical panel (AZGP1, PIP, S100A8, UBE2C) together with established clinicopathological features. Third, we showed that several putative markers previously identified by us may not only be useful for breast cancer prognostication, but may also be clinically relevant in oral squamous cell carcinoma, a cancer form bearing biological similarities to breast carcinoma. Lastly, we found that the 8p11-p12 genomic region is a hotspot for DNA amplification in breast cancer, where the *WHSC1L1* gene may be one of several genes located in region with oncogenic potential and a substantial contributor to the aggressive breast cancer phenotype.

Taken together, these findings further emphasize the importance of complementing established clinicopathological features with tumor-specific molecular markers to improve breast cancer risk assessment and develop more individualized treatment regimens.

Keywords: breast cancer, outcome prediction, molecular biomarker, 8p11-p12 amplification
ISBN: 978-91-628-8841-1

Populärvetenskaplig sammanfattning

Bröstcancer är den vanligaste cancerformen hos kvinnor och nyligen gick den om lungcancer som den cancersjukdom som orsakar flest dödsfall i västvärlden. År 2010 insjuknade cirka 7900 svenska kvinnor i bröstcancer, vilket är 30% av alla nya cancerfall i kvinnor. I allmänhet är bröstcancer betraktad som en åldersrelaterad sjukdom. De flesta kvinnor som insjuknar är i eller redan förbi klimakteriet. Däremot är bröstcancer ganska ovanlig hos kvinnor under 30 år och hos män, med bara cirka 30 fall inrapporterade under 2010. Antal nya fall i bröstcancer ökade ganska rejält vid mammografiinförandet i slutet på 80-talet och början på 90-talet. Dessutom kan uppgången bero på ökad livslängd och förändrad livsstil (rör på oss allt mindre, fetma, dricker allt mer alkohol, rökning, föder färre barn osv.). Patienter lever dock allt längre med sin sjukdom. För fyrtio år sedan var 5-årsöverlevnaden cirka 65% men idag är den närmare 90%. Återigen kan detta bero på att de flesta nya fallen upptäcks i ett tidigt skede samt bättre behandlingsmetoder (kirurgi, kemoterapi, strålbehandling, hormonbehandling och målinriktade behandlingar). Trots detta kommer cirka 20% av patienterna att få tillbaka sin cancer.

Arvsmassan (DNA molekyler) uppfyller en viktig funktion i kroppen. Den ger inte bara information om vilka celler som ska fungera som t.ex. hjärtceller eller hudceller men är också grunden till varför varje individ har ett visst utseende med särskild hudfärg, hårfärg osv. Cellerna förstår sin roll i kroppen genom att DNA molekyler (som innehåller informationen) för vidare instruktioner om vilka jobb som ska utföras inom en cell till RNA molekyler (gener) som i sin tur för vidare informationen till proteiner som utför jobbet. Varje dag uppstår det små fel i normala cellernas arvsmassa (genetiska avvikelser) men dessa brukar repareras, annars dör felaktiga celler. Däremot har cancerceller en speciell förmåga att kringgå celldöd när det uppstått fel i arvsmassan. Istället för att dö fortsätter cancerceller att växa och producera kopior av sig själv tills en samling av tumörceller (en tumör) bildas. Bröstcancer är en komplex sjukdom som är inte identisk hos två olika individer vilket kan förklara varför två patienter med samma symptom kanske svarar olika på samma behandling.

I denna avhandling hade vi två huvudmål. Först ville vi identifiera vilka genetiska avvikelser som ligger till grunden till varför vissa patienter har mer svårartade brösttumörer än andra. Sedan ville vi veta om denna information kunde hjälpa oss att hitta vilka patienter som troligen kommer få återfall eller som riskerar att dö på grund av sin sjukdom och därmed skulle behöva en tuffare behandling eller mer målinriktad behandling. Denna avhandling baseras på fyra delarbeten. Samtliga arbeten har utförts vid Sahlgrenska akademin vid Göteborgs universitet och Sahlgrenska universitetssjukhuset. Informationen om patienten och tumören erhölls från Regionalt Cancercentrum Västs cancerregister, Socialstyrelsens dödsregister, patologutlåtande och patientjournaler i enlighet med den regionala etikprövningsnämnden i Göteborgs bestämmelser.

I studie I upptäckte vi att bröstcancerpatienter som löper högre risk för återfall eller cancer-relaterad död har haft tumörer med ett stort antal genetiska avvikelser (dubbelt så många som patienter med mindre elakartade tumörer), specifika avvikelser på kromosomer 3, 18 och 20 samt förändrat genuttryck för 13 gener (på RNA nivåer). Detta var första gången dessa 13 generna associerades med kliniskt utfall för bröstcancerpatienter.

I studie II ville vi validera våra fynd i ett större utomstående material från bröstcancerpatienter. Eftersom forskare är skyldiga att dela med sig sina forskningsresultat kunde vi enkelt hitta 1141 bröstcancerprover som kunde användas för att jämföra våra resultat. Alla 13 generna kunde inte hittas i detta nya patientmaterial men vi kunde visa att genuttrycket för 6 av de 13 generna var väldigt likt det vi såg tidigare samt att denna förändring av genuttryck återigen påverkade kliniskt utfall för patienterna. I andra delen av studie II tyckte vi det var viktigt att undersöka om denna förändring av genuttrycket (på RNA nivå) hade förts vidare till cellens verksamma byggstenar (proteiner). Det finns tyvärr få metoder som kan undersöka proteinuttryck för de cirka 1 miljon kända proteinerna i ett experiment. Däremot finns det bra metoder för RNA molekyler. Trots att mängden RNA molekyler inte alltid ger upphov till samma mängd protein molekyler i samma cell kan vi ändå få en bild av vad som sker i en cell eller vävnad. På detta sätt kan vi begränsa de cirka 25000 mänskliga generna som kan ha betydelse för bröstcancer. Att validera ett tiotal gener på protein nivå är mer hanterlig än 25000 gener. I denna del av studien undersöktes proteinuttrycket för de 13 kandidatgenerna i fysiska tumörsnitt från patienttumörerna i studie I. Vi upptäckte att bara 3 av de 13 generna hade samma uttryck på RNA och proteinnivå. Viktigast av allt var att vi bättre kunde förutse högriskpatienter genom att kombinera information från 4 av de 13 proteinerna (AZGP1, PIP, S100A8, UBE2C) tillsammans med etablerade kliniska parametrar än när man endast använde de kliniska parametrarna.

I studie III undersökte vi om våra fynd i bröstcancer också kunde vara av betydelse i andra cancerformer. Här undersökte vi 16 bröstcancerrelaterade gener i munhålecancer, en cancerform som är biologiskt lik bröstcancer och som har ganska dålig 5-årsöverlevnad (50-60%). I denna studie analyserade vi proteinnivåer för de 16 markörerna i tumörsnitt från 43 patienter med munhålecancer. Vi visade att proteinuttrycket för de 16 markörerna var ganska likt i både bröstcancer och munhålscancer (bara 5 proteiner uttrycktes på olika sätt i de två cancerformerna). Detta tyder på att bröstcancer och munhålecancer är ganska lika trots att de härstammar från två olika delar av kroppen. Dessutom kunde vi se att två proteiner (CNTNAP2, S100A8) visade prognos medan tre andra proteiner (CBX2, SCUBE2, STK32B) var starkt kopplade med etablerade kliniska parametrar.

I det sista arbetet ville vi undersöka en av de mest förekommande genetiska avvikelserna i bröstcancer, dvs. avvikelser på kromosom 8 (närmare bestämt det 8p11-p12 genetiska området som finns i 15-25% av bröstcancerfall). Tidigare studier har visat att bröstcancerpatienter vars tumörer innehåller för mycket DNA runt det 8p11-p12 genetiska området har sämre prognos. Målinriktade behandlingar kunde utvecklas för denna genetiska avvikelse om genen eller generna som ligger i området hittas som ger upphov till denna elakartade typen av bröstcancer. Denna studie är fortfarande pågående men vi har hittat den *WHSC1L1* genen som är en stark kandidat. I normala celler påverkar genen celldelning. Eftersom vi har visat att ökad mängd DNA för *WHSC1L1* genen gett upphov till ökad mängd RNA och protein, tror vi att detta kan förklara varför brösttumörer som innehåller denna genetiska avvikelse växer snabbare än tumörer som inte har avvikelsen. Nu undersöker vi *WHSC1L1* genens roll i olika typer av bröstcancerceller som antingen innehåller eller saknar avvikelser på kromosom 8.

Sammanfattningsvis har vi kombinerat information från tumören och patienten för att identifiera nya mål för framtida behandlingar för bröstcancer. Våra fynd måste således bekräftas på proteinnivå med ett större ingående patientmaterial.

List of publications

This doctoral thesis is based on the following papers, which will be referred to in the text by Roman numerals:

I. **Parris T.Z.**, Danielsson A., Nemes S., Kovács A., Delle U., Fallenius G., Möllerström E., Karlsson P., and Helou K. Clinical implications of gene dosage and gene expression patterns in diploid breast carcinoma. *Clinical Cancer Research*. 2010 Aug 1;16(15):3860-74.

II. **Parris T.Z.**, Kovács A., Aziz L., Hajizadeh S., Nemes S., Semaan M., Forssell-Aronsson E., Karlsson P., and Helou K. Additive effect of the AZGP1, PIP, S100A8, and UBE2C molecular biomarkers improves outcome prediction in breast carcinoma. *International Journal of Cancer*. 2013 *in press*.

III. **Parris T.Z.***, Aziz L.*, Kovács A., Hajizadeh S., Nemes S., Semaan M., Karlsson P., and Helou K. Clinical relevance of breast cancer-related genes as potential biomarkers for oral squamous cell carcinoma. *Submitted*.

IV. **Parris T.Z.**, Kovács A., Hajizadeh S., Nemes S., Semaan M., Karlsson P., and Helou K. Functional significance of *WHSC1L1* gene amplification and/or over-expression in breast carcinoma. *Manuscript*.

* These two authors contributed equally to the work.

The following papers are not included in the thesis but are of relevance to the field.

1. **Parris T.**, Nik A.M., Kotecha S., Langston C., Helou K., Platt C., and Carlsson P. Inversion upstream of *FOXF1* in a case of lethal alveolar capillary dysplasia with misalignment of pulmonary veins. Am J Med Genet Part A 2013;161A:764-770.

2. Langen B., Rudqvist N., **Parris T.Z.**, Schüler E., Helou K., and Forssell-Aronsson E. Comparative analysis of transcriptional gene regulation indicates similar physiologic response in mouse tissues at low absorbed doses from intravenously administered ^{211}At. J Nucl Med 2013;54:1-9.

3. Shubbar E., Kovács A., Hajizadeh S., **Parris T.Z.**, Nemes S., Gunnarsdóttir K., Einbeigi Z., Karlsson P., Helou K. Elevated cyclin B2 expression in invasive breast carcinoma is associated with unfavorable clinical outcome. BMC Cancer 2013;13:1.

4. Danielsson A., Claesson K., **Parris T.Z.**, Helou K., Nemes S., Elmroth K., Elgqvist J., Jensen H., Hultborn R. Differential gene expression in human fibroblasts after alpha-particle emitter (211)At compared with (60)Co irradiation. Int J Radiat Biol. 2013;89(4):250-8.

5. Nemes S., Danielsson A., **Parris T.Z.**, Jonasson J.M., Karlsson P., Steineck G., and Helou K. A diagnostic algorithm to identify paired tumors with clonal origin. Genes Chromosomes Cancer 2013;52(11): 1007-16.

6. Nemes S., **Parris T.Z.**, Danielsson A., Einbeigi Z., Steineck G., Jonasson J.M., and Helou K. Integrative genomics with mediation analysis in a survival context. Comput Math Methods Med. 2013. *In press*.

7. Nemes S., **Parris T.Z.**, Danielsson A., Kannius-Janson M., Jonasson J.M., Steineck G., and Helou K. Segmented regression, a versatile tool to analyze mRNA levels in relation to DNA copy number aberrations. Genes Chromosomes Cancer 2012;51(1):77-82.

8. Rudqvist N., **Parris T.Z.**, Schüler E., Helou K., Forssell-Aronsson E. Transcriptional response of BALB/c mouse thyroids following in vivo astatine-211 exposure reveals distinct gene expression profiles. EJNMMI Res. 2012;2(1):32.

9. Schüler E., **Parris T.Z.**, Rudqvist N., Helou K., Forssell-Aronsson E. Effects of internal low-dose irradiation from 131I on gene expression in normal tissues in Balb/c mice. EJNMMI Res. 2011;1(1):29.

10. Möllerström E., Kovács A., Lövgren K., Nemes S., Delle U., Danielsson A., **Parris T.**, Brennan D.J., Jirström K., Karlsson P., Helou K. Up-regulation of cell cycle arrest protein BTG2 correlates with increased overall survival in breast cancer as detected by immunohistochemistry using tissue microarray. BMC Cancer. 2010;10:296.

11. Möllerström E., Delle U., Danielsson A., **Parris T.**, Olsson B., Karlsson P., Helou K. High-resolution genomic profiling to predict 10-year overall survival in node-negative breast cancer. Cancer Genet Cytogenet. 2010;198(2):79-89.

12. Nemes S., **Parris T.Z.**, Danielsson A., Karlsson P., Jonasson J.M., Steineck G., and Helou K. Differential gene expression networks and prognosis for breast cancer-specific survival. Manuscript.

Table of contents

Abbreviations

Array-CGH	microarray-based comparative genomic hybridization
BAC	bacterial artificial chromosome
BASE	BioArray Software Environment
BRE	Bloom-Richardson-Elston tumor grading system
bp	base pairs
cDNA	complementary DNA
CNA	copy number alteration
CNV	copy number variation
Cy3	cyanine 3
Cy5	cyanine 5
Da	dalton
DAPI	4', 6-diamidino-2-phenylindole
DBC	diploid breast carcinoma
DMEM	Dulbecco's Modified Eagle Medium
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DFS	disease-free survival
DMFS	disease metastasis-free survival
DSS	disease-specific survival
ER	estrogen receptor
FBS	fetal bovine serum
FDR	false discovery rate
FFPE	formalin-fixed, paraffin embedded specimens
FISH	fluorescence *in situ* hybridization
FITC	fluorescein isothiocyanate
gDNA	genomic DNA
HER2	human epidermal growth factor receptor type 2
IHC	immunohistochemistry
kb	kilobases
kDa	kilodalton
L-15	Leibovitz medium
Mb	megabases
mRNA	messenger ribonucleic acid
NaPyr	sodium pyruvate
NEAA	non-essential amino acids
OS	overall survival
OSCC	oral squamous cell carcinoma

PAD	pathological/anatomical diagnostics
PCR	polymerase chain reaction
pN0	axillary lymph node-negative
pN1	axillary lymph node-positive
PEST	penicillin-streptomycin
PgR	progesterone receptor
qPCR	quantitative real-time PCR
RIN	RNA integrity number
RNA	ribonucleic acid
RNase	ribonuclease
RPMI	Roswell Park Memorial Institute medium
SCC	squamous cell carcinoma
TNM	T: tumor size; N: lymph node; M: distant metastasis

Introduction

Cancer development

Because of its crab-like appearance, the father of Western medicine, Hippocrates, called cancer “karkinos” which is the Greek word for crab. Carcinoma stems from this word and is the medical term for cancers that originate in the epithelial component. For centuries, breast cancer was considered “cancer” as it was among the only cancer form which was outwardly visible. A connection between a concealed cancer of the inner organs and the impending death that followed was not an easy task as autopsies were rarely performed in the ancient world. “Cancer” was well-known long before Hippocrates gave it a name. Two ancient Egyptian papyri from 1500 B.C. (the Edwin and Ebers Papyri) provide specific accounts of cancer diagnoses and treatment, including untreatable “bulging tumors” in the breast. Hippocrates made an attempt to describe cancer and postulated the humoral theory. According to this theory, the body contains four humors (or fluids): blood, phlegm, yellow bile (choler or anger), and black bile (melancholy). He believed that cancer derives from black bile, which isn’t confined to just the tumor mass but rather is systemic and spreads throughout the body (1). Although the ancients were wrong about many aspects of the disease, Hippocrates was right when he described cancer as a systemic disease.

Today, there are several hundred known cancer forms, each characterized by the cell type of origin. Cancer (malignant epithelial neoplasm) is a heterogeneous group of diseases characterized by uncontrolled cellular growth, invasion of surrounding normal tissue, and spread of abnormal cells to distant sites via lymphatic and/or blood vessels. If metastasis remains unchecked, it can result in death. Malignant epithelial neoplasms are distinguishable from benign tumors, which remain confined within the basement membrane and lack the ability to metastasize.

The heterogeneous nature of tumors has been explained by two models: the clonal evolution model and the cancer stem cell (CSC) model. The two models differ in that the clonal evolution model postulates that the majority of tumor cells within a tumor mass are potentially tumorigenic due to the

accumulation of deleterious genetic alterations and that all cells within a tumor must be killed to eliminate the cancer. On the other hand, the cancer stem cell model hypothesizes that only a small fraction of tumor cells are the driving force of the malignancy and only these need be eradicated to prevent tumor recurrence and metastasis.

Cancer is caused by a variety of multifactorial characters including hereditary, external environmental (infections, chemicals, radiation, pollutants), and lifestyle factors (tobacco, alcohol, obesity). These causal factors may together or independently lead to the generation and accumulation of genetic abnormalities in normal cells (mutations). It is estimated that dividing cells produce thousands of somatic mutations every day, but these are usually repaired otherwise cell cycle progression may be inhibited (2). However, during the multistep process of tumorigenesis, inherited or acquired somatic mutations (sporadic) are accumulated in genes that convert normal cells into malignant neoplastic cells. The majority of cancer cases are sporadic, while cases attributed to inherited genes, e.g. the *BRCA1* and *BRCA2* breast/ovarian cancer risk genes, alone only account for 5-10% of cancers (3). The oncogenic process occurs in both sexes, any organ of the body, and in all age groups (Figure 1). However, this process takes time which may explain why cancer is a relatively rare occurrence during an average human lifetime and is frequently an age-related disease.

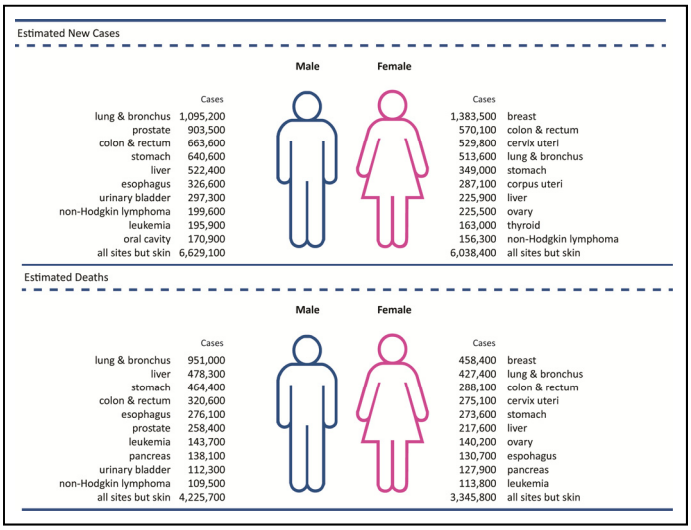

Figure 1. Leading sites of estimated new cancer cases and deaths worldwide in 2008. The figure is adapted from (4). Source: GLOBOCON 2008.

Cancer genetics and epigenetics

The deregulation of gene functions maintaining cell proliferation, apoptosis, genome stability, angiogenesis, invasion, metastasis, cellular metabolism, and tumor-promoting inflammation support tumor development and progression (5, 6). It is estimated that almost 2% of the 20,000-25,000 protein-coding genes in the human genome contribute to tumorigenesis (7, 8). The normal function of cancer-causing genes is to regulate cell proliferation and differentiation. However, during tumorigenesis these normal functions are altered to enhance cellular growth and promote metastatic dissemination. The three main types of cancer-causing genes are oncogenes which stimulate cell growth, tumor suppressor genes which inhibit cell growth, and DNA repair genes which repair errors in the DNA sequence during cell division. Proto-oncogenes can be activated by gene amplification, point mutation, partial deletion, inversion, chromosomal translocation, virus integration, or hypomethylation resulting in the formation of an oncogene with new or enhanced protein activities. These mutations are dominant, gain-of-function mutations and it is sufficient to mutate one allele to observe a change in the phenotype. Secondly, tumor suppressor genes are inactivated by chromosomal deletion, point mutation, mitotic recombination, gene conversion, or hypermethylation resulting in proteins with little or no activity. These mutations are recessive, loss-of-function mutations and both alleles must be mutated to silence normal activity. When referring to oncogenes and tumor suppressor genes, it is important to distinguish between the inherited DNA copy number found in the germline (copy number variations, CNV) or DNA copy number changes acquired (copy number alterations, CNA) during the oncogenic process. Lastly, mutations in genes that repair errors in replicated DNA during cell division (DNA repair genes) can promote the accumulation of additional genetic mutations.

Mechanisms that produce chromosomal aberrations associated with gene amplifications have been explained by various models. The most widely accepted model is the breakage-fusion-bridge model, although alternative mechanisms for oncogene amplification have recently been proposed (9). Breakage-fusion-bridge (also called anaphase bridges) cycles produce amplification of specific regions via double-stranded breaks as a result of replication stress, whereby generating genetic heterogeneity within a cell population (10). These breaks occur at chromosomal fragile sites, which are specific regions in mammalian chromosomes that are prone to breakage and rearrangements. There are two types of fragile sites: [1] common fragile

sites which are present in all cells of all individuals, and [2] rare fragile sites, such as the fragile X site, that are only present in a few individuals (10-15). There are approximately 100 common fragile sites in the human genome which encompass an estimated >100 Mb of DNA (14). All fragile sites do not form breaks at the same frequency. This is reflected in the variety of amplicon forms and sizes in the human genome. Amplified DNA can be present as double minutes, homogeneously staining regions (HSRs), or distributed throughout the genome and can range in size from kilobases to tens of megabases (16).

