NTNU – Trondheim
Norwegian University of
Science and Technology

Molecular profiling of Ductal Carcinoma *In Situ*

Hanne Håberg Mørk

Chemical Engineering and Biotechnology
Submission date: June 2012
Supervisor: Per Bruheim, IBT
Co-supervisor: Therese Sørlie, Oslo Universitetssykehus, Radiumhospitalet

Norwegian University of Science and Technology
Department of Biotechnology

Acknowledgements

The work presented in this thesis has been performed in the Department of Genetics, Institute for Cancer Research at The Norwegian Radium Hospital Spring 2012. This thesis was made possible due to Research Scientist Therese Sørlie, who I am very grateful to. I would like to express sincere gratitude to her for being an excellent supervisor, always encouraging and available for scientific guidance. I am grateful to Professor Anne-Lise Børresen-Dale, the head of the Department of Genetics, for opening the research facilities to master students to huge inspiration and motivation.

I would like to thank PhD Candidate Miriam Ragle Aure, the engineers Eldri Undlien Due, Phuong Vu, Rita Halvorsen and Anja Valen for taking the time to teach and help me in the laboratory and research associate Daniel Nebdal and PhD Candidate Robert Lesurf for invaluable help and guidance with different software. I wish to express gratitude to Dr. Fredrik Warnberg, Professor Ingrid Gribbestad's group and Research Scientist Vessela Kristensen's group for providing valuable tumor material and clinical information. I gratefully acknowledge Associate Professor Per Bruheim for being my contact supervisor at the Department of Biotechnology, Norwegian University of Science and Technology.

Finally, I would like to thank everybody in the Department of Genetics for contributing to a positive and inspirational work environment making it a great pleasure to write this thesis.

Oslo, June 2012

Hanne Håberg Mørk

Abstract

Breast cancer develops through multiple stages from hyperplasia to invasive and finally metastatic disease. Ductal carcinoma *in situ* (DCIS) is an abnormal proliferation of epithelial cells within the milk ducts in the breast without invasion beyond the basement membrane. The incidence of DCIS accounts for about 20-25% of newly diagnosed breast cancer cases. Some *in situ* lesions are believed to rapidly transit to invasive ductal carcinoma (IDC), while others remain unchanged or disappear. Nowadays, women who would never experience invasive breast cancer are undergoing unnecessary and potentially harmful treatment. Studies have revealed that the invasive phenotype of breast cancer is determined at the preinvasive stages of the tumor. Molecular studies of DCIS are therefore important in order to identify those lesions that have a greater risk of developing into invasive disease.

The objective of this thesis was to characterize *in situ* and invasive breast carcinomas by gene expression profiling. Differences in gene expression within DCIS and between DCIS and invasive breast carcinomas were examined to gain insights about molecular mechanisms underlying tumor progression and to identify potential progression markers. 58 tumor tissues from 37 pure DCIS and 21 pure invasive cancers were subjected to microarray gene expression analysis using Agilent One-Color Microarray 8×60K.

Hierarchical clustering proved that the samples related more to subtype than diagnosis. The most significant genes separating the invasive cancers from DCIS were found to be involved in functions related to the extracellular matrix and tumor-stromal interaction. A subgroup of eight DCIS tumors separated from the other DCIS by high expression of genes characteristic of the invasive tumors. These genes could be potential progression markers if validated in other studies. Heterogeneity was observed among the DCIS patients and two subgroups of *in situ* lesions were clearly differentiated based on upregulated immune response. Elevated levels of immune signaling were found in HER2+, basal-like, normal-like and luminal B subtypes, but were completely absent in luminal A tumors. The suppressing role of the immune system compared with the promoting role needs to be further investigated, and could potentially increase our knowledge concerning the progression of *in situ* lesions to invasive breast cancer.

Sammendrag

Brystkreft utvikles gjennom flere stadier fra hyperplasia til invasiv og til sist metastatisk sykdom. Duktalt karsinom *in situ* (DCIS) er en unormal celledeling av epitelcellene i melkekanalene i brystet uten invasjon utover basalmembranen. Forekomsten av DCIS utgjør om lag 20-25% av nydiagnostiserte brystkrefttilfeller. Noen *in situ* lesjoner antas å raskt utvikle seg til invasiv duktalt karsinom (IDC), mens andre forblir uendret eller forsvinner. I dag gjennomgår kvinner som aldri kommer til å utvikle invasiv brystkreft unødvendig og potensielt skadelig behandling. Undersøkelser har avdekket at den invasive fenotypen av brystkreft antagelig bestemmes allerede ved preinvasive stadier av svulsten. Molekylære studier av DCIS er derfor viktig for å identifisere de lesjoner som har en større risiko for å utvikle seg til invasiv sykdom.