Although genetic aberrations have for many years been one of the primary focuses of cancer research, it is now recognized that epigenetic modulations (e.g. DNA methylation and histone modification) also play a central role in the tumorigenic process without altering the DNA sequence. Subsequent epigenetic modulations can either enhance (e.g. hypomethylation, removal of methyl groups from cytosine sites) or silence gene expression (e.g. hypermetylation, accumulation of methylated cytosines) by affecting the appearance of the DNA major groove and thereby interfere with the binding of proteins to DNA molecules. DNA methylation can either increase or decrease the level of transcription, depending on whether a positive or negative regulatory element is inactivated (17). Research during the past few decades has provided evidence showing crosstalk between genetic and epigenetic abnormalities driving tumorigenesis (18-22). Therefore, the merging of the genetic and epigenetic fields has provided us with a better understanding of cancer development and progression.

Normal mammary development

Normal breast tissue is extremely dynamic as its composition, form, function, and pathology changes continuously during the lifetime of a female. In placental mammals, the mammary gland is essentially a highly evolved skin gland that resembles hair rudiments during the early stages of fetal development. The primary function of the mammary gland is to produce milk during late pregnancy and after childbirth. Breast milk serves two purposes: to provide nourishment and immunological protection for the infant (23).

Human mammary gland development (mammogenesis) occurs in three defined stages closely interconnected with sexual development and reproduction: embryonic, pubertal, and adult (pregnancy, lactation, and

involution) (23, 24). At birth, breast tissue is identical in both sexes with a collection of 4-18 lactiferous ducts which lead to the nipple (25). During the onset of puberty, the mammary gland remains rudimentary in males while tissue maturation continues in females under the cyclic hormonal control of steroid and peptide hormones (mainly ovarian estrogen). It is during puberty or within 1-2 years after menarche that enlargement of the breasts occurs as adipose cells accumulate in the connective tissue and elongation and extensive branching of the ducts is stimulated by the secretion of estrogen (23, 24). Breast tissue is sensitive to hormonal stimuli during the three phases of the menstrual cycle: estrogen levels peak during the first half of the cycle, which leads to ovulation halfway through the cycle, and lastly progesterone production stimulates the formation of milk glands. It is not until pregnancy and lactation have occurred that the breasts are considered to be fully developed. At this stage, progesterone stimulation causes the lobules to enlarge significantly but at the end of the lactational stage the breasts resemble that seen in nulliparious women. During menopause, secretion of estrogen decreases dramatically while progesterone levels remain constant which leads to progressive atrophy (involution) of the breast tissue: the connective tissue dehydrates, becomes inelastic, and is replaced by adipocytes, causing the breast tissue to shrink and lose shape as the size and number of lobules decreases.

Breast anatomy in the mature female

Our conception of the breast anatomy in the lactating female has changed dramatically since surgeon and anatomist Sir Astley Cooper first published his work, “Anatomy of the Breast”, in 1840. For 165 years, the breasts were described as containing lactiferous sinuses used for milk storage in close proximity of the nipple, but this model has recently been discredited (25). In fact, we now know that sinuses do not exist and that breast milk is stored at the site of synthesis, in the clusters of secretory alveoli (also called acini) that make up the lobules.

The branched duct system of the breast begins in the distally located terminal duct-lobular unit (TDLU) which exits at the summit of the nipple through the lactiferous ducts; TDLUs are absent in males. Histologically, breast tissue is composed of an epithelial (TDLU and surrounding myoepithelium) and stromal component. The epithelium is composed of a cell bilayer consisting of a layer of secretory luminal cells surrounded by an outer layer of contractile basal myoepithelial cells (MEC). The MEC are located in close proximity to the basement membrane which separate the

epithelium from the underlying stromal component (26). Only about 10% of the breast volume is comprised of epithelial cells, whereas the rest consists of sheets of connective tissue (fascia), broad fibrous bands of connective tissue (Cooper's ligaments), and adipose tissue. The fascia, which is composed of blood and lymphatic vessels, nerves, adipose and fibrous tissue separates breast tissue from the overlying dermis of the skin and underlying pectoral muscles. The Cooper's ligaments, which provide support are interwoven among the lobes and run from the deep fascia to the dermis.

The blood supply and lymphatic drainage of the breasts are quite extensive. The breasts are highly vascular with major veins leading to the pulmonary capillaries or the network of vertebral veins. The lymphatic fluid from the various quadrants of the breast travels to specific lymph nodes. A large proportion of the lymph (about 75%) is transported from the lateral quadrants of the breasts to the ipsilateral axillary lymph nodes, while the rest is either transported from the medial quadrants to the parasternal nodes or to the other breast, or from the lower quadrants to the abdominal lymph nodes.

Breast carcinoma

The primary focus of this thesis will be breast carcinoma, in particular, sporadic cases of invasive breast carcinoma in females. Figure 2 illustrates the pathobiological events, including the accumulation of epigenetic and genetic alterations, involved in the transformation of normal TDLU to invasive breast carcinoma.

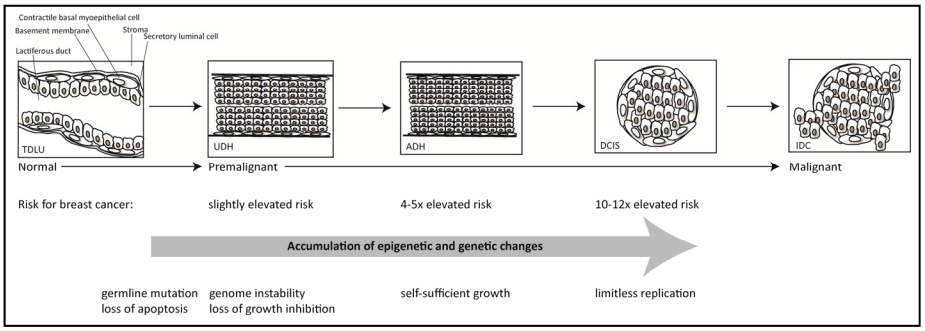

Figure 2. The pathobiological events associated with invasive breast cancer. Primary invasive breast tumors arise from normal breast epithelia that have acquired and accumulated epigenetic and genetic alterations, resulting in the deregulation of cancer-related functions, morphological changes, and elevated risk for breast cancer. UDH denotes usual ductal hyperplasia; atypical ductal hyperplasia (ADH), ductal carcinoma in situ (DCIS), invasive ductal carcinoma (IDC).

ETIOLOGY

Breast cancer is the most common cancer form and leading cause of cancer-related death among women worldwide (4). Incidence rates for breast cancer have increase dramatically since World War II. In the United States alone, an estimated 230,480 new invasive breast cancer cases and 39,520 deaths from the disease were projected to occur among women in 2011 (27), while approximately 8,000 new breast cancer cases are reported in Sweden annually (28). Although males can develop the disease, breast cancer only develops in the ducts (rare) and papillary cancer because males lack TDLU and cannot develop cancer in the lobules. Recently, Sir Cooper's observations have formed the basis for the proposed hypotheses (the Sick Lobe and the Biological Timing Theories) that breast carcinoma derives from mutated stem or progenitor cells present at multiple foci within a single lobe either simultaneously or asynchronously and that malignant transformation is also time-dependent (29).

The etiology of breast cancer is not fully known and may vary from patient to patient. The risk for developing breast cancer increases with certain lifestyles, diet choices, body size, obesity, alcohol consumption, stress, exposure to toxins, ionizing radiation in the chest area, reproductive factors, high breast density, and hormonal imbalances. The drastic increase in incidence rates since World War II has been associated with three major changes in Western lifestyles: increased consumption of processed foods (highly refined sugar, bleached flour, and vegetables oils), altered farming procedures and food preparation, and increased exposure to chemicals (30). When sugars and bleached flour are metabolized, glucose is produced. In 1956, Otto Warburg, Nobel Prize awardee and renowned biochemist, showed that cancer cells are addicted to glucose and aerobic conditions for growth and survival (31). Recent studies have shown that organic fruits and vegetables inhibit proliferation of certain cancer cell lines (32, 33).

Today, we have a better understanding of the impact cyclical estrogen levels have on breast carcinogenesis because of innovative experiments performed by Colonel Sir George Beatson in 1896 showing the regression of advanced breast cancer following an oophorectomy (surgical removal of the ovaries) (34). Since then, a relationship has been established between increased breast cancer risk and prolonged endogenous and/or exogenous estrogen exposure, e.g. early age at menarche (<12 years), nulliparity or late age at first full-term pregnancy (>30 years), late age at menopause, mammographic density, oral contraceptives, and hormonal replacement

therapy (35-37). A prime example of the effects of nulliparity is the high incidence of breast cancer in Roman Catholic nuns (1).

DIAGNOSIS

Today, about a half million Swedish women between 40-74 years of age participate in the mammography screening program annually. Consequently, about 65-70% of breast cancer cases in Sweden are detected early with mammography before any visible symptoms develop and the screening program has thereby reduced breast cancer-related deaths by 21% (38). Mammography screening is particularly beneficial because early stage breast cancer is usually asymptomatic. Although the most common finding in symptomatic patients is a lump in the breast, about 90% of breast lumps are benign (fluid filled cysts or fibroadenoma). On the other hand, advanced breast cancer may produce a variety of symptoms:

- A lump in the breast or a lump or swelling in the armpit persisting after a menstruation cycle
- Tenderness in the breast (although lumps are usually painless) persisting after a menstruation cycle
- Changes in the surface of the breast (size, shape, skin color, texture, temperature) such as skin dimpling
- Changes in the appearance of the nipple such as size or shape, a rash, or an unusual discharge (clear or bloody fluid) from the nipple

Guidelines for diagnosis of breast cancer include using a four-point system with clinical (palpation of the breast and local regional lymph nodes), pathological (core needle biopsy or fine needle aspiration), and radiological examinations (mammography of the breasts and ultrasound of the breast and local regional lymph nodes). A final diagnosis is made by a pathologist according to the World Health Organization (WHO) histological classification of tumors of the breast, the tumor-node-metastasis (TNM) staging system, and the Nottingham (Elston-Ellis) histologic tumor grading system (BRE). The TNM staging system (stage I-IV) provides a description of how advanced a tumor is by taking into account the size of the tumor (T1-T4), the extent of spread to the regional lymph nodes (N0-N3), and the extent of spread (M0-M1) to other parts of the body (Figure 3). The M1 category can be further subclassified according to the anatomical region of distant metastasis.

Malignant breast cancers can derive from the three main components of breast tissue: epithelium, myoepithelium, and the mesenchymal stroma.

However, the majority of malignant breast cancers are derived from luminal epithelial cells (>90%), whereas neoplasms of MEC origin (e.g. malignant invasive myoepithelial cancer) are the most uncommon and benign form of breast tumors (e.g. benign myoepithelioma), possibly because MECs secrete suppressor proteins (e.g. maspin) which limit tumor growth, invasiveness, and neoangiogenesis (26, 39). Soft tissue sarcomas originating from mesenchymal stromal cells also occur in breast tissue but are extremely rare (<1%). Secondary malignancies in the breast, such as secondary angiosarcomas or fibrosarcomas, frequently occur following radiotherapy. Fibroepithelial tumors, e.g. fibroadenoma and phyllodes tumors, are biphasic lesions originating from both the stromal and epithelial components.

The most common metastatic sites in advanced breast cancer are the chest wall, regional lymph nodes, and/or the skeleton. The liver, lung, and central nervous system are less common sites of recurrence. The liver is most likely a site of distant metastasis because the blood supply is filtered by the liver. Metastases to specific sites are frequently associated with different breast cancer subtypes (40). In patients with far advanced disease, metastasis can be found in almost any organ.

A pathological assessment of tumor grade (Grade I-III), or degree of differentiation, measures how closely neoplastic breast cells resemble normal breast epithelium by calculating a combined score including three morphological features (tubule formation, nuclear pleomorphism, and mitotic count). Well differentiated tumors (Grade I, favorable prognosis) are organized similarly to the tissue of origin. However, during tumor evolution a tumor becomes more dedifferentiated and disorganized resulting in neoplastic tissue that does not resemble the tissue of origin (moderately differentiated tumors (Grade II) and poorly differentiated tumors (Grade III, unfavorable prognosis)).

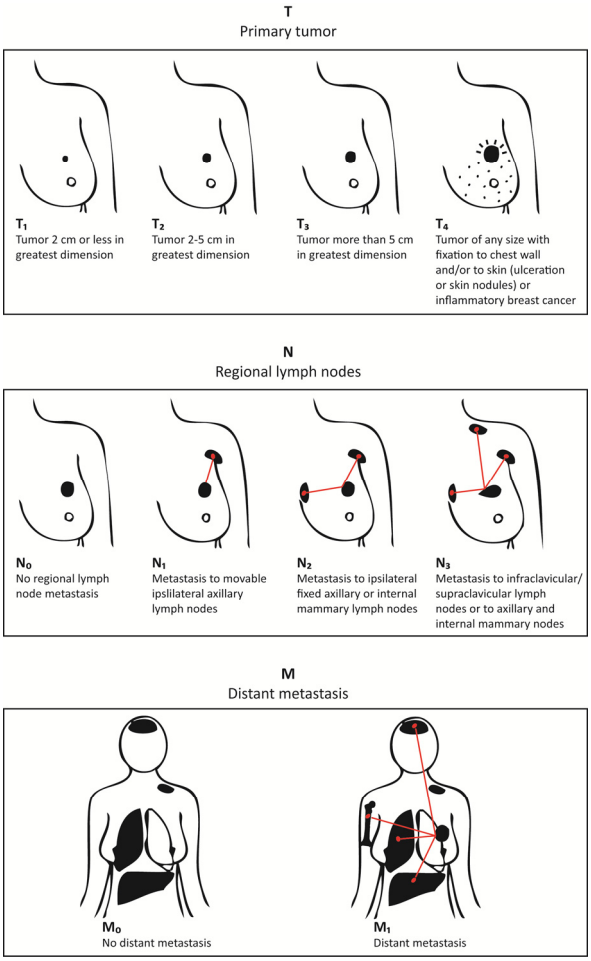

Figure 3. The TNM classification of malignant breast tumors based on clinical characteristics. Schematic illustration showing three aspects determining staging, including tumor size and the extent of metastasis to the regional lymph nodes and distant parts of the body. T_X and N_X denote that the primary tumor and regional lymph nodes (e.g., previously removed) cannot be assessed, respectively (not shown). T_{is} denotes a primary tumor in situ. Figure adapted from **http://1st-in-breastcancer.com/the-importance-of-knowing-the-stages-of-breast-cancer**.

BREAST CANCER STEM CELLS

In breast cancer, unequivocal proof of the existence of cancer stem cells (CSCs) was only first reported in 2003 (41), although the existence of CSCs within hematologic cancers and solid tumors is not a new concept (42). CSCs are relatively rare within a tumor cell population and are thought to share some of the same characteristics as normal stem cells, including self-

renewal (by symmetric and asymmetric division) and differentiation into mature cells representing all cell types within a tumor. CSCs have to be carefully isolated and purified prospectively to study their properties. To take advantage of this principle, Al-Hajj and colleagues developed a model which showed that only a small population of breast tumor cells is capable of initiating tumor formation (tumorigenic) in immunodeficient mice, while the bulk of the tumor mass is non-tumorigenic. The authors were also able to effectively distinguish tumorigenic and non-tumorigenic cells by isolating cell fractions with a flow cytometer on the basis of cell surface marker expression. They found that tumorigenic breast cancer cells could be characterized by three cell surface markers: CD44+ (adhesion molecule), CD24- (adhesion molecule), and B38.1+ (breast and ovarian cancer-specific marker). Recently, high ALDH activity has also been shown to be a good marker for CSCs and a predictor of patient clinical outcome (43). In addition, Ali *et al.* (2011) showed that a composite analysis of the CD44/CD24, ALDH1A1, ALDH1A3, and ITGA6 breast cancer stem cell markers has the most effective prognostic value. Furthermore, high ALDH activity has also been shown to be a good marker for CSCs and a predictor of patient clinical outcome (44).

Today, we have a better understanding of the biology of breast CSCs. The identification and characterization of CSCs in tumors has revolutionized cancer research and may help improve clinical treatment of cancer. A fundamental assumption is that tumors can be eradicated by targeting the tumorigenic CSC fraction. However, CSCs possess characteristics which hinder their elimination with conventional breast cancer therapies (Figure 4). CSCs are long-lived, slow growing cells that are frequently in a dormant, quiescent cell cycle state. However, conventional breast cancer therapies (surgery, chemotherapy, radiotherapy, endocrine therapy) target rapidly dividing cells which may eliminate the bulk of tumor cells, but will probably have no effect on CSCs. The majority of CSCs may be eradicated by surgery, however CSCs have several characteristics which deem them resistant to other treatment regimens: (a) CSCs express elevated levels of multi-drug resistant proteins (chemotherapy resistant), (b) CSCs express elevated levels of reactive oxygen species (ROS) scavenging (45) and DNA damage response genes (radiotherapy resistant) (46), and (c) the majority of CSCs do not express the sex hormone receptors, estrogen and progesterone (endocrine therapy resistant). By failing to eliminate CSCs, these long-lived cells will have time to accumulate additional mutations which may promote tumor recurrence and metastasis of the primary tumor.

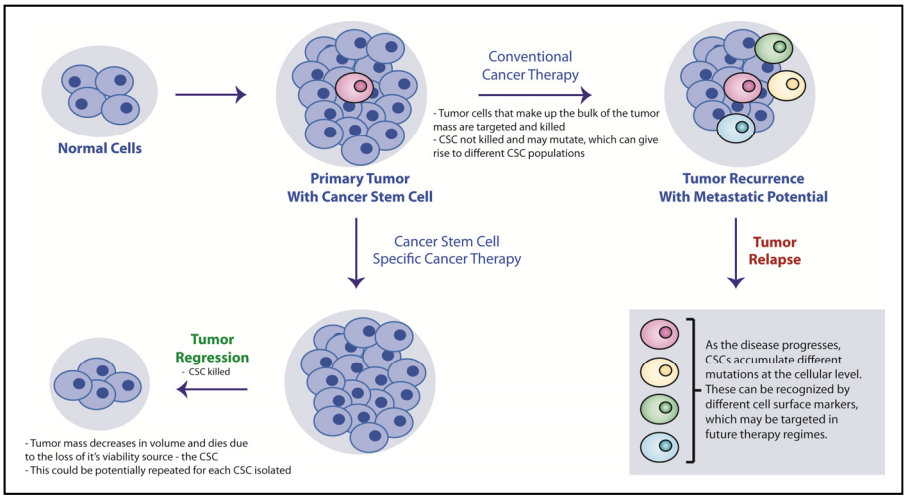

Figure 4. Schematic illustration showing the Cancer Stem Cell Theory (adapted from The European Cancer Stem Cell Research Institute).

MOLECULAR CLASSIFICATION OF BREAST CANCER

During this revolutionizing age of microarrays, our knowledge of tumor biology has increased dramatically. The last ten years has shown the introduction, refinement, and continued delineation of the major molecular subtypes of breast cancer which illustrate the heterogeneous nature of the disease and its association with clinical outcome and differences in response to treatment. This classification system, consisting of five intrinsic molecular subtypes (luminal A, luminal B, basal-like, HER2/ER-, and normal-like), was first developed by Perou *et al.* (47, 48) and Sorlie *et al.* (49), but has since been questioned for using too few tumor specimens (less than 100) to develop the signature and the relevance of the normal-like, HER2/ER-, and ER-positive (luminal A/B) subtypes (Figure 5). Recently, three subtypes of HER2-positive breast tumors were identified, each varying in clinicopathological features and clinical outcome, which may explain why many HER2-positive do not classify within the HER2/ER- molecular subtype (50). In addition, the subtypes have been associated with distinct DNA copy number aberrations and methylation patterns (51, 52).

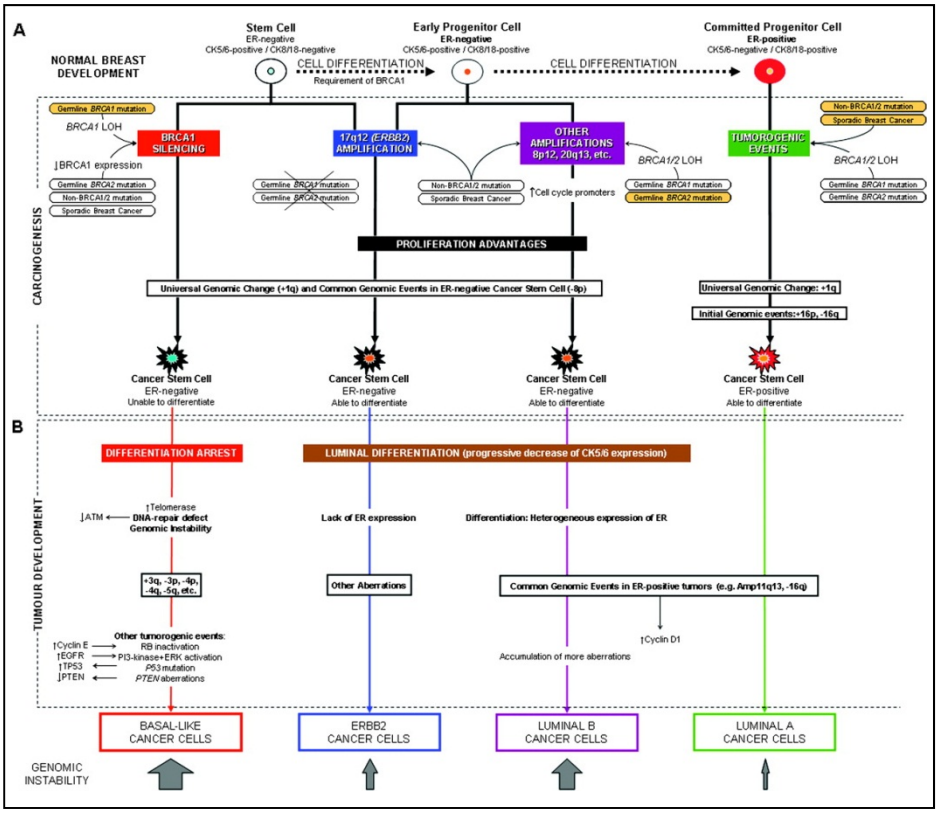

Figure 5. Integrative model of carcinogenesis and tumor development in breast CSCs. This figure was published in (53), reprinted by permission from Oxford University Press, copyright 2008. Genomic stability is represented by arrows at the bottom of the figure, where thick arrows depict high genetic instability.

The recent introduction of three novel subtypes (molecular apocrine, interferon-rich, and claudin-low) has illustrated that this field is still evolving and there may be several aspects of breast cancer tumor biology that are still poorly understood (Table 1) (54-57). Perou *et al.* and Melchor *et al.* have since extended the theory of the molecular subtypes to propose a connection between normal mammary development, the claudin-low molecular breast tumor subtype, and breast CSCs (53, 58). This theory illustrates that during normal mammary development, undifferentiated breast CSCs with mesenchymal features become differentiated myoepithelial and luminal cells: breast CSCs are enriched in the Claudin-low subtype, *BRCA1* mutations in the basal-like subtype, and differentiated myoepithelial and luminal cells in the luminal A/B subtypes. The Cancer Genome Atlas Network recently performed a comprehensive characterization of breast cancer by integrating data obtained from over 800 breast cancer patients using six different microarray platforms measuring

DNA copy number, DNA methylation, mutation status, gene expression, microRNA expression, and cancer-related protein expression (59). A consensus between the platforms revealed four major subtypes which were not only highly correlated with hierarchal clustering of the gene and protein expression data, but also correlated well with the previously reported subtypes (luminal A, luminal B, basal-like, HER2/ER-).

Table 1. Characteristics of the molecular breast cancer subtypes

Molecular subtype	Characteristics	Prevalence	Prognosis
Luminal A	ER+ and/or PgR+, HER2-, low Ki67, express high amounts of luminal cytokeratins, and well differentiated	42-59%	Favorable
Luminal B	ER+ and/or PgR+, HER2+ (or HER2- with high Ki67), express high amounts of luminal cytokeratins, and poorly differentiated	6-19%	Intermediate/ unfavorable
Basal-like	ER-, PgR-, HER2-, cytokeratin 5/6+, EGFR+, c-KIT+, express high levels of growth factors such as HGF and IGF, and/or HER1+	14-20%	Unfavorable
HER2/ER-	ER-, PgR-, HER2+	7-12%	Unfavorable
Claudin-low	ER-, PgR-, HER2-, Claudin 3/4/7-, low E-cadherin		
Molecular apocrine	ER-, AR+, FOXA1+, HER2+, PIP+	8-14%	Unfavorable

Abbreviations: ER+ = Estrogen receptor-positive; ER- = Estrogen receptor-negative; PgR+ = Progesterone receptor-positive; PgR- = Progesterone receptor-negative; HER2+ = HER2/*neu*-positive; HER2- = HER2/*neu*-negative; AR+ = Androgen receptor-positive; PIP+ = prolactin-inducible protein-positive.

RISK ASSESSMENT AND TREATMENT OPTIONS

Guidelines for standardized breast cancer treatment have been developed by several professional organizations in the U.S. and Europe, including The American Society of Clinical Oncology and the St. Gallen International Expert Consensus (60). In general, two main questions are considered when deciding which patients should receive adjuvant systemic therapy. First, does the patient have a high risk of recurrence and breast cancer-related death? Second, does systemic therapy reduce this risk? Currently, a one-size-fits-all principle is applied to breast cancer treatment. Following breast conserving surgery or mastectomy, adjuvant therapy is determined using patient characteristics (age and/or menopausal status) and tumor

characteristics (TNM status, histological grade, hormone receptor status, HER2/*neu* receptor status). To reduce the risk of recurrence, adjuvant therapy is administered combining chemotherapy, radiotherapy, endocrine treatment (Tamoxifen or aromatase inhibitors), or Trastuzumab (Herceptin) therapy. Similar treatment regimens are administered to patients exhibiting similar "characteristics". However, two patients with similar "characteristics" may also respond differently to the same treatment. Therefore, three aspects of current breast cancer therapy are insufficient. First, patient stratification and risk assessment needs to be improved to determine which patients would most benefit from adjuvant therapy to reduce tumor burden. Second, targeted therapy is also needed to actively mitigate tumor expansion and spread without affecting surrounding normal tissue function, such as treatment of HER2/*neu*-positive breast cancer with Trastuzumab. Lastly, individualized treatment regimens need to be applied in order to select the right treatment for the right patient, at the right time and at the right dose.

During the past decade, molecular profiling of breast carcinoma has been used to illustrate and describe tumor heterogeneity. A few of these genetic signatures are listed in Table 2. However, biomarkers identified using these methods are seldom accepted for clinical decision making. For example, among the multigene assays Oncotype DX™, Mammaprint®, and uPA/PAI-1, only Oncotype DX™ is commonly accepted by clinicians (61). Nevertheless, treatment of early invasive breast cancer is still evolving; the St. Gallen conferences strive to establish treatment recommendations including clinical parameters showing reproducible results, but also an increasing number of tumor biology-based biomarkers, e.g. endocrine and HER2-targeted therapy, intrinsic subtypes, mutational analysis, biomarkers targeting frequently perturbed signaling pathways.

Table 2. Multigene signatures for breast carcinoma

Assay description	Biomarkers (n)	Method and sample conditions	Clinical use	Biomarkers in the signature in common with this dissertation*
MAMMAPRINT® (62)				
FDA approved assay for pN0 breast cancer patients of all ages with tumors less than 5 cm and either ER- or ER+ to predict risk for metastasis and determine which patients will benefit from chemotherapy	70	Microarray, fresh/frozen or formalin-fixed tissue	Tamoxifen, adjuvant and neoadjuvant chemotherapy	SCUBE2, STK32B
CLINICAL TREATMENT SCORE				
An algorithm which consisting of axillary lymph node status, tumor size, histological grade, age, and treatment	N/A	N/A		N/A
IHC4 (63)				
A four-marker signature (ER, PgR, HER2, Ki67) which provides independent prognostic information after adjusting for established clinical variables	4	IHC, FFPE		None
ONCOTYPE Dx™ RECURRENCE SCORE (64)				
Signature used to calculate a disease recurrence score in early-stage ER-positive breast cancer	21	qPCR, FFPE	Tamoxifen, Adjuvant, CMF	SCUBE2
PREDICTOR ANALYSIS OF MICROARRAY (PAM50) (65, 66)				
Signature to classify breast tumors into the intrinsic molecular subtypes	50	qPCR, fresh/frozen	Neo-adjuvant chemotherapy	UBE2C
BREAST CANCER INDEX (67)				
Signature combining the molecular grade index and *HOXB13:IL17BR* to identify a subgroup of early-stage ER-positive breast cancer patients with an unfavorable prognosis despite adjuvant endocrine therapy	5	qPCR, FFPE		None
ENDOPREDICT (68)				

Assay description	Biomarkers (n)	Method and sample conditions	Clinical use	Biomarkers in the signature in common with this dissertation*
Signature to predict distant recurrence in ER-positive, HER2/*neu*-negative breast cancer treated with adjuvant endocrine therapy	8	qPCR		AZGP1, STC2, UBE2C
GENOMIC GRADE INDEX (69, 70)				
Signature to define histological grade (high or low genomic grade) in estrogen receptor-positive breast cancer using molecular profiling.	97	Microarray, fresh/frozen or formalin-fixed tissue		NME5, UBE2C
ROTTERDAM 76-GENE SIGNATURE (71)				
Signature to predict distant recurrence in axillary lymph node-negative breast cancer patients	76	Microarray, fresh/frozen or formalin-fixed tissue		None
WOUND RESPONSE SIGNATURE (72)				
Signature depicting a wound healing response in fibroblasts from ten anatomical regions after serum exposure	446	Microarray, fresh/frozen or formalin-fixed tissue		
INVASIVENESS GENE SIGNATURE (73)				
Signature that can differentiate highly tumorigenic CD44+CD24-/low cells from normal breast epithelium	186	Microarray, fresh/frozen or formalin-fixed tissue		STC2

*Note: Gene targets discussed in this dissertation.
Abbreviations: CMF = Cyclophosphamide, methotrexate and 5-fluorouracil; FFPE = Formalin-fixed, paraffin-embedded; IHC = Immunohistochemistry; qPCR = Quantitative real-time PCR; N/A = Not applicable.

Specific aims

The main objectives of this doctoral thesis have been to study…

PAPER I

To identify recurrent DNA copy number alterations and gene expression patterns associated with diploid breast tumor malignancy and clinical outcome

PAPER II

To assess the clinical significance of the 12 novel candidate biomarkers (AZGP1, CBX2, DNALI1, NME5, PIP, S100A8, SCUBE2, SERPINA11, STC2, SUSD3, STK32B, UBE2C) in breast carcinoma

PAPER III

To assess whether the candidate biomarkers (AZGP1, BTG2, CBX2, CNTNAP2, DNALI1, NME5, PIP, S100A8, SCUBE2, SERPINA11, STC2, SUSD3, STK32B, UBE2C, WHSC1L1) are breast cancer-specific or general key cancer genes

PAPER IV

To identify putative oncogenes located within the 8p11-p12 amplification region whose amplification and/or over-expression promote an aggressive form of breast carcinoma

Patients, tumor specimens, and cell lines

All procedures were performed in accordance with the Declaration of Helsinki and approved by the Medical Faculty Research Ethics Committee (Gothenburg, Sweden).

PAPER I

Of the 2,541 breast carcinoma specimens available (diagnosed between 1991 and 1999 in Western Sweden) in the fresh-frozen tissue tumor bank at the Sahlgrenska University Hospital Oncology Lab (Gothenburg, Sweden), 97 diploid primary invasive breast cancers (DBC) fulfilled the criteria for inclusion in the study as shown in Figure 6. The patients were then stratified according to axillary lymph node status (node-negative, pN0; node-positive, pN1) and breast cancer-specific survival (>8-year survivors, henceforth termed long-term survivors; breast cancer-specific mortality within 8 years of diagnosis, henceforth termed short-term survivors). Disease-specific survival (DSS) for breast cancer was defined as the time from initial diagnosis to breast cancer-related death. Three of the four groups used in the study consisted of 25 DBC cases each (1. pN0, long-term survivors, 2. pN1, long-term survivors, 3. pN1, short-term survivors), while the remaining group (pN0, short-term survivors) consisted of 22 DBC cases.

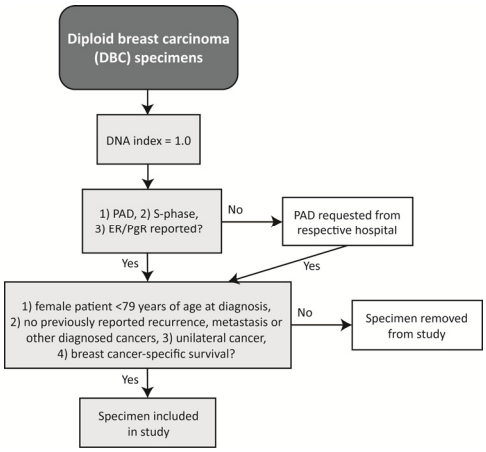

Figure 6. Criteria for inclusion in study for Paper I. The flowchart shows the criteria for inclusion of DBC specimens in the study.

PAPER II

Ninety-four formalin-fixed, paraffin-embedded tissues (FFPE) were obtained from the Department of Pathology at Sahlgrenska University Hospital. The tissue specimens corresponded to 71 of the 97 diploid breast cancer patients used in **paper I**. In addition, transcriptome data from six publicly available datasets consisting of 1,141 breast cancer patient samples were used to validate the 13-marker signature identified in **paper I**. Breast cancer survival rates were defined as a) the time from initial diagnosis to breast cancer-related death for DSS, b) time from initial diagnosis to first relapse (loco-regional or distant) or breast cancer-related death for disease-free survival (DFS), c) time from initial diagnosis to death from any cause for overall survival (OS); d) time from initial diagnosis to distant metastasis for distant metastasis-free survival (DMFS), and e) time from operative lesion removal to detection of tumor recurrence for recurrence-free survival (RFS).

PAPER III

A retrospective study was conducted using 43 consecutive cases of oral squamous cell carcinoma (OSCC) with clinically negative neck lymph nodes (pN0) at the time of diagnosis between 1997-2004 at Sahlgrenska University Hospital in Gothenburg, Sweden. All patients underwent diagnostic battery inclusive biopsy of the primary tumor, palpation of the neck, radiological examination with MRT and/or CT, and classified according to the American Joint Committee on Cancer (AJCC) staging system. Surgical excision of the primary tumor and supraomohyoidal neck dissection (SOHND) were performed, followed by examination of the lymph nodes with hematoxylin and eosin (H&E) staining. This analysis showed that 11/43 patients had positive cervical lymph nodes (pN+), all of whom received post-operative radiotherapy to the neck. Thirty-two patients were still considered neck negative (pN0) and followed up clinically. Forty-five FFPE samples corresponding to the 43 patients were obtained from the Department of Pathology at Sahlgrenska University Hospital. Histological classification and staging of the tumor specimens were performed according to the WHO classification and International Union Against Cancer (UICC), respectively.

PAPER IV

Fresh-frozen (n = 229) and FFPE (n = 127) invasive breast carcinoma specimens corresponding to 185 and 95 patients were obtained from the Oncology Lab and Department of Pathology at Sahlgrenska University Hospital (Figure 7), respectively. These samples included the 97 fresh-frozen samples described in **paper I** and the 94 FFPE samples described in **paper II**.

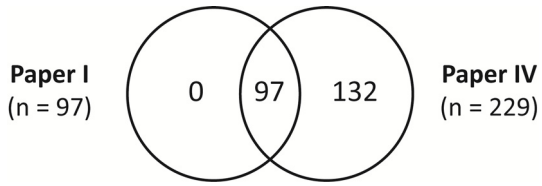

Figure 7. Fresh-frozen primary invasive breast carcinomas (n = 229) derived from patients diagnosed between 1991-1999 in Western Sweden were analyzed for genetic alterations in relation to clinicopathological parameters, as described in papers I and IV.

Human breast cancer cell lines (BT-20, BT-474, BT-483, CAL-120, CAMA-1, HCC38, Hs578T, JIMT-1, MCF-7, MDA-MB-134-VI, MDA-MB-231, MDA-MB-361, MDA-MB-436, MDA-MB-468, T47D, ZR-75-1), the non-tumorigenic breast epithelial cell line (MCF-10A), the S68 breast cancer cell line kindly provided by Dr. Verónique Catros, and the human HEK293T cell line were cultured as described in Table 3.

Table 3. Human cell lines and culture conditions

	Origin[§]	Cancer type[†]	ER	PgR	HER2	Culture conditions[‡]
BREAST CANCER CELL LINES						
BT-20[a]	P	IDC (74)	-	-		DMEM, 10% FBS, 1% NEAA
BT-474[a]	P	IDC (75)	+	+	+	RPMI 1640, 10% FBS
BT-483[a]	P	IDC (75)	+	+		RPMI 1640, 20% FBS, 10µg/ml bovine insulin, 1% NaPyr
CAL-120[b]	M (Pl)	AC				DMEM, 10% FBS
CAMA-1[a]	M (Pl)	C	+	-		DMEM, 10% FBS
HCC38[c]		(76)	-	-		RPMI 1640, 10% FBS
Hs578T[a]	P	CS (77)	-	-		DMEM, 10% FBS, 10µg/ml bovine insulin
JIMT-1[b]	M (Pl)	IDC (78)				DMEM, 10% FBS
MCF-7[a]	M (Pl)	IDC	+	+		DMEM, 10% FBS
MDA-MB-134-VI[a]	M (Pl)	IDC (79)	+	-		L-15, 20% FBS
MDA-MB-231[a]	M (Pl)	IDC (79)	-	-		DMEM, 10% FBS
MDA-MB-361[a]	M (Br)	AC (79)	+	-	+	DMEM, 10% FBS
MDA-MB-436[a]	M (Pl)	AC (80)	-	-		L-15, 10% FBS
MDA-MB-468[a]	M (Pl)	AC (80)	-	-		L-15, 10% FBS
S68[d]	M (Pl)					RPMI 1640, 2mM L-Glutamine, 100IU/ml penicillin, 100µg/ml

	Origin[§]	Cancer type[†]	ER	PgR	HER2	Culture conditions[‡]
						Streptomycin, 1mM NaPyr, 10mM Hepes, 10% FBS
T47D[a]	M (Pl)	IDC (81)	+	+		RPMI 1640, 10% FBS
ZR-75-1[a]	M (As)	IDC (82)	+	-		RPMI 1640, 10% FBS, 5µg/ml bovine insulin, 200uM L-Glutamine, 1% PEST
KIDNEY CELL LINE						
HEK293T[e]		AC				
NON-TUMORIGENIC CELL LINE						
MCF-10A[a]		FD (83)	-	-		RPMI 1640, 10% FBS, 5µg/ml bovine insulin, 1mM L-Glutamin, 1% PEST

Source: [a]Department of Oncology, Gothenburg University; [b]DSMZ, Braunschweig, Germany; [c]American Type Culture Collection (ATCC), Manassas, VA, USA; [d]Gift from V. Catros; [e]GeneCopoeia, Rockville, MD, USA

[§]P: primary; M: metastasis; Pl: pleural effusion; As: ascites; Br: brain.

[†]IDC: invasive ductal carcinoma; C: carcinoma; FD: fibrocystic disease; AC: adenocarcinoma; OSCC: oral squamous cell carcinoma; TAC; thyroid anaplastic carcinoma.

[‡]Cells cultured at 37°C, 5% CO_2.

Nucleic acid isolation

Nucleic acids (DNA and total RNA) were isolated from 10-20 mg sections of fresh-frozen tumor specimens. The presence of malignant cells was assessed in all samples by evaluation of touch preparation imprints stained with May-Grünwald Giemsa (Chemicon). Highly representative specimens containing more than 70% neoplastic cell content were included in the analyses.

DNA ISOLATION

In **papers I and IV**, genomic DNA (gDNA) was isolated from fresh-frozen tumor sections using the Wizard Genomic DNA extraction kit (Promega), including proteinase K treatment (Roche) followed by phenol-chloroform purification (Sigma). The DNA concentration was measured using a Nanodrop ND-1000 (Nanodrop Technologies). The gDNA samples were stored at 4°C.

RNA ISOLATION

In **papers I and IV**, fresh-frozen tumor sections were homogenized into a fine powder using the Mikro-Dismembrator S ball mill (Sartorius Stedim Biotech), followed by isolation of total RNA using the RNeasy Lipid Tissue

Mini Kit (Qiagen) according to the manufacturer's instructions with minor modifications. The RNA concentration was measured using Nanodrop and RNA integrity assessed using the RNA 6000 Nano LabChip Kit with Agilent 2100 Bioanalyzer (Agilent Technologies). RNA quality and integrity are critical to obtaining reliable results in RNA-dependent downstream experiments. The Agilent 2100 Bioanalyzer Expert software generates the RNA Integrity Number (RIN), an estimate of RNA quantification, and a calculation of the ribosomal ratios using only 1 µl of each total RNA sample. As previously determined for frozen breast tumor samples, a RIN threshold of at least 6.0 was used (84). Total RNA samples with RIN values lower than this threshold were discarded and reisolated.

Protein isolation

In **papers II and IV**, total proteins were isolated from 10-20 mg sections of fresh-frozen tumor specimens and/or cultured cells. Fresh-frozen tissue samples were homogenized in Mammalian Cell Lysis Buffer including Benzonase® Nuclease and Protease Inhibitor Solution using the TissueLyser ball mill (Qiagen) while cell pellets were simply resuspended in lysis buffer, followed by protein isolation using the Qproteome Mammalian Protein Prep Kit (Qiagen) according to the manufacturer's instructions. The supernatant was transferred to a new 1.5 ml microcentrifuge tube and placed on dry ice until use or stored at -80°C. The protein concentration was measured on a spectrophotometer using the Bradford assay and Bio-Rad Protein Assay Dye Reagent Concentrate (Bio-Rad).

Microarrays

ARRAY COMPARATIVE GENOMIC HYBRIDIZATION

In **papers I and IV**, whole-genome tiling arrays with 38,043 BAC reporters mapping to the UCSC May 2004 hg17: NCBI Build 35 were manufactured as previously described (85) at the SCIBLU Genomics DNA Microarray Resource Center (SCIBLU), Department of Oncology, Lund University. The clone set consisted of the 32K BAC clone library (BacPac Resources), the 3.4K FISH Mapped Clones Version 1.3 (BacPac Resources), clones located in telomeric regions (86), and clones covering microdeletion syndromes (87). Images and raw signal intensities were acquired using an Agilent G2505B DNA microarray scanner (Agilent Technologies) and GenePix Pro 6.0.1.22 (Axon Instruments) image analysis software.

TRANSCRIPTIONAL ANALYSIS

In **paper I**, the 97 total RNA samples were processed at SCIBLU using Illumina HumanHT-12 Whole-Genome Expression BeadChips (Illumina),

according to the manufacturer's instructions. In **paper IV**, total RNA samples from 150/229 tumor specimens (45 specimens harboring 8p11-p12 amplification, 18 specimens with 8p11-p12 low-level gain, 16 specimens with 8p11-p12 heterozygous loss, 71 specimens lacking 8p11-p12 aberrations) and 12 cell lines were processed at SCIBLU using Illumina HumanHT-12 BeadChips. The 150 samples include the 97 samples described in **paper I**. The expression microarrays contained approximately 49,000 probes representing more than 25,400 RefSeq (Build 36.2, Release 22) and Unigene (Build 199) annotated genes. Images and raw signal intensities were acquired using the Illumina BeadArray Reader scanner and BeadScan 3.5.31.17122 (Illumina) image analysis software, respectively.

In **paper II**, six publicly available Affymetrix U133A gene expression microarray datasets (n = 1,141) were used to validate the 13-marker signature (*AZGP1, CBX2*, *DNALI1*, *LOC389033*, *NME5*, *PIP*, *S100A8*, *SCUBE2*, *SERPINA11*, *STC2*, *STK32B*, *SUSD3,* and *UBE2C*) described in **paper I**.

DNA METHYLATION ANALYSIS

In **paper IV**, 22/229 tumor specimens (11 specimens harboring 8p11-p12 amplification and 11 specimens lacking 8p11-p12 aberrations) were processed at SCIBLU using Illumina Infinium Human Methylation 450 Beadchips. The beadchips contained more than 485,000 specific cytosine sites representing more than 99% of RefSeq annotated genes (Build 37), 96% of CpG islands, and sites flanking CpG islands (up to 4kb on either side of CpG islands).