Målet med denne avhandlingen var å karakterisere *in situ* og invasive karsinomer ved å måle genekspresjonsnivået i tumorvevet. Forskjeller innad i DCIS og mellom DCIS og invasiv karsinomer ble undersøkt for å få innsikt i deres molekylære profiler og for å identifisere potensielle markører for progresjon. 58 tumorvev fra 37 rene DCIS og 21 rene invasive duktale karsinom ble målt ved hel-genom genekspresjonsmikromatrise analyse (Agilent Human GE 8×60K microarray).

Alle tumorene ble klassifisert til en av de fem molekylære subtypene med PAM50. Hierarkisk klustering viste at det var mer sammenheng mellom prøvene av samme subtype enn samme diagnose. De mest signifikante genene som skilte de invasive duktale karsinomene fra DCIS viste seg å være involvert i funksjoner relatert til ekstracellulær matrix og tumor-stromal interaksjon. En undergruppe av åtte DCIS svulster skilte seg ut fra de andre DCIS tumorene ved høye uttrykk av gener karakteristiske for de invasive svulstene. Disse genene kan være potensielle progresjonsmarkører dersom de kan valideres i andre studier. Heterogenitet ble observert blant DCIS pasientene og to undergrupper av *in situ* lesjoner var tydelig forskjellige basert på oppregulert immunrespons. Forhøyede nivåer av immunsignal ble funnet i HER2+, basal-liknende, normal-liknende og luminal B subtyper, men var helt fraværende i luminal A subtyper. Den kreftdempende rollen til immunsystemet sammenlignet med den kreftfremmende må undersøkes videre, og kan potensielt øke vår kunnskap om utviklingen av *in situ* lesjoner til invasiv brystkreft.

Contents

Chapter 1

Introduction

1.1 Breast cancer

1.1.1 Incidence

Breast cancer affects the lives of millions of women worldwide. Over the past several decades, the incidence of breast cancer has increased, while the death rate has steadily decreased [1]. This observation may be explained by increased mammographic screening and early detection of the preinvasive stages of breast cancer. In 2009, 2760 Norwegian women were diagnosed with breast cancer [1]. The disease is most frequent among women older than 50 years. The prognosis of the patients is mostly dependent on tumor stage at the time of diagnosis. Without spread to axillary lymph nodes, five year survival rate is reported to be 95 %. The survival rate is decreasing to 18 % if distant metastasis is present [2].

1.1.2 Breast cancer types

Breast cancers are separated after origin of disease. The mammary gland consists of lobules (milk producing glands) and branching ducts (milk channels) (Figure 1.1). The ends of the ducts are termed the terminal ductal–lobular units (TDLUs) [3]. Most breast cancers are thought to arise in the TDLU. Tumors that arise in the ducts or the lobules are termed ductal carcinoma and lobular carcinoma, respectively. The TDLUs consist of two types of epithelial cells: the inner luminal epithelial cells and the outer myoepithelial cells. Luminal epithelial cells line the normal breast duct and have secretory properties. Myoepithelial cells have both contractile muscle and epithelial properties. The two cell types are distinct and the precursors to various forms of breast cancer. Cancers from luminal epithelial cells are most common. The basement membrane surrounds the epithelial

cells and works as a mechanical barrier. Its function is to anchor the epithelial layer to the connective tissue underneath.

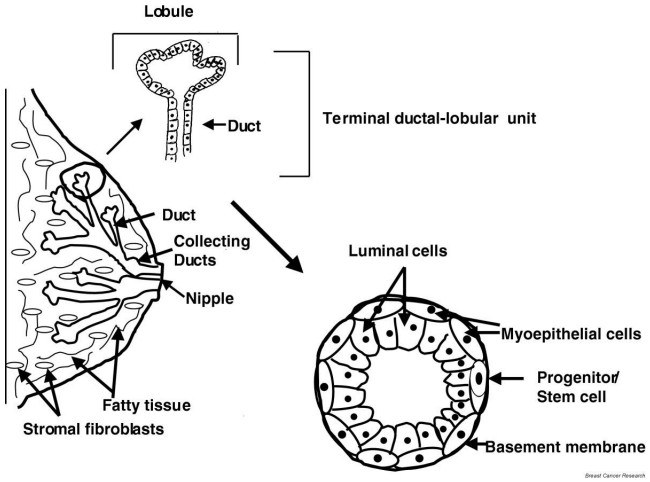

Figure 1.1: **Anatomy of the human breast.** The mammary gland consists of ducts and lobules surrounded by fatty connective tissue and fibroblasts. The terminal ductal–lobular unit (TDLU) is the unit thought to be the origin of most breast cancers and consists of a hollow central lumen, the inner luminal epithelial cell layer, the outer basal myoepithelium cell layer and the surrounding basement membrane. (Modified after [3]).