Immunohistochemistry

Optimal antibody dilutions and assay conditions were achieved for immunohistochemistry using breast carcinoma (**papers II and IV**) or oral squamous cell carcinoma (**paper III**) as positive controls. Four micrometer FFPE sections were subsequently immunostained with primary antibodies (Table 4). The sections were pretreated using the Dako PTLink system (Dako, Carpinteria, CA, USA) and processed on an automated Dako Autostainer platform using the Dako Envision™ FLEX High pH Link Kit (pH 9). Peroxidase-catalyzed diaminobenzidine was used as the chromogen, followed by hematoxylin counterstain. The slides were then rinsed with deionized water, dehydrated in absolute alcohol, followed by 95% alcohol, cleared in xylene, and mounted. One FFPE section was also stained with H&E to assess the degree of inflammatory infiltration and to determine histological grading (according to the Nottingham (Elston-Ellis) modification of the Scarff-Bloom-Richardson grading system).

Table 4. Primary antibodies for immunohistochemistry

	Host	Dilution‡	Sources	Study
PRIMARY ANTIBODIES				
Anti-AZGP1 (Cat. HPA012582)	Rabbit	1:500, 1:250	Sigma-Aldrich	II, III
Anti-BTG2 (Cat. 18-003-42396)	Rabbit	1:100	GenWay Biotech	III
Anti-CBX2 (Cat. AV51628)	Rabbit	1:250, 1:150	Sigma-Aldrich	II, III
Anti-CNTNAP2 (Cat. ab104525)	Rabbit	1:300	Abcam	III
Anti-DNALI1 (Cat. ab58213)	Mouse	1:25	Abcam	II
Anti-GCET1* (Cat. ab86673)	Rabbit	1:50, 1:25	Abcam	II, III
Anti-NME5 (Cat. 12923-1-AP)	Rabbit	1:400, 1:300	ProteinTech Group	II, III
Anti-p16 (Cat. sc-56330)	Mouse	1:200	Santa Cruz Biotechnology	III
Anti-PIP (Cat. HPA009177)	Rabbit	1:25	Sigma-Aldrich	II
Anti-S100A8 (Cat. HPA024372)	Rabbit	1:2000	Sigma-Aldrich	II, III
Anti-SCUBE2 (Cat. HPA006353)	Rabbit	1:25, 1:100	Sigma-Aldrich	II, III
Anti-STC2 (Cat. 10314-1-AP)	Rabbit	1:500, 1:100	ProteinTech Group	II, III
Anti-STK32B (Cat. HPA015820)	Rabbit	1:50, 1:300	Sigma-Aldrich	II, III
Anti-SUSD3 (Cat. HPA042310)	Rabbit	1:100, 1:150	Sigma-Aldrich	II, III
Anti-UBE2C (Cat. ab56861)	Mouse	1:500	Abcam	II, III
Anti-WHSC1L1 (Cat. HPA018893)	Rabbit	1:200, 1:100	Sigma-Aldrich	III, IV

Note: * GCET1 is an alias for the SERPINA11 protein; ‡ Occasionally, the same antibody was used for several studies. If the same antibody dilution was used for multiple studies, then one only dilution is given. If different antibody dilutions were used for different studies, then the first dilution is given followed by a comma and the second dilution for the first and second studies, respectively.

Fluorescence *in situ* hybridization

In **paper I**, dual-color FISH was performed on touch preparations using the Zyto*Light* SPEC HER2/CEN 17 Dual Color Probe (ZytoVision GmbH), according to the manufacturer's instructions (directly labeled probe). In **paper IV**, dual-color FISH was performed on metaphase chromosomes and touch preparations using probes prepared from a contig of 42 overlapping BAC clones (indirectly labeled probes) to validate gene amplification observed on chromosome band 8p11-p12 using array-CGH.

Touch preparations consisting of interphase cells were prepared by pressing fresh cuts of frozen tumor specimens onto a Kim-wipe to remove excess liquid and then onto SuperFrost plus charged microscope slides (Erie Scientific Company). Metaphase chromosome preparations were prepared from cultured peripheral blood cells and cancer cell lines on SuperFrost microscope slides according to standard protocols. The hybridized slides were analyzed using a Leica DMRA2 fluorescent microscope (Leica) equipped with an ORCA Hamamatsu CCD (charged-couple devices) camera

(Hamamatsu City, Japan) and filter cubes specific for FITC, Rhodamine, and UV for DAPI visualization. Digitalized black and white images were captured using the Leica CW4000 software package (Cambridge, UK).

In order to distinguish cells displaying aneuploidy from gene amplification, only gene copy number increases equal to or higher than 2.5-fold were scored as amplification, corresponding to >5 copies/diploid chromosome set equivalent. At least 100 nuclei with intact morphology on the basis of DAPI counterstaining were scored from each clinical specimen. Clumped, damaged, overlapping, and nuclei located outside the area of probe hybridization were excluded from the analysis. If less than 100 cells were present on the slide, then as many nuclei as possible were evaluated. To determine the gene copy number and the pattern of gene amplification, a gene amplification classification system was applied where specimens were classified as *low level amplification* when the number of specific hybridization signals varies from 5 to 10 signals, *moderate amplification* when the number of signals varies from 10 to 20, and as *high level amplification* when the number of signals is > 20 (88). An average DNA copy number were calculated taken in consideration of the clonal heterogeneity seen in the analyzed specimens.

Quantitative real-time polymerase chain reaction

Validation of the expression microarray data was performed using qPCR with pre-designed TaqMan® Gene Expression Assays (Applied Biosystems). To prevent any interference by residual gDNA contamination in the qPCR application, TaqMan assays were primarily chosen which annealed to intron splice junctions so that gDNA would not be amplified. Three endogenous controls, *i.e. PPIA*, *PUM1*, and *HPRT1*, were selected based on their constitutive expression using the Illumina HumanHT-12 platform. In **paper I**, 82/97 total RNA samples were used to validate the expression patterns of 16 transcripts, *i.e. AFF3*, *AGR3*, *CA12*, *DNAJC12*, *ESR1*, *GATA3*, *GLYATL2*, *PIP*, *S100A8*, *S100A9*, *S100P*, *SCUBE2*, *STC2*, *TAT*, *TFF1*, and *TFF3*. In **paper IV**, 77/83 total RNA samples from tumors harboring 8p11-p12 aberrations and 7/71 samples lacking aberrations were used to validate the expression patterns of 11 transcripts, *i.e. ASH2L*, *BAG4*, *BRF2*, *DDHD2*, *LSM1*, *PPAPDC1B*, *PROSC*, *RAB11FIP1*, *TM2D2*, *WHSC1L1*, and *ZNF703*. The geometric mean of the three endogenous controls was used to normalize the data and relative gene expression levels were calculated with the relative standard curve method. The Student's t-test was calculated to

determine the difference in expression between studied groups and the Spearman correlation coefficients (two-tailed) to establish the relationship between microarray and qPCR expression patterns.

Immunoblot

In **papers II and IV**, extracts (50 µg) were resolved by SDS-PAGE on 4-12% Bis-Tris gels (Invitrogen), transferred to nitrocellulose membranes (Fisher Scientific), probed with primary and secondary antibodies (Table 5), and detected using the SuperSignal® West Femto Maximum Sensitivity Substrate (Thermo Scientific). Digitalized images were acquired using Fujifilm Luminescent Image Analyzer LAS-4000 and analysis performed with the Image Gauge v4.0 software.

Table 5. Primary and secondary antibodies for immunoblot

	Host	Dilution‡	Sources	Study
PRIMARY ANTIBODIES				
Anti-ACTB (Cat. ab6276)	Mouse	1:2000	Abcam	II, IV
Anti-AZGP1 (Cat. HPA012582)	Rabbit	1:500	Sigma-Aldrich	II
Anti-PIP (Cat. HPA009177)	Rabbit	1:200	Sigma-Aldrich	II
Anti-S100A8 (Cat. HPA024372)	Rabbit	1:1000	Sigma-Aldrich	II
Anti-UBE2C (Cat. ab56861)	Mouse	1:200	Abcam	II
Anti-NSD3 (Cat. ab4514)	Rabbit	1:500	Abcam	IV
SECONDARY ANTIBODIES				
ECL anti-mouse IgG HRP-linked Whole Ab (Cat. NA931V)	Sheep	1:2000	Amersham	II, IV
ECL anti-rabbit IgG HRP-linked Whole Ab (Cat. NA934V)	Donkey	1:2000	Amersham	II, IV

‡ Occasionally, the same antibody was used for several studies. If the same antibody dilution was used for multiple studies, then one only dilution is given. If different antibody dilutions were used for different studies, then the first dilution is given followed by a comma and the second dilution for the first and second studies, respectively.

Lentiviral vector-based stable transfection

In **paper IV**, long-term stable knockdown or up-regulation of individual candidate genes were performed using short hairpin (shRNA) expression constructs (containing a seven base loop and 19-29 base stem) or open reading frame (ORF) cDNA clones (derived from full-length human genes but lacking the 5' or 3'UTRs), respectively. For a specific target gene, one shRNA clone set was used containing four separate shRNA expression constructs to ensure high knockdown efficiency with minimal off-target effects. Depending on the construct, the puromycin marker was used to enable stable cell line selection and the eGFP reporter gene for monitoring transfection efficiency.

Lentiviral particles were produced in the 293Ta packaging cell line with the Lenti-Pac HIV Expression Packaging Kit (GeneCopoeia), and harvested and filtered 48 h post transfection. Target cells ($10x10^4$) were infected in 24-well plates for 24 h at 37°C with complete medium supplemented with lentiviral particles and polybrene (Sigma-Aldrich). Stable cell selection was performed by maintaining cultures in 1 µg/ml puromycin (Sigma-Aldrich). Transduction efficiency and knockdown of target protein expression were determined by GFP imaging and immunoblot analysis, respectively. Downstream analyses were performed using stably transduced cells.

Table 6. Lentiviral short hairpin RNA (shRNA) and open reading frame (ORF) cDNA expression clone sets

	Promoter	Selection Marker	Reporter Gene	Vector	Viral Type	Study
shRNA clone set						
Human WHSC1L1-S (Cat. HSH013847-HIVU6)	U6	Puromycin	eGFP	psiHIV-U6	HIV	IV
Human WHSC1L1-L (Cat. HSH013848-HIVU6)	U6	Puromycin	eGFP	psiHIV-U6	HIV	IV
ORF cDNA expression clone set						
Human WHSC1L1-S (Cat. EX-X0597-Lv105)	CMV	Puromycin		pReceiver-Lv105		IV
Human WHSC1L1-L (Cat. EX-Z7852-Lv105)	CMV	Puromycin		pReceiver-Lv105		IV

Note: The lentiviral clone sets were purchased from GeneCopoeia, Inc.

Functional assays

In **paper IV**, the effects of individual candidate gene knockdown and/or overexpression cell proliferation (WST-1 assay), cell cycle phase distribution (flow cytometry), apoptosis (TUNEL assay), cell invasion (basement membrane invasion assay), anchorage-independency (agar growth assay), and colony formation were examined using stably transduced clones. All experiments were performed in triplicate.

Computational analysis

ARRAY-CGH ANALYSIS

In **papers I and IV**, data preprocessing and pin-based Lowess normalization were performed using the web-based BioArray Software Environment system (BASE, http://base2.thep.lu.se/onk/) provided by SCIBLU (89). In BASE, spot intensities representing the tumor (cy3) and reference (cy5)

channels were background-corrected by subtracting the median background from the median foreground signal intensities. Spot filtering was performed to remove unreliable features flagged as bad, absent, or not found during image analysis, not showing signal-to-noise ratios ≥ 5 for both channels, smaller than 50 µm in diameter, and features displaying saturated signal intensities over 65,000. Following filtering, 35,718 reporters remained for use in downstream analyses. Finally, reporters were normalized using an implementation of a pin-based Lowess algorithm (90) with a sliding window of 0.33, followed by smoothing and spike analysis to remove outliers and already merged replicates. Further analysis to segment the data into regions of gains and losses was performed using the Rank Segmentation algorithm with Nexus Copy Number Professional 4.1 software (BioDiscovery), as previously described (91). In brief, the basic Nexus settings suitable for BAC arrays were as follows: 5.0E-5 significance threshold, 1000 kb maximum contiguous probe spacing, and a minimum of 5 probes per segment. Log_2ratio thresholds for low-level gain, high-level gain/amplification, heterozygous loss, and homozygous deletion were set at +0.2, ≥+0.5, -0.2, and ≤-1.0, respectively, using a 0.01 p-value cutoff. Genomic regions covered entirely by previously reported copy number variations (CNV) in the human genome were removed (92). Minimal common regions of copy number imbalances were identified when observed in at least 25% of the tumor samples, while regions of high-level gain were identified when observed in at least 10% of the tumor samples using $P<0.01$. Differential chromosomal regions distinguishing patient groups were defined at $P<0.01$ (Fisher's exact test, two-tailed). Unsupervised hierarchical clustering was applied using complete linkage with Pearson correlation to group tumors based on detected genetic alterations (low-level gain and heterozygous loss). The association between clinicopathological features and CNAs was analyzed using a 0.05 p-value cutoff with the Mann–Whitney U or Kolmogorov-Smirnov test, as appropriate.

TRANSCRIPTIONAL ANALYSIS

In **papers I and IV**, data preprocessing and quantile normalization were applied to the raw signal intensities using BASE. Each gene is represented by about 30 beads on the Illumina HumanHT-12 Expression Beadchips. In the scanner, a local background correction was performed for each bead to remove background from the foreground. A bead was removed, if it was an outlier in comparison with the others for a specific gene. The remaining beads for a gene were averaged together to produce the Mean GRN value used for quantile normalization in BASE. Further data processing was

performed in Nexus Expression 2.0 (BioDiscovery) using log2-transformed, normalized expression values and a variance filter. Normalized values from five normal breast samples profiled with Illumina HumanWG-6 Expression Beadchips (GEO, accession number GSE17072) were used as reference (93). Differentially expressed genes were determined using the Benjamini-Hochberg method (94) to control for the false discovery rate (FDR) with FDR-corrected p-values <0.01.

In **papers I and IV**, the datasets were stratified into the molecular breast cancer subtypes using the five centroids (Normal-like, Basal-like, Luminal subtype A, Luminal subtype B, and HER2/ER-) and Genomic Grade Index (GGI; low, high) using ER-positive tumors as previously described (56, 69, 70). In brief, the centroids described by Hu *et al.* were downloaded from https://genome.unc.edu/pubsup/breastTumor/data/306genes-X-249samples-X-5subtypes+5centroids.xls and 289 Illumina probes were identified which matched the 306 probes used in the centroids. Several probes representing different splicing variants of a gene are contained on the Illumina platform. Therefore, one Illumina probe was selected which matched each intrinsic gene (preferably probes targeting all splice isoforms of a gene). Relative gene expression levels (normalized to normal breast tissue) for the 289 Illumina probes were used to assign the samples to the centroid with the highest positive Pearson correlation (P<0.05).

DNA METHYLATION ANALYSIS

In **paper IV**, data preprocessing and differential DNA methylation analysis were applied to the raw signal intensities using the IMA package in R/Bioconductor with thresholds set at: ≥±0.14 delta beta value and Bonferroni adjusted P<0.05 (95). Cytosine sites located on the Y chromosome or containing SNPs were removed. The estimated methylation level for specific cytosine sites (average beta, AVB) was calculated as a ratio between the intensities of methylated and unmethylated alleles and ranged from 0 (null methylated) to 1 (completely methylated). Delta beta values were calculated using ($AVB\ values_{8p11\text{-}p12\ amplified\ tumors} - AVB\ values_{8p11\text{-}p12\ non\text{-}amplified\ tumors}$).

INTEGRATIVE GENOMICS ANALYSIS

Illumina HumanHT-12 probe nucleotide sequences were mapped to genomic locations (NCBI Build 35) using sequences downloaded from the UCSC Genome Browser (96) at http://hgdownload.cse.ucsc.edu/goldenPath/hg17/chromosomes. A pair-wise comparison of the Illumina probe and BAC clone nucleotide sequences

was then conducted to generate Illumina-BAC probe pairs with 100% sequence similarity. In **paper I**, Illumina-BAC probes spanning the recurrent aberrations were selected from smoothed array-CGH data using $\log_2$ratio ±0.2. CNA-induced genes were assessed by Pearson correlation (Q<0.05). Further, the change in mRNA levels as a function of DNA copy numbers was estimated by robust piecewise linear regression. Regression slopes significantly higher than 1, i.e. positive allometric change, indicated higher gene expression levels in comparison with CNA levels; the opposite for slopes significantly smaller than 1, i.e. negative allometric change; a slope of 1 indicated a proportional gain/loss in CNA and gene expression, i.e. isometric change. Lastly, the difference between the mean relative mRNA values for tumors containing either gain/loss versus tumors without these aberrations was estimated with t-test for Illumina-BAC probes showing significant association between CNA and expression patterns.

In **paper IV**, 327 Illumina-BAC probe pairs spanning the 8p11-p12 genomic region were selected from smoothed array-CGH data. The statistical analysis was performed in three sequential steps. Firstly, CNA-induced genes were assessed by Pearson correlation (Q<0.05) and the change in relative mRNA levels as a function of DNA copy numbers was estimated by robust piecewise linear regression. Secondly, for Illumina-BAC probes showing significant association between CNA and expression patterns, t-test was performed to estimate the difference between the mean relative mRNA values for tumors containing either gene amplification versus tumors without this aberration. Lastly, probe pairs with Pearson correlation $r > 0.7$ were selected for further analysis to assess the number of tumors displaying gene up-regulation (mRNA $\log_2 > 0.58$) when amplified (DNA $\log_2$ratio > 0.5), as well as, the number of tumors displaying up-regulation in the absence of gene amplification (-0.2 < DNA $\log_2$ratio < 0.5). Statistical analyses were performed in R/Bioconductor.

IMMUNOHISTOCHEMICAL ANALYSIS

In **papers II-IV**, immunostaining was scored as previously described using a semi-quantitative H-score method to calculate the sum of the percentage and intensity of positively stained invasive tumor cells (negative staining = 0; weak staining = 1+; moderate staining = 2+; strong staining = 3+). The H-score ranged from 0 to 300, where H-score = (1 x %1+) + (2 x %2+) + (3 x %3+) (97). FFPE specimens lacking an invasive tissue component were removed from the analysis. Each tumor specimen was scored once, where multiple FFPE sections representing the same tumor were averaged. The X-

tile software (version 3.6.1) was used to determine an H-score cut-off for positive staining by dichotomizing patients according to H-score value (98).

Results and discussion

PAPER I: Comprehensive molecular classification of diploid breast carcinoma

It is well established that structural genetic instability (genetic aberrations) precedes numerical chromosomal instability (chromosome count) during breast tumorigenesis, implying that similar molecular mechanisms take place during tumor progression regardless of ploidy status (99-105). Furthermore, common genetic aberrations are found in diploid and aneuploid breast tumors, despite an approximate 3-fold increase in genetic events following aneuploidization (101). Therefore, to identify recurrent CNAs and abnormal gene expression patterns characteristic of breast carcinoma that may contribute to tumor malignancy and clinical outcome without the added complication of aneuploidization, we conducted genome-wide screening of 97 DBC cases (DNA index = 1, as characterized by flow cytometry) using array-CGH and transcriptional microarrays, respectively. Other forms of structural genetic aberrations such inversions and translocations are not detected using array-CGH and were therefore not reported in the current study.

In the current dataset, genetic alterations of varying size were detected on all chromosomes (ranging from 103 kb to 109 Mb constituting small regions on chromosomes or occasionally several of the smaller chromosomes in their entirety). On average, 40 ± 3.9 (± SEM) chromosomal aberrations (range, 2-204) were detected per diploid tumor sample, including 20 ± 2.0 low-level gains (range, 0-117), 4.9 ± 0.6 high-level gains/amplifications (range, 0-31), 15 ± 1.7 heterozygous losses (range, 0-104), and 0.03 ± 0.02 homozygous deletions (range, 0-1). We identified both genetic alterations characteristic of breast cancer (gain on chromosome cytobands 1q, 8q, and 16p, losses on 11q, 16q, and 17p in at least 25% of cases, as well as high-level gain/amplification on 11q in at least 10% of cases) and novel findings including gain on 8q22.3 and high-level gain/amplification on 1q32.1-q32.2 (Figure 8a) (100, 106-111). In addition, high-level amplification was also identified on cytobands 3q, 8p, 8q, 17q, and 20q in 5-9% of cases.

Unsupervised clustering of the dataset according to the identified genetic alterations showed that tumor malignancy and thereby clinical outcome is influenced by the number of CNAs and the specific genomic location of the CNA within a tumor. This analysis stratified patients into two groups

(henceforth termed genomic clusters 1 and 2). Genomic cluster 2, consisting of tumors with the highest number of alterations, had a more unfavorable prognosis and an overrepresentation of genetic alterations on chromosomes 3, 18, and 20. These findings suggest that the degree of genetic instability may contribute to clinical outcome (5-year DSS for cluster 1 (90%) and cluster 2 (70%), $P = 0.036$).

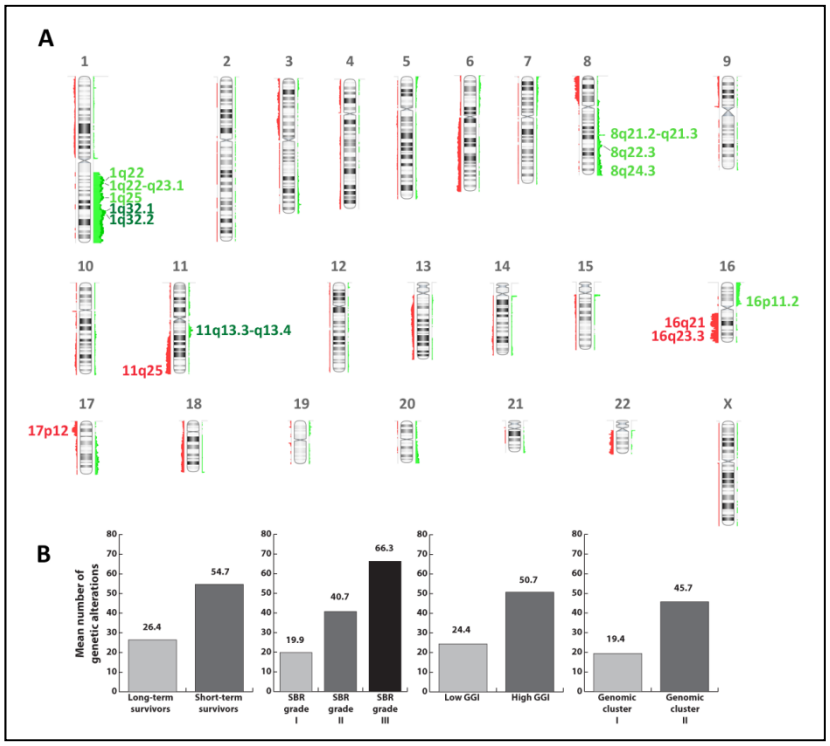

Figure 8. Recurrent copy number alterations (CNAs) in 97 diploid breast carcinomas. A, heterozygous losses (red) are shown to the left and low-level gains (light green) and high-level gains/amplifications (dark green) to the right of the chromosome ideograms. Heterozygous losses (red) and low-level gains (light green) in at least 25% of cases and high-level gains/amplifications (dark green) in at least 10% of cases are shown. B, higher numbers of genetic alterations and thereby high genetic instability in the tumor dataset are associated with molecular and clinicopathological characteristics.

In relation to established clinicopathological features, we observed a 2-fold increase in the average number of genetic alterations in tumors from short-term survivors ($P = 0.001$), tumors with high GGI ($P = 0.0004$), tumors in genomic cluster 2 ($P = 0.001$), and high-grade tumors ($P = 0.007$), in comparison with their corresponding counterparts. In addition, the number of genetic alterations increased with pathologic tumor size, but this change

was borderline significant (P = 0.07). Although genetic alterations generally accumulate over time, these findings suggest that genomic instability may also be strongly linked with tumor differentiation regardless of tumor size and that the number of genetic alterations may have a detrimental effect on clinical outcome (Figure 8b).

Several clinicopathological parameters for breast carcinoma are associated with aggressive tumor behavior and thereby impact clinical outcome, e.g. high histological grade, axillary lymph node status, hormone receptor negativity (ER- and PgR-negativity), HER2/*neu* gene amplification and/or overexpression, and triple negative status (ER-, PgR-, and HER2/*neu*-negativity). The collective transcriptome analysis showed distinct gene expression profiles for tumors displaying aggressive features, affecting key cellular functions including regulation of cell cycle, cell proliferation, cell differentiation, immune response, cell-cell signaling, hormone stimulation, and response to DNA damage stimulus. However, there was no significant difference in the expression patterns between lymph node-positive and node-negative primary tumors; only one gene (*CNTNAP2*) was differentially regulated (at least 1.5-fold difference). These findings suggest that primary tumors may be biologically similar, whether they have the capacity to spread to the axillary lymph nodes or not. To develop a panel of genes significantly deregulated in aggressive tumors, the expression patterns for four tumor groups (short-term DSS with <8-year survival, receptor-negative status, high histological grade, and triple negative status) were compared (Figure 9). Thirteen genes (upregulation of *CBX2*, *S100A8*, and *UBE2C*, and downregulation of *AZGP1*, *DNALI1*, *LOC389033*, *NME5*, *PIP*, *SCUBE2*, *SERPINA11*, *STC2*, *STK32B*, and *SUSD3*) were recurrently differentially regulated in all four groups. The results suggest that DSS with the 8-year cutoff used in this study may be dependent in part on receptor status, tumor grade, and triple negative status as all thirteen malignancy genes found in all four groups were also found in groups B-D. These malignancy genes have been previously implicated in cellular mechanisms including cell growth, motility, and progression in several different cancer forms (112-119). In addition, our results indicate that these genes are ER/PgR antagonists/responsive, with the exception of *UBE2C* which has been shown to crosstalk with HER2/*neu* (120, 121).

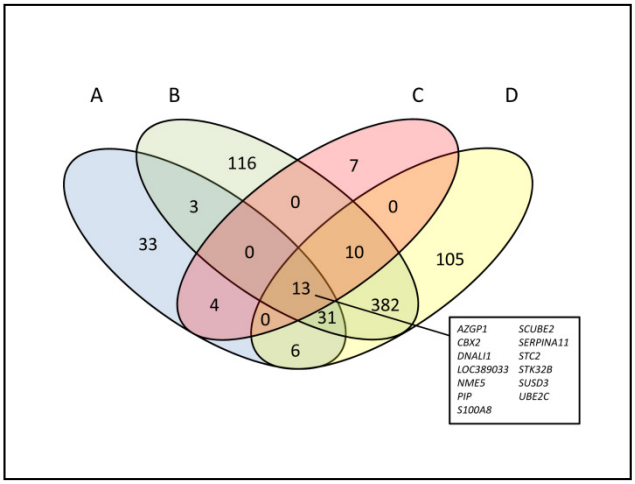

Figure 9. Number of altered genes in different malignancy groups in the 97 diploid breast carcinomas. The gene distribution is shown for <8-year DSS (A), receptor-negative status (B), SBR grade III (C), and triple negative status (D). Thirteen genes were commonly altered in all four groups, of which all thirteen were also found in groups B-D. Unigene clusters are not included in the gene count.

Lastly, the identified genetic alterations and transcriptome profiles were integrated to identify genes with the potential to function as oncogenes (high DNA copy number and elevated gene expression levels) and tumor suppressor genes (loss of one or both DNA copies and low gene expression levels). There was a significant difference in the size of the genetic alterations identified in this study, from a few hundred kilobases to over a hundred megabases and frequently containing a large number of genes within the same DNA segment. To put this into perspective, the largest gene in the human genome, *CNTNAP2*, is 2.3 Mb and would require 50 juxtaposed *CNTNAP2* genes to span the largest identified DNA segment in its entirety without any gaps. To narrow down the number of genes with the potential to play an important role in the breast cancer phenotype, we identified common gains and losses in at least 25% of the patient population and high-level gains/amplifications in at least 10% of cases. This analysis showed that few of the transcripts located within recurrent genetic alterations (13%) were also altered on the mRNA level, which is consistent with previous reports (108, 122, 123). These results suggest that many of the genes located within an altered DNA segment may be bystander genes with abnormal DNA copy numbers because they are located in close vicinity to key cancer genes. Important cancer-related genes are frequently selected for DNA amplification or deletion during the tumorigenic process. However, genes adjacent to cancer-related genes may also be included within an amplicon or a breakpoint because breakage-fusion-bridges cycles

mechanically break chromosomes at fragile sites with different frequencies. Recurrent amplicons are therefore often heterogeneous in different tumors, with regard to amplicon length and the genes which it contains. Nevertheless, only key genes should show altered gene expression levels. In the current study, we were able to show three different expression level patterns in relation to observed CNA log_2ratios: (a) 25% of transcripts had lower expression levels than expected (termed a negative allometric change) suggesting that other cellular mechanisms than genetic alterations, such as epigenetics, may have an effect on gene expression levels; (b) the majority of transcript expression levels (69%) were expected given the observed log2ratios (isometric change), which may imply that CNAs were the dominating factor for altered gene expression levels; and (c) 5% of the transcripts (e.g. the HER2/*neu* oncogene) had higher expression levels than expected (positive allometric change) indicating that more than one cellular mechanism besides genetic alterations may control gene expression levels (Figure 10). In addition, the majority of transcripts with altered DNA copy numbers and associated gene expression levels (40/48 transcripts, 83%) were located on chromosome 1q, where five clusters of adjacent genes were altered. Gain and/or high-level gains/amplifications of genes on chromosome 1q is a common event in breast carcinoma (124).

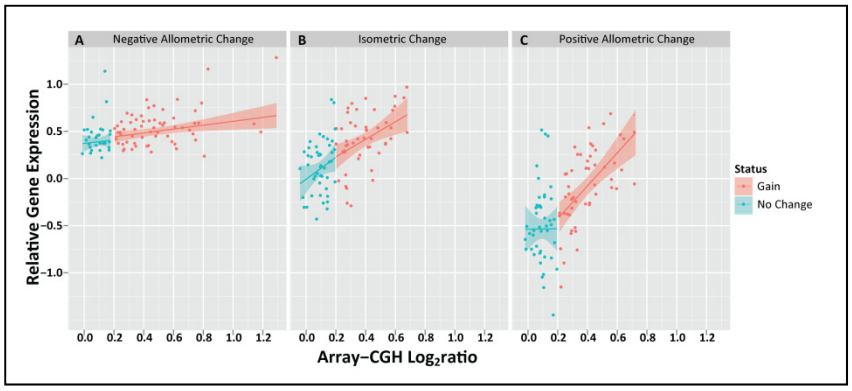

Figure 10. Robust piecewise linear regression analysis showing three different gene expression patterns given the observed CNA log_2ratios determined using array-CGH. (A) Negative allometric change depicts lower transcript expression levels given array-CGH log2ratio. (B) Isometric change depicts expected expression levels in relation to array-CGH log2ratio. (C) Positive allometric change depicts significantly higher transcript expression levels given the observed array-CGH log2ratio. Individual breast tumors are depicted as dots. Shaded areas, 95% confidence intervals of the regression line for gain (array-CGH log2ratio ≥ 0.2) and no change in gene dosage (0 < log2ratio < 0.2).

PAPER II: Validation of the 13-marker signature in breast carcinoma

Since its introduction in the 1990s, microarray technology (e.g. DNA, epigenome, mRNA, miRNA, proteome) has provided a wealth of new knowledge about the biology of cancer. During the past 20 years, a multitude of different gene expression signatures have been reported for breast cancer to predict metastasis, prognosis, and treatment efficacy (62, 64-66, 68-70, 73). In **paper II**, the 13-marker gene expression signature (overexpression of *CBX2*, *S100A8*, and *UBE2C* as well as underexpression of *AZGP1, DNALI1*, *LOC389033*, *NME5*, *PIP*, *SCUBE2*, *SERPINA11*, *STC2*, *STK32B*, *and SUSD3*) identified in **paper I** was validated using six external microarray datasets containing 1,141 invasive breast cancer samples and further evaluated with immunohistochemistry on multiple FFPE samples from 71/97 patients from **paper I**.

It would have been ideal to validate our microarray results from **paper I** (training set) using previously published breast cancer microarray datasets containing clinical data and preferably profiled with the same microarray platform, Illumina HumanHT-12 Beadchips. However, there are not many Illumina datasets publicly available for breast cancer. Therefore, six Affymetrix U133A datasets were used as the test cohort, which contained clinical data and four different clinical endpoints (DFS, OS, DMFS, and RFS). Because of differences in the Illumina and Affymetrix platforms, we were not able to evaluate the full 13-marker signature in the validation cohort; three genes (*LOC389033, SERPINA11,* and *SUSD3*) were not found on the Affymetrix platform and were excluded from the analysis.

Nevertheless, similar transcriptional patterns for the evaluated genes were observed in both the training and test sets. In addition, univariate Cox regression analysis showed that the full 13-marker signature was associated with DSS in the training cohort, whereas 6/10 evaluated genes (*AZGP1, NME5, S100A8, SCUBE2, STC2,* and *UBE2C*) were associated with DFS in the test cohort. These findings suggest that although transcriptional levels for specific genes may be consistent in different datasets, the use of a biomarker to determine prognosis may also be affected by the endpoint used. The DFS endpoint used in the test cohort is a combination of DSS, i.e. the time from initial diagnosis to breast cancer-related death, and the time from initial diagnosis to first loco-regional or distant recurrence.

Hierarchical clustering of the training and test cohorts (performed separately) using the six common markers (*AZGP1, NME5, S100A8, SCUBE2, STC2,* and *UBE2C*) stratified both datasets into two main clusters. In general,

the first cluster (termed the high-risk group) showed upregulation of *S100A8* and *UBE2C* as well as downregulation of *AZGP1, NME5, SCUBE2,* and *STC2*. In general, the six genes were inversely regulated in the second cluster (termed the low-risk group). However, 11 of the 1,141 samples from the test dataset were classified in a third cluster because of disparate transcriptional levels. In addition, the high-risk group was associated with aggressive breast cancer features in both cohorts, e.g. high histological grade, steroid hormone negativity (ER and PgR), HER2/*neu*-positivity, triple negative status, and the HER2/ER- and basal-like intrinsic subtypes ($P < 0.05$). Furthermore, the high-risk group was significantly associated with high S-phase fraction in the training cohort. Kaplan-Meier survival analysis showed that the six-marker signature was a predictor of DSS for the training cohort, as well as DFS, OS, and DMFS for the test cohort. No significant relationship was found between RFS and the six-marker signature (Figure 11). Taken together, these results suggest that DSS, which does not take into consideration the time to first recurrence as do DFS and RFS but rather only breast cancer-related death, may be a one of the strongest endpoints to determine the prognostic potential of the six-marker signature.

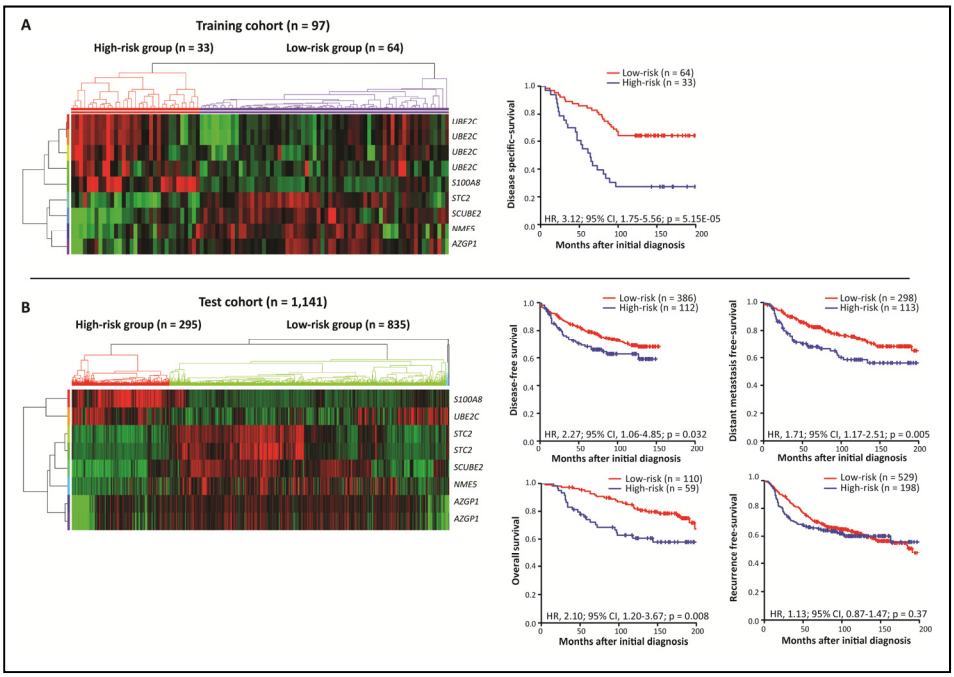

Figure 11. Prognostic potential of the six-marker signature in the training and test cohorts. (A) Hierarchical clustering and risk assessment of the training cohort (n = 97) according to transcriptional levels stratified patients into statistically significant high- and low-risk groups.

In the left panel, up-regulation is shown in red, down-regulation in green, and no change in transcriptional levels in black. Estimates of the probability of disease-specific survival according to the risk groups is shown (right panel), where the x-axes depict Months after initial diagnosis and the y-axes depict disease-specific survival. P-values, hazard ratios (HR), and 95% confidence intervals (95% CI) were calculated using the log-rank test and Cox proportional hazards regression, respectively. (B) Hierarchical clustering and risk assessment of the test cohort (n = 1,141). Estimates of the probability of disease-free, distant metastasis-free, overall, and recurrence-free survival are displayed, where the x-axes depict Months after initial diagnosis and the y-axes depict survival rates.

Despite the fact that transcriptional and protein expression levels do not always correlate, few genetic signatures are validated at the protein level. This is astonishing because proteins are macromolecules that perform crucial processes within cells and therefore the best determinants of biological activity within a cell or tissue. In contrast, specific mRNA transcripts may not be translated into functional protein units because of mRNA stability and degradation. Here, protein expression levels for the full 13-marker signature were also evaluated using immunohistochemistry with multiple FFPE samples from patients in the training set (71/97 patient samples were available at the Sahlgrenska University Hospital, Department of Pathology) and significant markers were further validated in representative samples using Western blot. Two genes were excluded from this analysis (LOC389033 does not produce a protein product and the antibody for STK32B could not be optimized). We found that AZGP1, S100A8, NME5 were consistently expressed at the mRNA and protein expression levels (Figure 12). The differences in mRNA and protein levels for the remaining markers may be the result of analyzing mRNA levels in the whole tumor mass (containing both malignant and non-malignant cells) and protein levels solely in invasive tumor cells, the choice of antibody for immunohistochemistry, and differences in expression levels at these two biological levels which has been shown previously for the PIP protein, where transcription is induced by androgens and is induced post-transcriptionally by prolactin (125).

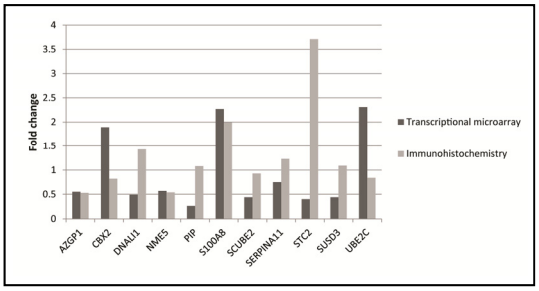

Figure 12. Correlation of mRNA and protein expression levels for the markers. The markers are shown on the X-axis and the fold change on the Y-axis.

The protein expression levels for the 11 analyzed markers were evaluated in relation to clinical outcome using DSS as endpoint. Kaplan-Meier analysis showed that aberrant AZGP1, PIP, S100A8, and UBE2C protein expression levels were predictive of clinical outcome. Significantly shorter survival rates were shown for patients with tumors expressing low levels of AZGP1 and PIP, as well as enhanced levels of S100A8 and UBE2C in invasive neoplastic cells. AZGP1 and PIP showed cytoplasmic and membranous staining, whereas S100A8 and UBE2C showed cytoplasmic and nuclear staining (Figure 13a). To stratify the 71 patients into high- and low-risk groups, a risk score was calculated for each patient using the combined protein expression patterns for the four-marker panel (AZGP1, PIP, S100A8, and UBE2C). Patients with missing values for one or more of the four proteins were excluded from the analysis, leaving 59 patients. The risk groups were significantly associated with DSS, where approximately 50% of patients were characterized as high-risk patients and had shorter survival rates (Figure 13b). Lastly, a concordance index (C-index) for the time-dependent area under the ROC curve [AUC(t)] was calculated for individual markers, the four-marker panel, and established clinicopathological features to establish the predictive performance of different models over time. This analysis showed that the four-marker panel was relatively stable over time and performed better than either of the markers individually, increasing the C-index from 0.586-0.628 to 0.735. In addition, outcome prediction was further improved by combining the four-marker panel together with established clinicopathological parameters (patient age at diagnosis, histological grade, number of positive axillary lymph nodes, pathological tumor size, ER/PgR status, and HER2/*neu* status; Figure 13c). Taken together, these findings suggest that outcome prediction of breast cancer can be improved by combining molecular markers for tumor behavior and clinicopathological features.

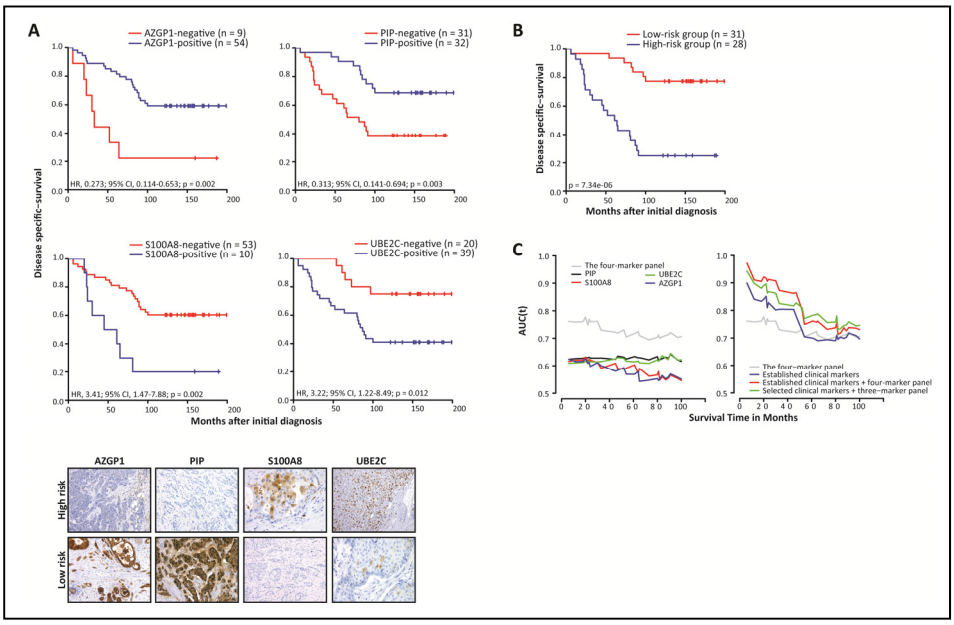

Figure 13. Prognostic potential of aberrant protein levels of the four-marker panel in breast carcinoma. (A) Estimates of the probability of disease-specific survival according to dichotomized protein expression levels of AZGP1, PIP, S100A8 and UBE2C (n =71). The x-axes depict Months after initial diagnosis and the y-axes depict disease-specific survival. P-values, hazard ratios (HR), and 95% confidence intervals (95% CI) were calculated using the log-rank test and Cox proportional hazards regression, respectively. Representative immunohistochemical staining in the invasive component of tumors according to the risk groups is shown in the bottom panel. (B) Estimates of the probability of disease-specific survival according to dichotomized protein expression levels of AZGP1, PIP, S100A8 and UBE2C (n = 59). (C) The lines represent the time-dependent area under the ROC curve [AUC(t)] for AZGP1, PIP, S100A8 and UBE2C protein expression assessed individually and combined as a four-marker panel. Outcome prediction was improved by combining the four markers together as a single model (C-index from 0.586-0.628 to 0.735) and by combining established clinicopathological features (patient age at diagnosis, histological grade, number of positive axillary lymph nodes, pathological tumor size, ER/PgR status and HER2/neu status) with the four-marker panel (C-index from 0.773 to 0.836). The x-axes depict survival time in months and the y-axes depict the time-dependent area under the ROC curve [AUC(t)].