Breast cancers can be either non-invasive or invasive. Invasive breast cancer cells have penetrated the basement membrane and invaded surrounding breast tissue. At this point, the cancer cells have the ability to spread to the lymph nodes and blood stream and metastasize to all organs of the body. Among the invasive forms of breast cancer, invasive ductal carcinoma (IDC) accounts for 70-80% of all breast carcinomas, while invasive lobular carcinoma (ILC) is the second most common type and accounts for 10-20% of all breast cancer cases. Other less common invasive types make up the remaining percents and include mucinous carcinoma, papillary carcinoma and tubular carcinoma among others [2].

Preinvasive breast cancer possesses some malignant properties, but is still confined to its original site and has not broken through the basement membrane. The term *in situ* means "in place", which characterize both ductal carcinoma *in situ* (DCIS) and lobular carcinoma *in situ* (LCIS). Ductal carcinoma *in situ* accounts for 20-25% of all breast cancer cases detected by mammography screening [2].

1.1.3 Breast cancer progression

Progression from normal epithelial cells in the duct wall of the breast to metastatic cancer cells is thought to develop through multiple stages (Figure 1.2). Atypical ductal hyperplasia (ADH) is the precursor to DCIS and is used to describe increased proliferation of

the epithelial cells [4]. Progression from ADH to DCIS marks the transition from benign stage to malignant disease. Some DCIS lesions are believed to rapidly transit to invasive ductal carcinoma, while others remain unchanged or disappear [5].

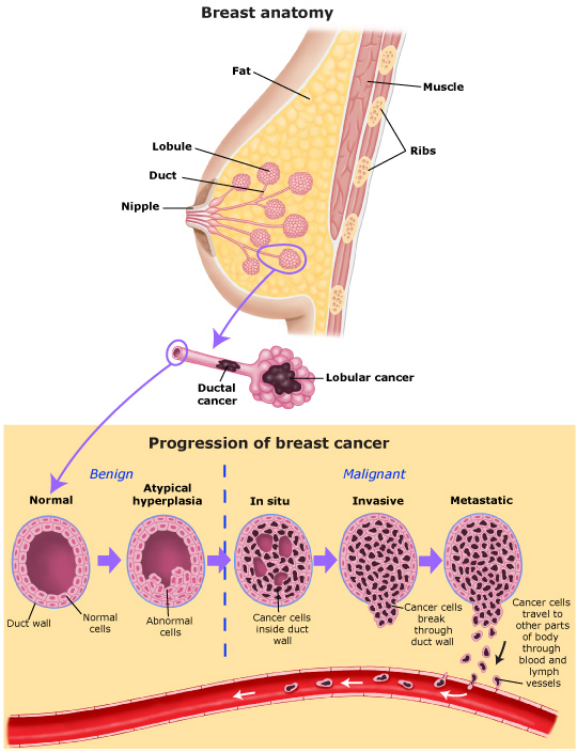

Figure 1.2: **The multistep progression of ductal carcinoma.** A schematic view of the progression of normal epithelial cells in the duct wall through hyperplasia, *in situ* to invasive and metastatic disease. *In situ* ductal breast cancer is a non-invasive form of ductal breast cancer that consists of a clonal proliferation of malignant epithelial cells that accumulate within the lumen of the breast duct. Invasive forms of breast cancer cells have the ability to become metastatic by travelling to other organs of the body through blood and lymph vessels (Modified after [6]).

1.1.4 Histopathological features

Breast cancer is a complex and heterogeneous disease with distinct histopathological features [7]. Markers for classifying the different types of breast cancer involve tumor type, tumor grade, tumor stage, expression of hormone receptors (estrogen and progesterone) and HER2 receptor status.

1.1.4.1 Tumor grading

Histological grade describes proliferation and differentiation of breast cancer cells and is considered an important prognostic factor. The grading system is based on three morphological features: 1) mitotic count (rate of cell division), 2) tubule formation and

3) nuclear pleomorphism (change in cell size and uniformity) [8]. Each feature is assigned a score from 1 to 3, indicating slow and fast cell growth, respectively. All invasive forms of breast cancer are graded after the same criteria [9].

1.1.4.2 Tumor staging

Tumor staging is useful to estimate breast cancer prognosis. The TNM (tumor-node-metastasis) system was developed by Pierre Denoix in 1942 and is based on size of the primary tumor (T), spread to axillary lymph nodes (N) and presence of distant metastases (M) [10]. There are five tumor stages. Stage 0 represents non-invasive breast cancer (DCIS). Stage I describes small tumors that are localized to the breast. Stage II describes larger tumors with possible spread to the axillary lymph nodes. Large tumors that have invaded tissues around the breast fall into Stage III and Stage IV represents disease with metastases throughout the body. Stage 0/I/II patients have a significantly better prognosis than Stage III/IV patients [10].