Gene expression (transcriptional and protein expression) is frequently regulated by more than one mechanism. DNA amplification of the 1q (*S100A8*) and 20q13.13 (*UBE2C*) genomic regions are frequent genetic events in breast carcinoma, as well as anaplastic thyroid carcinoma cell lines and primary colon tumors and liver metastases, respectively (126-129). In **paper I**, we showed that the aberrant transcriptional patterns for the genes in the 13-marker signature were not due to abnormal DNA copy number. Here, we found that only 3/10 S100A8-positive tumors displayed DNA

amplification or low-level gain, while DNA copy numbers for *AZGP1*, *PIP*, and *UBE2C* were normal. In pancreatic cancer, epigenetic regulation of AZGP1 resulting in lower levels of the protein were associated with aggressive tumor features and the induction of epithelial-mesenchymal transition (130). Here, we found that AZGP1, PIP, and UBE2C were frequently co-expressed, whereas S100A8 was expressed independent of the other proteins. Interestingly, common functions have been previously found for AZGP1, PIP, S100A8, and UBE2C: AZGP1 and PIP are both secreted glycoproteins frequently co-expressed in exocrine glandular epithelia of various normal tissues, adipose tissue, and cancer (131-140); AZGP1 and UBE2C are both involved in cell cycle regulation and thus have an effect on cell proliferation and inhibition of cdc2 activity (141, 142); elevated levels of AZGP1, PIP and S100A8 are found in human fluids composed of proteins involved in inflammatory and immune responses (143); the FDA approved proteasome inhibitor bortezomib targets both UBE2C and PIP expression (144, 145).

PAPER III: The prognostic potential of breast-cancer related genes in oral squamous cell carcinoma, two biologically similar cancer types

The molecular characterization of different cancer types into intrinsic subtypes has become a popular approach to describe the heterogeneous nature of cancer. Recently, similarities between several different breast cancer subtypes and other cancer subtypes have been shown, e.g. ovarian cancer, glioma, hepatocellular carcinoma, lung squamous cell carcinoma, nasopharyngeal carcinoma, and head and neck squamous cell carcinoma (146, 147). Therefore, cancers derived from different anatomical sites may display similar biological characteristics because common signaling pathways may be perturbed (7). To test this concept, we evaluated the prognostic potential of 16 breast cancer-related genes (*AZGP1, BTG2, CBX2, CNTNAP2, DNALI1, LOC389033, NME5, PIP, S100A8, SCUBE2, SERPINA11, STC2, STK32B, SUSD3, UBE2C,* and *WHSC1L1*) identified by us in oral squamous cell carcinoma, a cancer form with relatively low 5-year survival rates (148-150).

We evaluated the 16-marker signature in oral squamous cell carcinoma (OSCC) at the protein level using FFPE samples from 43 OSCC patients. PIP and DNALI1 were excluded from the immunohistochemical analysis because of overall low protein expression levels in OSCC samples, and *LOC389033* was again excluded because it does not produce a protein product. In

addition, AZGP1 cytoplasmic and nuclear staining were evaluated separately. The majority of the markers in the signature had similar transcriptional patterns for both breast carcinoma and OSCC, with the exception of the S100A8, STC2, STK32B, SUSD3, and WHSC1L1 proteins which were inversely regulated in OSCC samples.

A correlation analysis between protein expression patterns and established clinicopathological features showed an association between S100A8-negativity and poor tumor differentiation (high histological grade), SCUBE2-negativity and cervical lymph node metastasis (only 1/12 SCUBE2-positive tumors was lymph node-positive), CBX2-positivity and strong tumor inflammatory infiltration, and STK32B-negativity and larger tumor size (T3-T4). In addition, borderline significance was found between UBE2C and SCUBE2 expression and tumor differentiation, SERPINA11 expression and tumor inflammatory infiltration, as well as NME5 and p16 expression. Kaplan-Meier analysis showed that CNTNAP2-positivity was associated with significantly shorter DSS rates, whereas S100A8-negativity was associated with OS (Figure 14a). Predictive models for DSS and OS were then calculated using Cox proportional hazards models. In the univariate analysis, CNTNAP2 protein expression, cervical lymph node status, and tumor size were statistically significant for DSS. However, a model containing all three variables (C-index = 0.949) was not significantly better than a model containing only lymph node status and tumor size (C-index 0.941). These findings further illustrate the profound effect lymph node metastases have on clinical outcome for OSCC patients, i.e. early locoregional recurrences frequently occur within two years of initial treatment and have an adverse effect on survival rates (151). To further evaluate the role aberrant CNTNAP2 protein expression patterns have on OSCC clinical outcome, it may be necessary to limit the analysis to a larger cohort of lymph node-negative patients. Univariate analysis for OS showed an association with patient age at diagnosis, tumor size, cervical lymph node status, and S100A8 protein expression. Outcome prediction was significantly improved by including S100A8 protein expression, increasing the C-index from 0.697 to 0.849. In addition, the model was stable even after eight years after initial diagnosis (Figure 14b).

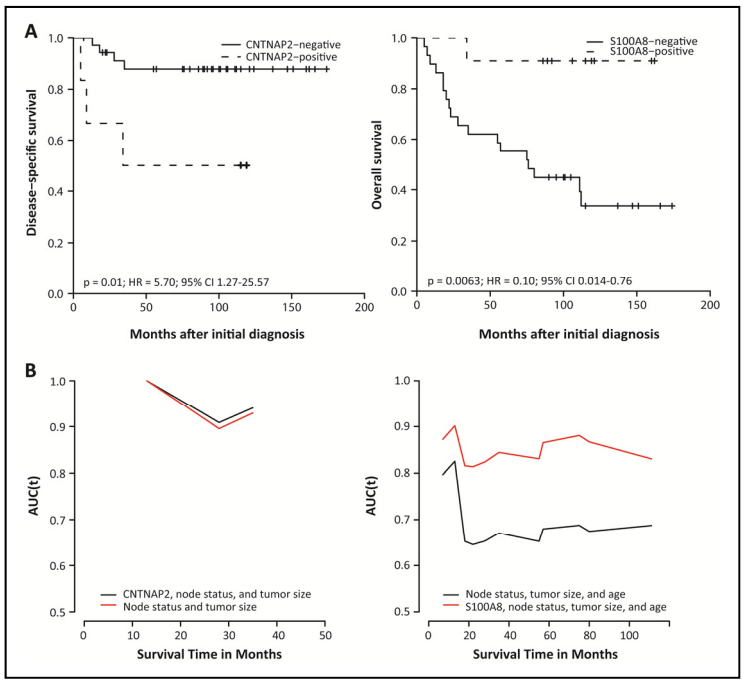

Figure 14. Clinical significance of aberrant CNTNAP2 and S100A8 protein expression in OSCC. (A) Estimates of the probability of disease-specific survival and overall survival according to dichotomized protein expression levels of CNTNAP2 and S100A8, respectively (n = 43). The x-axes depict Months after initial diagnosis and the y-axes depict survival rates. P-values, hazard ratios (HR), and 95% confidence intervals (95% CI) were calculated using the log-rank test and Cox proportional hazards regression, respectively. (B) The lines represent the time-dependent area under the ROC curve [AUC(t)] for CNTNAP2 protein expression adjusted for node status and tumor size, as well as S100A8 protein expression adjusted for node status, tumor size, and age. Outcome prediction was improved by combining established clinicopathological features with S100A8 expression (C-index from 0.697 to 0.849), but not by adding CNTNAP2 expression to a model for DSS (C-index from 0.941 to 0.949). The x-axes depict survival time in months and the y-axes depict the time-dependent area under the ROC curve [AUC(t)].

PAPER IV: Isolation of a putative driver gene for 8p11-p12 DNA amplification in breast carcinoma

Specific regions in the human genome are hotspots for amplicon formation because they contain one or more genes with oncogenic potential (16). The 8p11-p12 amplicon is a common genetic event in solid tumors, e.g. breast carcinoma (152, 153), pleuropulmonary blastoma (154), invasive micropapillary carcinoma (155), lung cancer and esophageal squamous cell

carcinomas (156-159), urothelial carcinoma (160), urinary bladder cancer (161), osteosarcoma (124), and pancreatic adenocarcinoma (158). In breast carcinoma (both familial and sporadic cases), chromosome 8 is a frequent target for DNA amplification and loss, occurring in 10-25% of cases (162-166). However, this genomic region spans over 10 Megabase (Mb) and encompasses approximately 53 known genes, many of which have been shown to have oncogenic potential. In addition, analysis of breast cancer cell lines harboring 8p11-p12 amplification have shown that the initiation site and structure of the 8p11-p12 DNA rearrangement involves different mechanisms of gene activation, i.e. translocations and gene amplifications, thereby result in the activation of different combinations of candidate genes (167). Therefore, the aggressive phenotype (significantly shorter survival rates, high histological grade, high Ki-67, and elevated cyclin E levels) imposed by the 8p11-p12 amplicon may be the result of one or more interacting genes in this genomic region. To date, eight genes have emerged as promising candidates, e.g. *BAG4*, *C8orf4*, *DDHD2*, *ERLIN2, LSM1*, *PPAPDC1B*, *WHSC1L1*, and *ZNF703* (166, 168-173).

To investigate the role recurrent DNA amplification of the 8p11-p12 genomic region has on breast tumorigenesis, we compiled 229 array-CGH datasets from three separate studies, including the 97 diploid breast carcinoma samples described in **paper I** (149, 174, 175). Recurrent CNAs spanning the 8p11-p12 region were observed in 83/229 breast tumors (36%), including 47/83 high-level gains/amplifications, 20/83 low-level gains, and 16/83 heterozygous losses. In general, two different patterns of regional amplification were observed, i.e. cases with focal amplification spanning the 8p11-p12 region and cases harboring both amplification of the 8p11-p12 region in conjunction with amplification of the 8q arm (Figure 15). In addition to co-amplification with the 8q region, the 8p11-p12 amplicon was also co-amplified with 1q, 11q, 12p, 16p, 17q, or 20q; the 8p11-p12 amplicon occurred as the sole region of amplification in two cases. Approximately 53% of 8p11-p12 amplified tumors also harbored DNA amplification of the *MYC* gene (8q24), in comparison with 20% and 18% of co-amplification with *CCND1* (11q13) and *ERBB2* (17q12), respectively. Interestingly, we observed a five-fold increase in the number of amplifications in lesions containing the 8p11-p12 amplicon compared to tumors lacking the amplicon (P = 1.8E-13).

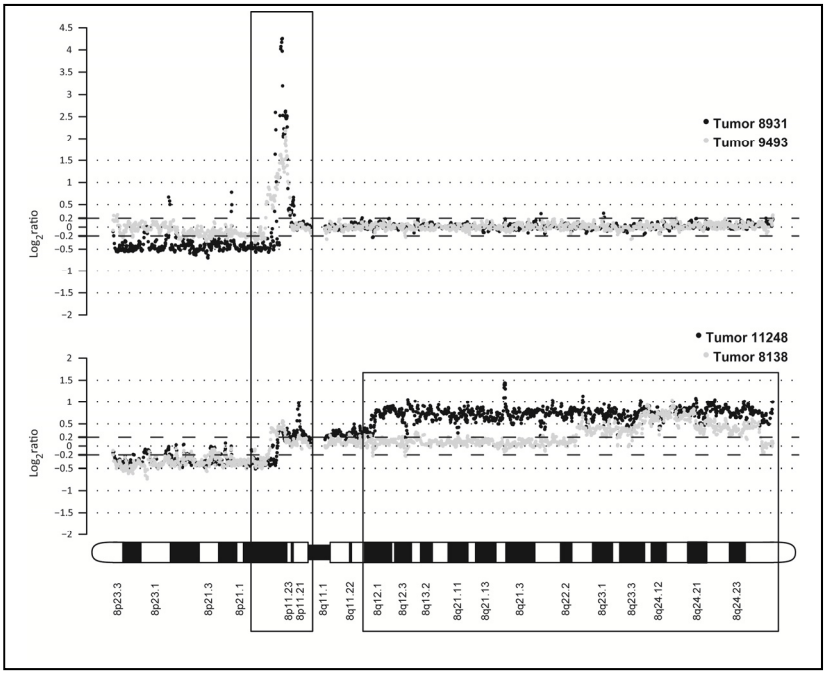

Figure 15. Array-CGH genomic profiles showing recurrent DNA amplification of the 8p11-p12 genomic region in breast carcinoma. The top panel shows focal amplification of the 8p11-p12 region in two breast tumors. Black dots depict BAC clones spanning chromosome 8 for tumor 8931 and gray dots for tumor 9493. The bottom panel shows amplification of the 8p11-p12 and 8q regions. Black dots depict BAC clones spanning chromosome 8 for tumor 11248 and gray dots for tumor 8138. The x-axis shows chromosome 8 from the 8p telomere to the 8q telomere. The y-axis shows the $\log_2$ratio value for each BAC clone (tumor gDNA versus normal control gDNA).

All types of CNAs at the 8p11-p12 region (gain, loss, and amplification; *P* = 5.0E-6), including amplification (*P* = 4.0E-5) or loss (*P* = 0.005), were associated with reduced OS rates. In contrast, low-level gain of the region was not indicative of unfavorable prognosis (*P* = 0.08). In addition, the amplicon was most prevalent in tumors of large pathologic size (*P* = 0.0002), high GGI status (*P* = 0.0004), and high S-phase fraction (*P* = 0.02). There was no significant difference in histologic type, estrogen/progesterone receptor status, HER2/*neu* receptor status, triple negative status, or molecular breast cancer subtype. Nevertheless, OS rates were affected by molecular subtype; basal-like tumors containing the 8p11-p12 amplicon had significantly shorter survival rates, followed by the Luminal B subtype (*P* = 0.0003).

As expected, we observed both intratumoral (differences in individual cells within a tumor) and intertumoral (differences between different tumors) heterogeneity in cases harboring the 8p11-p12 amplicon. Comparison of genomic profiles generated using array-CGH showed that the amplicon could be divided into 5 subregions mapping to a 12.1 Mb region from 31.9-43.9 Mb (from telomere to centromere on the 8p arm), which was further refined to nine minimal common amplification peaks (range, 41.2-377.4 kb) from 34.3-42.5 Mb (Figure 16) using the Nexus Copy Number Professional 4.1 software. One of the smallest peaks and notably the most common mapped to a 67.9 kb region spanning the *WHSC1L1* gene on chromosome band 8p12. The *WHSC1L1* gene was amplified in 32 of the 47 (68%) 8p11-p12 amplified samples, consisting of four Basal-like, three HER2/ER-, and twenty-five Luminal subtype B tumors. Dual-color interphase FISH performed using a contig of overlapping BAC clones covering the 8p11-p12 genomic region showed that specific regions within the amplicon could range from neutral DNA copy numbers (two copies of a gene) to high-level amplification (up to 50 copies of a gene) in different individual cells within the same tumor.

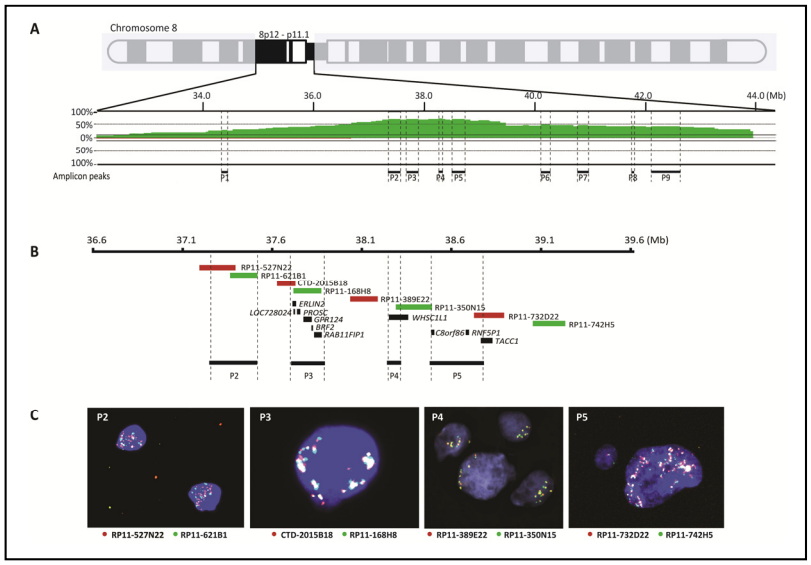

Figure 16. DNA amplification of the 8p11-p12 genomic region in breast carcinoma. (A) Frequency plot of 8p11-p12 amplification in 47 breast tumors. The x-axis depicts the genomic position on chromosome 8 in Mb according to UCSC May 2004 hg17: NCBI Build 35 from the telomere on the 8p arm to the centromere; y-axis, percentage of tumors with amplification (green) and loss (red). Vertical dashed lines indicate the nine most common amplification peaks (P1-P9). (B-C) Zoom-in of peaks P2-P5 showing BAC clones used in locus-specific dual-color FISH and significantly associated genes spanned each peak. Biotin-labeled (green) BAC

clones RP11-621B1 (P2), RP11-168H8 (P3), RP11-350N15 (P4), RP11-742H5 (P5) were combined with digoxigenin-labeled (red) BAC clones RP11-527N22 (P2), CTD-2015B18 (P3), RP11-389E22 (P4), RP11-732D22 (P5). Overlapping DNA sequences were detected as yellow hybridization signals. The interphase nuclei were counterstained with DAPI.

An integrative analysis of DNA copy numbers and transcriptional expression for genes spanning the 8p11-p12 amplicon was performed in two steps using 150 of the 229 profiled with both methods. First, 1933 differentially regulated transcripts (≥1.5 fold-change, $Q < 0.01$) were identified which distinguished tumors harboring the amplicon (n = 45) from tumors with neutral DNA dosage (n = 71), among which 24 transcripts (18 genes and 1 EST: *ZNF703*, *ERLIN2, PROSC, BRF2, RAB11FIP1, EIF4EBP1, BAG4, DDHD2, PPAPDC1B, WHSC1L1, Hs.156542, TM2D2, ADAM9, GOLGA7, MYST3, PLAT, IKBKB, THAP1, RNF170*) were located in the 8p11-p12 genomic region. Second, a correlation analysis ($Q < 0.05$, $r > 0.7$) was performed using robust piecewise linear regression to identify genes significantly associated with DNA and mRNA levels (up-regulation when amplified). Ten genes (*PROSC, BRF2, EIF4EBP1, LSM1, BAG4, DDHD2, PPAPDC1B, WHSC1L1, TM2D2, GOLGA7)* were identified, nine (*PROSC, BRF2, EIF4EBP1, BAG4, DDHD2, PPAPDC1B, WHSC1L1, TM2D2, GOLGA7*) of which were common for both methods. Interestingly, the *WHSC1L1* gene had the strongest association between DNA dosage and relative mRNA levels. The *WHSC1L1* gene, also known as *NSD3*, is a member of the NSD family of histone lysine methyltransferases. There are at least two co-expressed WHSC1L1 isoforms (the long and short isoforms) that may compete for binding sites on target proteins (176). The full-length WHSC1L1 protein contains several functional domains with methyltransferase and protein binding activity, which play a pivotal role in chromatin modification and regulation of transcription. On the other hand, the short WHSC1L1 isoform only contains one PWWP-domain (proline-tryptophan-tryptophan-proline) that may be involved in cell growth. Although both the long and short splicing variants for the *WHSC1L1* gene (*WHSC1L1-L* and *WHSC1L1-S*) were recurrently up-regulated in the presence of DNA amplification, *WHSC1L1-L* was also up-regulated in 45% of cases in the absence of amplification showing that this gene may be activated by more than one molecular mechanism. None of the other eight significantly expressed genes in the 8p11-p12 region showed this tendency, but Hs.105962 (96%), Hs.170296 (96%), and *ZNF703 (98%)* were generally up-regulated independent of DNA copy number status in tumors compared with normal breast tissue. However, Cox regression analysis showed that

five of the nine candidates (*BAG4*, *DDHD2*, *EIF4EBP1*, *GOLGA7,* and *WHSC1L1-L*) were predictive of overall survival.

WHSC1L1-expressing tumors also displayed distinct transcriptional patterns, with 64 differentially expressed transcripts (62 genes) in comparison with WHSC1L1-negative tumors, including four genes spanning the 8p11-p12 region (*ASH2L*, *DDHD2*, *RAB11FIP1, ZNF703*). The differentially expressed genes were primarily associated with crucial cellular processes, e.g. regulation of transcription, cell differentiation, cell adhesion, cell migration, immune response, negative regulation of apoptosis, metabolism, and methyltransferase activity. Up-regulation of WHSC1L1 was also associated with shorter OS (P = 1.35E-8). In addition, the *ASH2L* (r = 0.73), *DDHD2* (r = 0.75), *RAB11FIP1* (r = 0.58), and *WHSC1L1* genes clustered together and had the highest correlation coefficients, whereas the *ZNF703* gene clustered together with genes regulating cell adhesion and migration. Furthermore, the *WHSC1L1-L* and *ASH2L* genes not only have structural similarities (both contain SET and zinc finger domains), but are also included in histone methyltransferase complexes which activate transcription by preferential methylation of lysine-4 of histone H3 which makes these two genes interesting candidates. *WHSC1L1-L* also methylates lysine-27 of histone H3, which is an epigenetic tag denoting inhibition of transcription. *WHSC1L1-S* lacks the SET domain needed for methyltransferase activity. Additionally, the *WHSC1L1-L, ASH2L,* and *ZNF703* genes all contain a zinc finger domain, whereas DDHD2 and RAB11FIP1 are both phosoproteins.

Lastly, we performed a correlation analysis between CNAs, aberrant methylation patterns, and transcriptional patterns in 22/229 tumors (11 tumors harboring the 8p11-p12 amplicon and 11 tumors with neutral DNA dosage in the 8p11-p12 region). This analysis showed that the majority of cytosine sites were hypermethylated (89%) and only 11% were hypomethylated. As expected, hypermethylation occurred most frequently in the promoter regions of genes. In addition, hypermethylation ranged from 72-98% on all chromosomes, whereas hypomethylation was most prevalent on chromosomes 8 (28%) and 9 (24%). Interestingly, only 61 aberrantly methylated coding RNAs (4.5%) were differentially regulated, which play a role in cell differentiation, DNA replication, cell migration, and cell adhesion. Eleven of the 61 genes spanned the 8q12.1-q24.22 frequently co-amplified in 8p11-p12 amplified tumors, nine of which were hypomethylated and overexpressed (*IMPAD1*, *NDRG1, PLEKHF2*, *RRM2B*, *SQLE*, *TAF2*, *TATDN1*, *TRPS1*, *VPS13B*), the *ENPP2* gene was

hypermethylated and underexpressed, and the *FABP5* gene was hypermethylated but overexpressed. Only one gene (*BRF2*) in the 8p11-p12 genomic region showed aberrant methylation patterns, but this gene was overexpressed despite hypermethylation due to amplification in 7/11 cases. We found that hypomethylation alone frequently enhanced gene expression patterns. However, hypomethylation and DNA amplification of the same transcript further enhanced expression levels. These findings indicate that aberrant methylation patterns may be a secondary event to further lock genes in their inactive or active states only after they have already been silenced or activated by other means (177-180). The *ENPP2* gene was an exception to this phenomenon because hypermethylation occurred at four different cytosine sites in the promoter region of the gene, resulting in lower transcriptional levels despite amplification of the gene in 2/11 samples harboring the 8p11-p12 amplicon (Figure 17). Furthermore, 8/11 genes (*FABP5, NDRG1, PLEKHF2*, *RRM2B*, *SQLE*, *TAF2*, *TATDN1, TRPS1)* may not be distinctive of 8p11-p12 amplification as they were also differentially regulated in *MYC* co-amplified tumors.

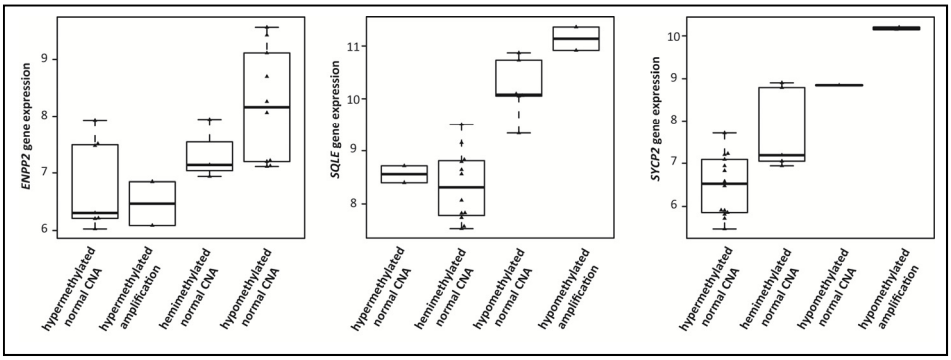

Figure 17. The effect of aberrant DNA copy number and DNA methylation patterns on transcriptional levels. Box plots showing the relationship between CNA status, methylation status, and transcriptional levels for three candidate genes (*ENPP2*, *SQLE*, and *SYCP2*) in 22 tumor samples. X-axis, methylation and CNA status; Y-axis, gene expression signal intensity.

Concluding remarks

In this doctoral thesis, translational research was performed by combining tumor biology with clinicopathological features to improve breast cancer risk assessment, but also for oral squamous cell carcinoma.

Genetic alterations characteristic of breast cancer include gain on chromosome cytobands 1q, 8q, and 16p, losses on 11q, 16q, and 17p, and high-level gain/amplification on 3q, 8p, 8q, 11q, 17q, and 20q. However, few genes located in these regions were altered at the mRNA level, suggesting that only specific genes may be targeted during the oncogenic process. Additionally, a 13-gene signature with normal DNA dosage levels (upregulation of *CBX2*, *S100A8*, and *UBE2C*, and downregulation of *AZGP1*, *DNALI1*, *LOC389033*, *NME5*, *PIP*, *SCUBE2*, *SERPINA11*, *STC2*, *STK32B*, and *SUSD3*) and associated with aggressive tumor behavior was identified, suggesting that genes are activated by several cellular mechanisms during tumorigenesis.

Because transcriptional and protein levels do not always correlate, molecular targets should also be validated at the protein level. Furthermore, outcome prediction for breast carcinoma is improved by combining established clinicopathological features (patient age at diagnosis, histological grade, number of positive axillary lymph nodes, pathological tumor size, ER/PgR status, and HER2/*neu* status) with molecular targets characteristic of aggressive tumor behavior (i.e. AZGP1, PIP, S100A8, and UBE2C).

Despite differences in anatomical site of origin, some clinically relevant biomarkers may be useful therapeutic targets (CBX2, CNTNAP2, S100A8, SCUBE2, and STK32B) for both breast carcinoma and oral squamous cell carcinoma because of biological similarities between the two cancer types.

The experiments on the 8p11-p12 amplicon in breast carcinoma are still ongoing. The 8p11-p12 amplicon contains several genes with oncogenic potential, which may interact to produce the aggressive phenotype. However, the *WHSC1L1* gene is a strong candidate which is frequently co-amplified with the *MYC* gene. Transcriptional levels are not only regulated by CNAs but also aberrant methylation patterns.

Future directions

The primary focus of this doctoral thesis was to identify potential therapeutic targets for breast carcinoma by combining tumor biology with clinicopathological features. In future work, several aspects of the current work can be explored further.

First and foremost, it will be necessary to validate clinically relevant targets using immunohistochemistry in large independent patient cohorts.

Second, genes in the four-marker panel (AZGP1, PIP, S100A8, and UBE2C) have been previously studied as individual targets, but not together. Therefore, it would be interesting to test the effect collective gene knockdown or overexpression of these genes has on breast cancer cell lines *in vitro* and *in vivo*, as well as whether gene up-regulation can transform normal human mammary epithelial cells. In addition, the FDA approved bortezomib, a proteasome inhibitor to disrupt cell cycle function, has been effective in treating multiple myeloma and mantle cell lymphoma and may also target breast tumors overexpressing the UBE2C and PIP proteins.

Third, we found a strong association between the basal-like phenotype and AZGP1 and S100A8 expression. In addition, high-risk patients identified using the four-marker signature frequently had tumors lacking ER/PgR expression, which may indicate the potential presence of a breast CSC fraction. It would therefore be interesting to know whether patients classified in high-risk groups (according to the four-marker panel) have a breast CSC population which may affect conventional breast cancer therapies (chemotherapy, radiotherapy, endocrine therapy) following surgery. This would be tested using immunohistochemical markers for breast CSCs (CD44/CD24, ALDH1A1, ALDH1A3, and ITGA6) rather than cell sorting which requires living cells.

Lastly, recent studies have reported similarities between breast carcinoma and both hormonally-driven (ovarian cancer) and non-hormonally-driven malignant tumors (glioma, hepatocellular carcinoma, lung SCC, nasopharyngeal carcinoma, and head and neck SCC). It would therefore be interesting to assess the prognostic potential of the full 13-marker signature in other cancer forms using immunohistochemistry.

Acknowledgements

This work was conducted at the Oncology Lab and Sahlgrenska Cancer Center, Department of Oncology, Sahlgrenska University Hospital, Sahlgrenska Academy at University of Gothenburg.

Financial support was provided by King Gustav V Jubilee Clinic Cancer Research Foundation, Swedish Cancer Society, LUA, Percy Falk Foundation, Johanna Breast Cancer Research Foundation, Inger and Birgit Hultén's Memorial Foundation, Lions Cancer Research Foundation, Wilhelm and Martina Lundgren Research Foundation, Serena Ehrenström Foundation for Cancer Research / Torsten and Sara Jansson Research Foundation, Assar Gabrielsson Research Foundation for Clinical Cancer Research, Lars Hierta's Memorial Research Foundation, and Sahlgrenska University Hospital Research Foundation.

This work would not have been possible without invaluable information from the Western Sweden Cancer Registry and the Cause of Death Register.

I would like to take this opportunity to express my deepest gratitude to all the people who have helped me through the years, and especially

Khalil Helou, my supervisor, for being both a very good friend and an excellent coach. Your perfectly timed "aha" moments have always kept me on my toes and encouraged me to keep my eyes open.

Per Karlsson and *Anikó Kovács*, my co-supervisors, for always believing in me and for sharing your limitless knowledge. Thank you!

What would I have done without the support of my fellow group members, *Szilárd Nemes, May Semaan*, *Luaay Aziz*, and *Shahin Hajizadeh*!?!

Fellow members of the "Radiobiology group", *Eva Forssell-Aronsson*, *Emil Schüler*, *Nils Rudqvist*, *Britta Langen*, *Johan Spetz*, *Lilian Karlsson*, and *Ann Wikström* for your help and encouragement, but especially for broadening my horizons.

Past and present staff at the Oncology Lab, Departments of Oncology and Pathology, and the Sahlgrenska Cancer Center for providing an inspiring work environment. Special thanks to *Anna Danielsson, Ulla Delle, Ghita Fallenius, Marcela Davila Lopez, and Ulric Pedersen* for invaluable technical support and encouraging words. *Helena Kahu,* for making me laugh so hard that I almost "literally" roll on the floor crying. You're the best storyteller of all time!

Thank you, *Srinivas "Sunny" Veerla* and *Ingrid Wilson* at SCIBLU Genomics DNA Microarray Resource Center (SCIBLU) for all of your help.

My family and friends, especially *my parents*, *sisters*, *Jörgen*, *mamma Margareta, pappa Sylve, pappa Arne*, and *Daniela* for always encouraging me to challenge myself and trusting my judgment. I love you all very much!

References

1. Olson JS. Bathsheba's breast: women, cancer, and history: The John Hopkins University Press; 2002.
2. Cairns J. Mutation selection and the natural history of cancer. Nature. 1975;255:197-200.
3. Claus EB, Schildkraut JM, Thompson WD, Risch NJ. The genetic attributable risk of breast and ovarian cancer. Cancer. 1996;77:2318-24.
4. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA: a cancer journal for clinicians. 2011;61:69-90.
5. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100:57-70.
6. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646-74.
7. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, et al. A census of human cancer genes. Nat Rev Cancer. 2004;4:177-83.
8. Santarius T, Shipley J, Brewer D, Stratton MR, Cooper CS. A census of amplified and overexpressed human cancer genes. Nat Rev Cancer. 2010;10:59-64.
9. Kuwahara Y, Tanabe C, Ikeuchi T, Aoyagi K, Nishigaki M, Sakamoto H, et al. Alternative mechanisms of gene amplification in human cancers. Genes Chromosomes Cancer. 2004;41:125-32.
10. Shimizu N, Shingaki K, Kaneko-Sasaguri Y, Hashizume T, Kanda T. When, where and how the bridge breaks: anaphase bridge breakage plays a crucial role in gene amplification and HSR generation. Experimental cell research. 2005;302:233-43.
11. Ciullo M, Debily MA, Rozier L, Autiero M, Billault A, Mayau V, et al. Initiation of the breakage-fusion-bridge mechanism through common fragile site activation in human breast cancer cells: the model of PIP gene duplication from a break at FRA7I. Human molecular genetics. 2002;11:2887-94.
12. Gisselsson D, Pettersson L, Hoglund M, Heidenblad M, Gorunova L, Wiegant J, et al. Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity. Proceedings of the National Academy of Sciences of the United States of America. 2000;97:5357-62.
13. Glover TW, Arlt MF, Casper AM, Durkin SG. Mechanisms of common fragile site instability. Human molecular genetics. 2005;14 Spec No. 2:R197-205.
14. Hellman A, Zlotorynski E, Scherer SW, Cheung J, Vincent JB, Smith DI, et al. A role for common fragile site induction in amplification of human oncogenes. Cancer Cell. 2002;1:89-97.
15. Terry PD, Rohan TE. Cigarette smoking and the risk of breast cancer in women: a review of the literature. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2002;11:953-71.
16. Albertson DG. Gene amplification in cancer. Trends Genet. 2006;22:447-55.
17. Jones PA, Takai D. The role of DNA methylation in mammalian epigenetics. Science (New York, NY). 2001;293:1068-70.
18. Baylin SB. The cancer epigenome: its origins, contributions to tumorigenesis, and translational implications. Proceedings of the American Thoracic Society. 2012;9:64-5.
19. Esteller M. Aberrant DNA methylation as a cancer-inducing mechanism. Annual review of pharmacology and toxicology. 2005;45:629-56.

20. Gal-Yam EN, Saito Y, Egger G, Jones PA. Cancer epigenetics: modifications, screening, and therapy. Annual review of medicine. 2008;59:267-80.
21. Huang Y, Nayak S, Jankowitz R, Davidson NE, Oesterreich S. Epigenetics in breast cancer: what's new? Breast Cancer Res. 2011;13:225.
22. Lund AH, van Lohuizen M. Epigenetics and cancer. Genes & development. 2004;18:2315-35.
23. Watson CJ, Khaled WT. Mammary development in the embryo and adult: a journey of morphogenesis and commitment. Development. 2008;135:995-1003.
24. Hennighausen L, Robinson GW. Signaling pathways in mammary gland development. Dev Cell. 2001;1:467-75.
25. Ramsay DT, Kent JC, Hartmann RA, Hartmann PE. Anatomy of the lactating human breast redefined with ultrasound imaging. J Anat. 2005;206:525-34.
26. Sopel M. The myoepithelial cell: its role in normal mammary glands and breast cancer. Folia Morphol (Warsz). 2010;69:1-14.
27. DeSantis C, Siegel R, Bandi P, Jemal A. Breast cancer statistics, 2011. CA: a cancer journal for clinicians. 2011;61:409-18.
28. Cancerfonden. Cancerfondsrapporten 2012. Stockholm; 2012.
29. Tot T. Breast Cancer: A Lobar Disease: Springer; 2011.
30. Servan-Schreiber D. Anti cancer: A New Way of Life. New York: Penguin Group; 2009.
31. Warburg O. On the origin of cancer cells. Science (New York, NY). 1956;123:309-14.
32. Ren H, Endo H, Hayashi T. The superiority of organically cultivated vegetables to general ones regarding antimutagenic activities. Mutat Res. 2001;496:83-8.
33. Olsson ME, Andersson CS, Oredsson S, Berglund RH, Gustavsson KE. Antioxidant levels and inhibition of cancer cell proliferation in vitro by extracts from organically and conventionally cultivated strawberries. Journal of agricultural and food chemistry. 2006;54:1248-55.
34. Beatson G. On the treatment of inoperable cases of carcinoma of the mamma: suggestions for a new method of treatment with illustrative cases. Lancet. 1896;148:104-407.
35. Becker S, Kaaks R. Exogenous and endogenous hormones, mammographic density and breast cancer risk: can mammographic density be considered an intermediate marker of risk? Recent results in cancer research Fortschritte der Krebsforschung Progres dans les recherches sur le cancer. 2009;181:135-57.
36. Chen WY. Exogenous and endogenous hormones and breast cancer. Best practice & research Clinical endocrinology & metabolism. 2008;22:573-85.
37. Dunn BK, Wickerham DL, Ford LG. Prevention of hormone-related cancers: breast cancer. J Clin Oncol. 2005;23:357-67.
38. Ringborg U, Dalianis, T., Henriksson, R. Onkologi. 2nd edition ed: Liber AB; 2008.
39. Guinebretiere JM, Menet E, Tardivon A, Cherel P, Vanel D. Normal and pathological breast, the histological basis. Eur J Radiol. 2005;54:6-14.
40. Kennecke H, Yerushalmi R, Woods R, Cheang MC, Voduc D, Speers CH, et al. Metastatic behavior of breast cancer subtypes. J Clin Oncol. 2010;28:3271-7.
41. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. Proceedings of the National Academy of Sciences of the United States of America. 2003;100:3983-8.
42. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med. 1997;3:730-7.

43. Ginestier C, Hur MH, Charafe-Jauffret E, Monville F, Dutcher J, Brown M, et al. ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Cell stem cell. 2007;1:555-67.
44. Ali HR, Dawson SJ, Blows FM, Provenzano E, Pharoah PD, Caldas C. Cancer stem cell markers in breast cancer: pathological, clinical and prognostic significance. Breast Cancer Res. 2011;13:R118.
45. Haraguchi N, Ishii H, Mimori K, Tanaka F, Ohkuma M, Kim HM, et al. CD13 is a therapeutic target in human liver cancer stem cells. The Journal of clinical investigation. 2010;120:3326-39.
46. Bao S, Wu Q, McLendon RE, Hao Y, Shi Q, Hjelmeland AB, et al. Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. Nature. 2006;444:756-60.
47. Perou CM, Jeffrey SS, van de Rijn M, Rees CA, Eisen MB, Ross DT, et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proceedings of the National Academy of Sciences of the United States of America. 1999;96:9212-7.
48. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406:747-52.
49. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proceedings of the National Academy of Sciences of the United States of America. 2001;98:10869-74.
50. Staaf J, Ringner M, Vallon-Christersson J, Jonsson G, Bendahl PO, Holm K, et al. Identification of subtypes in human epidermal growth factor receptor 2--positive breast cancer reveals a gene signature prognostic of outcome. J Clin Oncol. 2010;28:1813-20.
51. Bediaga NG, Acha-Sagredo A, Guerra I, Viguri A, Albaina C, Ruiz Diaz I, et al. DNA methylation epigenotypes in breast cancer molecular subtypes. Breast Cancer Res. 2010;12:R77.
52. Bergamaschi A, Kim YH, Wang P, Sorlie T, Hernandez-Boussard T, Lonning PE, et al. Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. Genes Chromosomes Cancer. 2006;45:1033-40.
53. Melchor L, Benitez J. An integrative hypothesis about the origin and development of sporadic and familial breast cancer subtypes. Carcinogenesis. 2008;29:1475-82.
54. Farmer P, Bonnefoi H, Becette V, Tubiana-Hulin M, Fumoleau P, Larsimont D, et al. Identification of molecular apocrine breast tumours by microarray analysis. Oncogene. 2005;24:4660-71.
55. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biol. 2007;8:R76.
56. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, et al. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics. 2006;7:96.
57. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. 2010;12:R68.
58. Perou CM. Molecular stratification of triple-negative breast cancers. The oncologist. 2010;15 Suppl 5:39-48.

59. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490:61-70.
60. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thurlimann B, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO. 2013;24:2206-23.
61. Gnant M, Harbeck N, Thomssen C. St. Gallen 2011: Summary of the Consensus Discussion. Breast care (Basel, Switzerland). 2011;6:136-41.
62. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature. 2002;415:530-6.
63. Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, Quinn E, et al. Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer. J Clin Oncol. 2011;29:4273-8.
64. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med. 2004;351:2817-26.
65. Chia SK, Bramwell VH, Tu D, Shepherd LE, Jiang S, Vickery T, et al. A 50-gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen. Clin Cancer Res. 2012;18:4465-72.
66. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009;27:1160-7.
67. Ma XJ, Salunga R, Dahiya S, Wang W, Carney E, Durbecq V, et al. A five-gene molecular grade index and HOXB13:IL17BR are complementary prognostic factors in early stage breast cancer. Clin Cancer Res. 2008;14:2601-8.
68. Filipits M, Rudas M, Jakesz R, Dubsky P, Fitzal F, Singer CF, et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. Clin Cancer Res. 2011;17:6012-20.
69. Loi S, Haibe-Kains B, Desmedt C, Lallemand F, Tutt AM, Gillet C, et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. J Clin Oncol. 2007;25:1239-46.
70. Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst. 2006;98:262-72.
71. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet. 2005;365:671-9.
72. Chang HY, Sneddon JB, Alizadeh AA, Sood R, West RB, Montgomery K, et al. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. PLoS biology. 2004;2:E7.
73. Liu R, Wang X, Chen GY, Dalerba P, Gurney A, Hoey T, et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. N Engl J Med. 2007;356:217-26.
74. Lasfargues EY, Ozzello L. Cultivation of human breast carcinomas. J Natl Cancer Inst. 1958;21:1131-47.
75. Lasfargues EY, Coutinho WG, Redfield ES. Isolation of two human tumor epithelial cell lines from solid breast carcinomas. J Natl Cancer Inst. 1978;61:967-78.

76. Gazdar AF, Kurvari V, Virmani A, Gollahon L, Sakaguchi M, Westerfield M, et al. Characterization of paired tumor and non-tumor cell lines established from patients with breast cancer. Int J Cancer. 1998;78:766-74.
77. Hackett AJ, Smith HS, Springer EL, Owens RB, Nelson-Rees WA, Riggs JL, et al. Two syngeneic cell lines from human breast tissue: the aneuploid mammary epithelial (Hs578T) and the diploid myoepithelial (Hs578Bst) cell lines. J Natl Cancer Inst. 1977;58:1795-806.
78. Tanner M, Kapanen AI, Junttila T, Raheem O, Grenman S, Elo J, et al. Characterization of a novel cell line established from a patient with Herceptin-resistant breast cancer. Mol Cancer Ther. 2004;3:1585-92.
79. Cailleau R, Young R, Olive M, Reeves WJ, Jr. Breast tumor cell lines from pleural effusions. J Natl Cancer Inst. 1974;53:661-74.
80. Cailleau R, Olive M, Cruciger QV. Long-term human breast carcinoma cell lines of metastatic origin: preliminary characterization. In vitro. 1978;14:911-5.
81. Keydar I, Chen L, Karby S, Weiss FR, Delarea J, Radu M, et al. Establishment and characterization of a cell line of human breast carcinoma origin. Eur J Cancer. 1979;15:659-70.
82. Engel LW, Young NA, Tralka TS, Lippman ME, O'Brien SJ, Joyce MJ. Establishment and characterization of three new continuous cell lines derived from human breast carcinomas. Cancer Res. 1978;38:3352-64.
83. Soule HD, Maloney TM, Wolman SR, Peterson WD, Jr., Brenz R, McGrath CM, et al. Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. Cancer Res. 1990;50:6075-86.
84. Strand C, Enell J, Hedenfalk I, Ferno M. RNA quality in frozen breast cancer samples and the influence on gene expression analysis--a comparison of three evaluation methods using microcapillary electrophoresis traces. BMC Mol Biol. 2007;8:38.
85. Jonsson G, Staaf J, Olsson E, Heidenblad M, Vallon-Christersson J, Osoegawa K, et al. High-resolution genomic profiles of breast cancer cell lines assessed by tiling BAC array comparative genomic hybridization. Genes Chromosomes Cancer. 2007;46:543-58.
86. Knight SJ, Lese CM, Precht KS, Kuc J, Ning Y, Lucas S, et al. An optimized set of human telomere clones for studying telomere integrity and architecture. Am J Hum Genet. 2000;67:320-32.
87. Vissers LE, de Vries BB, Osoegawa K, Janssen IM, Feuth T, Choy CO, et al. Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. Am J Hum Genet. 2003;73:1261-70.
88. Bozzetti C, Nizzoli R, Guazzi A, Flora M, Bassano C, Crafa P, et al. HER-2/neu amplification detected by fluorescence in situ hybridization in fine needle aspirates from primary breast cancer. Ann Oncol. 2002;13:1398-403.
89. Saal LH, Troein C, Vallon-Christersson 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.
90. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res. 2002;30:e15.
91. Moore SR, Persons DL, Sosman JA, Bobadilla D, Bedell V, Smith DD, et al. Detection of copy number alterations in metastatic melanoma by a DNA fluorescence in situ hybridization probe panel and array comparative genomic hybridization: a southwest oncology group study (S9431). Clin Cancer Res. 2008;14:2927-35.

92. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. Nat Genet. 2004;36:949-51.
93. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat Med. 2009;15:907-13.
94. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc. 1995;57:289-300.
95. Wang D, Yan L, Hu Q, Sucheston LE, Higgins MJ, Ambrosone CB, et al. IMA: an R package for high-throughput analysis of Illumina's 450K Infinium methylation data. Bioinformatics. 2012;28:729-30.
96. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res. 2002;12:996-1006.
97. McCarty KS, Jr., Miller LS, Cox EB, Konrath J, McCarty KS, Sr. Estrogen receptor analyses. Correlation of biochemical and immunohistochemical methods using monoclonal antireceptor antibodies. Archives of pathology & laboratory medicine. 1985;109:716-21.
98. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. Clin Cancer Res. 2004;10:7252-9.
99. Farabegoli F, Santini D, Ceccarelli C, Taffurelli M, Marrano D, Baldini N. Clone heterogeneity in diploid and aneuploid breast carcinomas as detected by FISH. Cytometry. 2001;46:50-6.
100. Hicks J, Krasnitz A, Lakshmi B, Navin NE, Riggs M, Leibu E, et al. Novel patterns of genome rearrangement and their association with survival in breast cancer. Genome Res. 2006;16:1465-79.
101. Hicks J, Muthuswamy L, Krasnitz A, Navin N, Riggs M, Grubor V, et al. High-resolution ROMA CGH and FISH analysis of aneuploid and diploid breast tumors. Cold Spring Harb Symp Quant Biol. 2005;70:51-63.
102. Rennstam K, Baldetorp B, Kytola S, Tanner M, Isola J. Chromosomal rearrangements and oncogene amplification precede aneuploidization in the genetic evolution of breast cancer. Cancer Res. 2001;61:1214-9.
103. Ried T, Heselmeyer-Haddad K, Blegen H, Schrock E, Auer G. Genomic changes defining the genesis, progression, and malignancy potential in solid human tumors: a phenotype/genotype correlation. Genes Chromosomes Cancer. 1999;25:195-204.
104. Tanner MM, Karhu RA, Nupponen NN, Borg A, Baldetorp B, Pejovic T, et al. Genetic aberrations in hypodiploid breast cancer: frequent loss of chromosome 4 and amplification of cyclin D1 oncogene. Am J Pathol. 1998;153:191-9.
105. Tirkkonen M, Tanner M, Karhu R, Kallioniemi A, Isola J, Kallioniemi OP. Molecular cytogenetics of primary breast cancer by CGH. Genes Chromosomes Cancer. 1998;21:177-84.
106. Andre F, Job B, Dessen P, Tordai A, Michiels S, Liedtke C, et al. Molecular characterization of breast cancer with high-resolution oligonucleotide comparative genomic hybridization array. Clin Cancer Res. 2009;15:441-51.
107. Fridlyand J, Snijders AM, Ylstra B, Li H, Olshen A, Segraves R, et al. Breast tumor copy number aberration phenotypes and genomic instability. BMC Cancer. 2006;6:96.
108. Haverty PM, Fridlyand J, Li L, Getz G, Beroukhim R, Lohr S, et al. High-resolution genomic and expression analyses of copy number alterations in breast tumors. Genes Chromosomes Cancer. 2008;47:530-42.

109. Loo LW, Grove DI, Williams EM, Neal CL, Cousens LA, Schubert EL, et al. Array comparative genomic hybridization analysis of genomic alterations in breast cancer subtypes. Cancer Res. 2004;64:8541-9.
110. Naylor TL, Greshock J, Wang Y, Colligon T, Yu QC, Clemmer V, et al. High resolution genomic analysis of sporadic breast cancer using array-based comparative genomic hybridization. Breast Cancer Res. 2005;7:R1186-98.
111. Nessling M, Richter K, Schwaenen C, Roerig P, Wrobel G, Wessendorf S, et al. Candidate genes in breast cancer revealed by microarray-based comparative genomic hybridization of archived tissue. Cancer Res. 2005;65:439-47.
112. Arai K, Takano S, Teratani T, Ito Y, Yamada T, Nozawa R. S100A8 and S100A9 overexpression is associated with poor pathological parameters in invasive ductal carcinoma of the breast. Curr Cancer Drug Targets. 2008;8:243-52.
113. Chapman EJ, Kelly G, Knowles MA. Genes involved in differentiation, stem cell renewal, and tumorigenesis are modulated in telomerase-immortalized human urothelial cells. Mol Cancer Res. 2008;6:1154-68.
114. Cheng CJ, Lin YC, Tsai MT, Chen CS, Hsieh MC, Chen CL, et al. SCUBE2 suppresses breast tumor cell proliferation and confers a favorable prognosis in invasive breast cancer. Cancer Res. 2009;69:3634-41.
115. Cunha IW, Carvalho KC, Martins WK, Marques SM, Muto NH, Falzoni R, et al. Identification of genes associated with local aggressiveness and metastatic behavior in soft tissue tumors. Transl Oncol. 2010;3:23-32.
116. Fujita T, Ikeda H, Taira N, Hatoh S, Naito M, Doihara H. Overexpression of UbcH10 alternates the cell cycle profile and accelerate the tumor proliferation in colon cancer. BMC Cancer. 2009;9:87.
117. Hale LP, Price DT, Sanchez LM, Demark-Wahnefried W, Madden JF. Zinc alpha-2-glycoprotein is expressed by malignant prostatic epithelium and may serve as a potential serum marker for prostate cancer. Clin Cancer Res. 2001;7:846-53.
118. Hassan MI, Waheed A, Yadav S, Singh TP, Ahmad F. Prolactin inducible protein in cancer, fertility and immunoregulation: structure, function and its clinical implications. Cell Mol Life Sci. 2009;66:447-59.
119. Okamoto Y, Ozaki T, Miyazaki K, Aoyama M, Miyazaki M, Nakagawara A. UbcH10 is the cancer-related E2 ubiquitin-conjugating enzyme. Cancer Res. 2003;63:4167-73.
120. Berlingieri MT, Pallante P, Sboner A, Barbareschi M, Bianco M, Ferraro A, et al. UbcH10 is overexpressed in malignant breast carcinomas. Eur J Cancer. 2007;43:2729-35.
121. Gruvberger S, Ringner M, Chen Y, Panavally S, Saal LH, Borg A, et al. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. Cancer Res. 2001;61:5979-84.
122. Hyman E, Kauraniemi P, Hautaniemi S, Wolf M, Mousses S, Rozenblum E, et al. Impact of DNA amplification on gene expression patterns in breast cancer. Cancer Res. 2002;62:6240-5.
123. Reyal F, Stransky N, Bernard-Pierrot I, Vincent-Salomon A, de Rycke Y, Elvin P, et al. Visualizing chromosomes as transcriptome correlation maps: evidence of chromosomal domains containing co-expressed genes--a study of 130 invasive ductal breast carcinomas. Cancer Res. 2005;65:1376-83.
124. Knuutila S, Bjorkqvist AM, Autio K, Tarkkanen M, Wolf M, Monni O, et al. DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies. Am J Pathol. 1998;152:1107-23.

125. Myal Y, Gregory C, Wang H, Hamerton JL, Shiu RP. The gene for prolactin-inducible protein (PIP), uniquely expressed in exocrine organs, maps to chromosome 7. Somatic cell and molecular genetics. 1989;15:265-70.
126. Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell. 2006;10:529-41.
127. Lee JJ, Foukakis T, Hashemi J, Grimelius L, Heldin NE, Wallin G, et al. Molecular cytogenetic profiles of novel and established human anaplastic thyroid carcinoma models. Thyroid : official journal of the American Thyroid Association. 2007;17:289-301.
128. Orsetti B, Nugoli M, Cervera N, Lasorsa L, Chuchana P, Rouge C, et al. Genetic profiling of chromosome 1 in breast cancer: mapping of regions of gains and losses and identification of candidate genes on 1q. British journal of cancer. 2006;95:1439-47.
129. Takahashi Y, Ishii Y, Nishida Y, Ikarashi M, Nagata T, Nakamura T, et al. Detection of aberrations of ubiquitin-conjugating enzyme E2C gene (UBE2C) in advanced colon cancer with liver metastases by DNA microarray and two-color FISH. Cancer Genet Cytogenet. 2006;168:30-5.
130. Kong B, Michalski CW, Hong X, Valkovskaya N, Rieder S, Abiatari I, et al. AZGP1 is a tumor suppressor in pancreatic cancer inducing mesenchymal-to-epithelial transdifferentiation by inhibiting TGF-beta-mediated ERK signaling. Oncogene. 2010;29:5146-58.
131. Abdul-Rahman PS, Lim BK, Hashim OH. Expression of high-abundance proteins in sera of patients with endometrial and cervical cancers: analysis using 2-DE with silver staining and lectin detection methods. Electrophoresis. 2007;28:1989-96.
132. Bing C, Bao Y, Jenkins J, Sanders P, Manieri M, Cinti S, et al. Zinc-alpha2-glycoprotein, a lipid mobilizing factor, is expressed in adipocytes and is up-regulated in mice with cancer cachexia. Proceedings of the National Academy of Sciences of the United States of America. 2004;101:2500-5.
133. Brysk MM, Lei G, Adler-Storthz K, Chen Z, Brysk H, Tyring SK, et al. Zinc-alpha2-glycoprotein expression as a marker of differentiation in human oral tumors. Cancer Lett. 1999;137:117-20.
134. Freije JP, Fueyo A, Uria J, Lopez-Otin C. Human Zn-alpha 2-glycoprotein cDNA cloning and expression analysis in benign and malignant breast tissues. FEBS letters. 1991;290:247-9.
135. Gagnon S, Tetu B, Dube JY, Tremblay RR. Expression of Zn-alpha 2-glycoprotein and PSP-94 in prostatic adenocarcinoma. An immunohistochemical study of 88 cases. Am J Pathol. 1990;136:1147-52.
136. Irmak S, Tilki D, Heukeshoven J, Oliveira-Ferrer L, Friedrich M, Huland H, et al. Stage-dependent increase of orosomucoid and zinc-alpha2-glycoprotein in urinary bladder cancer. Proteomics. 2005;5:4296-304.
137. Lei G, Arany I, Selvanayagam P, Rajaraman S, Ram S, Brysk H, et al. Detection and cloning of epidermal zinc-alpha 2-glycoprotein cDNA and expression in normal human skin and in tumors. J Cell Biochem. 1997;67:216-22.
138. Sanchez LM, Vizoso F, Diez-Itza I, Lopez-Otin C. Identification of the major protein components in breast secretions from women with benign and malignant breast diseases. Cancer Res. 1992;52:95-100.
139. Tada T, Ohkubo I, Niwa M, Sasaki M, Tateyama H, Eimoto T. Immunohistochemical localization of Zn-alpha 2-glycoprotein in normal human tissues. The journal of

histochemistry and cytochemistry : official journal of the Histochemistry Society. 1991;39:1221-6.
140. Todorov PT, McDevitt TM, Meyer DJ, Ueyama H, Ohkubo I, Tisdale MJ. Purification and characterization of a tumor lipid-mobilizing factor. Cancer Res. 1998;58:2353-8.
141. He N, Brysk H, Tyring SK, Ohkubo I, Brysk MM. Zinc-alpha(2)-glycoprotein hinders cell proliferation and reduces cdc2 expression. Journal of cellular biochemistry Supplement. 2001;Suppl 36:162-9.
142. Townsley FM, Aristarkhov A, Beck S, Hershko A, Ruderman JV. Dominant-negative cyclin-selective ubiquitin carrier protein E2-C/UbcH10 blocks cells in metaphase. Proceedings of the National Academy of Sciences of the United States of America. 1997;94:2362-7.
143. Ghafouri B, Tagesson C, Lindahl M. Mapping of proteins in human saliva using two-dimensional gel electrophoresis and peptide mass fingerprinting. Proteomics. 2003;3:1003-15.
144. Bavi P, Uddin S, Ahmed M, Jehan Z, Bu R, Abubaker J, et al. Bortezomib stabilizes mitotic cyclins and prevents cell cycle progression via inhibition of UBE2C in colorectal carcinoma. Am J Pathol. 2011;178:2109-20.
145. Jones MD, Liu JC, Barthel TK, Hussain S, Lovria E, Cheng D, et al. A proteasome inhibitor, bortezomib, inhibits breast cancer growth and reduces osteolysis by downregulating metastatic genes. Clin Cancer Res. 2010;16:4978-89.
146. Chung CH, Parker JS, Karaca G, Wu J, Funkhouser WK, Moore D, et al. Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. Cancer Cell. 2004;5:489-500.
147. Lin P, Huang Z. Correlation analysis connects cancer subtypes. PLoS One. 2013;8:e69747.
148. Mollerstrom E, Kovacs A, Lovgren K, Nemes S, Delle U, Danielsson A, et al. Up-regulation of cell cycle arrest protein BTG2 correlates with increased overall survival in breast cancer, as detected by immunohistochemistry using tissue microarray. BMC Cancer. 2010;10:296.
149. Parris TZ, Danielsson A, Nemes S, Kovacs A, Delle U, Fallenius G, et al. Clinical Implications of Gene Dosage and Gene Expression Patterns in Diploid Breast Carcinoma. Clin Cancer Res. 2010.
150. Parris TZ, Kovacs A, Aziz L, Hajizadeh S, Nemes S, Semaan M, et al. Additive effect of the AZGP1, PIP, S100A8, and UBE2C molecular biomarkers improves outcome prediction in breast carcinoma. Int J Cancer. 2013.
151. Kernohan MD, Clark JR, Gao K, Ebrahimi A, Milross CG. Predicting the prognosis of oral squamous cell carcinoma after first recurrence. Archives of otolaryngology--head & neck surgery. 2010;136:1235-9.
152. Adelaide J, Chaffanet M, Imbert A, Allione F, Geneix J, Popovici C, et al. Chromosome region 8p11-p21: refined mapping and molecular alterations in breast cancer. Genes Chromosomes Cancer. 1998;22:186-99.
153. Theillet C, Adelaide J, Louason G, Bonnet-Dorion F, Jacquemier J, Adnane J, et al. FGFRI and PLAT genes and DNA amplification at 8p12 in breast and ovarian cancers. Genes Chromosomes Cancer. 1993;7:219-26.
154. Roque L, Rodrigues R, Martins C, Ribeiro C, Ribeiro MJ, Martins AG, et al. Comparative genomic hybridization analysis of a pleuropulmonary blastoma. Cancer Genet Cytogenet. 2004;149:58-62.
155. Thor AD, Eng C, Devries S, Paterakos M, Watkin WG, Edgerton S, et al. Invasive micropapillary carcinoma of the breast is associated with chromosome 8

abnormalities detected by comparative genomic hybridization. Human pathology. 2002;33:628-31.
156. Bass AJ, Watanabe H, Mermel CH, Yu S, Perner S, Verhaak RG, et al. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. Nat Genet. 2009;41:1238-42.
157. Boelens MC, Kok K, van der Vlies P, van der Vries G, Sietsma H, Timens W, et al. Genomic aberrations in squamous cell lung carcinoma related to lymph node or distant metastasis. Lung cancer (Amsterdam, Netherlands). 2009;66:372-8.
158. Mahmood SF, Gruel N, Nicolle R, Chapeaublanc E, Delattre O, Radvanyi F, et al. PPAPDC1B and WHSC1L1 Are Common Drivers of the 8p11-12 Amplicon, Not Only in Breast Tumors But also in Pancreatic Adenocarcinomas and Lung Tumors. Am J Pathol. 2013.
159. Tonon G, Wong KK, Maulik G, Brennan C, Feng B, Zhang Y, et al. High-resolution genomic profiles of human lung cancer. Proceedings of the National Academy of Sciences of the United States of America. 2005;102:9625-30.
160. Williams SV, Platt FM, Hurst CD, Aveyard JS, Taylor CF, Pole JC, et al. High-resolution analysis of genomic alteration on chromosome arm 8p in urothelial carcinoma. Genes Chromosomes Cancer. 2010;49:642-59.
161. Simon R, Richter J, Wagner U, Fijan A, Bruderer J, Schmid U, et al. High-throughput tissue microarray analysis of 3p25 (RAF1) and 8p12 (FGFR1) copy number alterations in urinary bladder cancer. Cancer Res. 2001;61:4514-9.
162. Garcia MJ, Pole JC, Chin SF, Teschendorff A, Naderi A, Ozdag H, et al. A 1 Mb minimal amplicon at 8p11-12 in breast cancer identifies new candidate oncogenes. Oncogene. 2005;24:5235-45.
163. Gelsi-Boyer V, Orsetti B, Cervera N, Finetti P, Sircoulomb F, Rouge C, et al. Comprehensive profiling of 8p11-12 amplification in breast cancer. Mol Cancer Res. 2005;3:655-67.
164. Melchor L, Garcia MJ, Honrado E, Pole JC, Alvarez S, Edwards PA, et al. Genomic analysis of the 8p11-12 amplicon in familial breast cancer. Int J Cancer. 2007;120:714-7.
165. Ray ME, Yang ZQ, Albertson D, Kleer CG, Washburn JG, Macoska JA, et al. Genomic and expression analysis of the 8p11-12 amplicon in human breast cancer cell lines. Cancer Res. 2004;64:40-7.
166. Yang ZQ, Streicher KL, Ray ME, Abrams J, Ethier SP. Multiple interacting oncogenes on the 8p11-p12 amplicon in human breast cancer. Cancer Res. 2006;66:11632-43.
167. Paterson AL, Pole JC, Blood KA, Garcia MJ, Cooke SL, Teschendorff AE, et al. Co-amplification of 8p12 and 11q13 in breast cancers is not the result of a single genomic event. Genes Chromosomes Cancer. 2007;46:427-39.
168. Bernard-Pierrot I, Gruel N, Stransky N, Vincent-Salomon A, Reyal F, Raynal V, et al. Characterization of the recurrent 8p11-12 amplicon identifies PPAPDC1B, a phosphatase protein, as a new therapeutic target in breast cancer. Cancer Res. 2008;68:7165-75.
169. Holland DG, Burleigh A, Git A, Goldgraben MA, Perez-Mancera PA, Chin SF, et al. ZNF703 is a common Luminal B breast cancer oncogene that differentially regulates luminal and basal progenitors in human mammary epithelium. EMBO Mol Med. 2011;3:167-80.
170. Sircoulomb F, Nicolas N, Ferrari A, Finetti P, Bekhouche I, Rousselet E, et al. ZNF703 gene amplification at 8p12 specifies luminal B breast cancer. EMBO Mol Med. 2011;3:153-66.

171. Turner N, Pearson A, Sharpe R, Lambros M, Geyer F, Lopez-Garcia MA, et al. FGFR1 amplification drives endocrine therapy resistance and is a therapeutic target in breast cancer. Cancer Res. 2010;70:2085-94.
172. Yang ZQ, Liu G, Bollig-Fischer A, Giroux CN, Ethier SP. Transforming properties of 8p11-12 amplified genes in human breast cancer. Cancer Res. 2010;70:8487-97.
173. Zhang X, Mu X, Huang O, Xie Z, Jiang M, Geng M, et al. Luminal Breast Cancer Cell Lines Overexpressing ZNF703 Are Resistant to Tamoxifen through Activation of Akt/mTOR Signaling. PLoS One. 2013;8:e72053.
174. Karlsson E, Danielsson A, Delle U, Olsson B, Karlsson P, Helou K. Chromosomal changes associated with clinical outcome in lymph node-negative breast cancer. Cancer Genet Cytogenet. 2007;172:139-46.
175. Möllerstrom E, Delle U, Danielsson A, Parris T, Olsson B, Karlsson P, et al. High-resolution genomic profiling to predict 10-year overall survival in node-negative breast cancer. Cancer Genet Cytogenet. 2010;198:79-89.
176. Stec I, van Ommen GJ, den Dunnen JT. WHSC1L1, on human chromosome 8p11.2, closely resembles WHSC1 and maps to a duplicated region shared with 4p16.3. Genomics. 2001;76:5-8.
177. Bird A. DNA methylation patterns and epigenetic memory. Genes & development. 2002;16:6-21.
178. Costello JF, Plass C. Methylation matters. Journal of medical genetics. 2001;38:285-303.
179. Fuks F. DNA methylation and histone modifications: teaming up to silence genes. Curr Opin Genet Dev. 2005;15:490-5.
180. Heard E, Disteche CM. Dosage compensation in mammals: fine-tuning the expression of the X chromosome. Genes & development. 2006;20:1848-67.