1.1.4.3 Cellular receptor status

A hormone receptor positive breast cancer expresses any or both of the hormone receptors; estrogen receptor (ER) and/or progesterone receptor (PgR). Interaction between the female hormones and the hormone receptor stimulates proliferation, and this growth can be down-regulated by the use of hormone receptor inhibitors, such as tamoxifen. Estrogen receptor positive breast cancers also tend to be progesterone receptor positive, but exceptions occur. Hormone positive and negative tumors are associated with differences in survival. Patients with ER and PgR negative tumors have an increased risk of mortality compared to ER and PgR positive tumors [11].

HER2 (Human Epidermal growth factor Receptor 2) is a growth factor receptor found to be overexpressed in 25-30% of human breast cancers and is strongly associated with increased disease recurrence and worse prognosis [12]. Amplification of the HER2 protein stimulates cell growth and specific drugs have been invented to target and inhibited its activity. The most successful drug so far is Trastuzmab (Herceptin), a monoclonal antibody targeting the HER2 and inhibiting growth of the tumor [13].

1.1.5 "Intrinsic" gene expression classification

The heterogeneous properties characterizing breast cancer are reflected by genomic variations. Advanced microarray technology and complete sequencing of the human genome have made it possible to classify human breast tumors into subtypes based on their gene

expression patterns [14]. Five main molecular classes of breast cancer have been proposed to give an improved classification, regardless of stage. These five "intrinsic" subtypes are termed luminal A, luminal B, basal-like, HER2+ and normal-like breast cancer (Figure 1.3). Luminal A subtype is typically ER positive (ER+), PgR positive (PgR+) and often low-grade. Luminal B subtype is typically ER+ and PgR+, often high grade and may be HER2 positive (HER2+). Basal-like subtype is often ER negative (ER-), PgR negative (PgR-) and HER2 negative (HER2-) (sometimes referred to as "triple-negative" tumors). The different subtypes are conserved across populations and various diagnoses and found to be associated with different clinical outcome. Patients with tumors expressing ER (luminal A subtype) show a better prognosis, than patients with tumors of basal-like and HER2+ subtype [15].

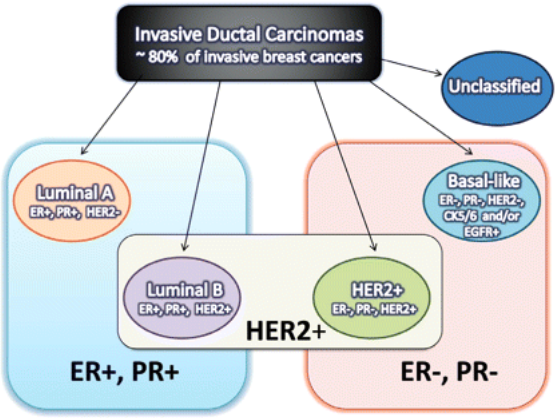

Figure 1.3: **"Intrinsic" gene expression classification of breast cancer subtypes**. The blue and pink rectangles group the subtypes based on ER and PgR status. Luminal A and luminal B subtypes express hormone receptors and are often distinguished by a negative and positive expression of HER2 receptor, respectively. HER2+ and basal-like subtypes are most often negative for ER and PgR, and distinguished by a positive and negative expression of the HER2 receptor. The different subtypes are also represented among DCIS tumors (Modified after [16]).

Risk models have been developed to incorporate the gene expression–based "intrinsic" subtypes to estimate prognosis. One example is the Breast BioClassifier® based on PAM50 (prediction analysis of microarrays), a 55-gene RT-qPCR assay [17]. The 55 genes were found from the originally "intrinsic" gene list of 534 genes [18]. The classifier is yielding risk of relapse based on tumor size and the molecular subtypes.The current utility of this model is to identify those patients with a very favorable outcome, who could be spared of adjuvant chemotherapy [19]. Nowadays, therapeutic decisions are mostly influenced by status of hormone receptors, HER2, grade and stage of the disease. In near future, gene expression profiling is thought to provide complementary prognostic and predictive information. Gene signatures, together with clinical and pathological factors

will help facilitate development of targeted drugs in specific groups of patients and lead to more individually tailored treatment [20, 19].

1.2 Breast cancer at the cellular and molecular level

1.2.1 Cancer hallmarks

Six common traits ("hallmarks") among cancers have been identified to explain the transformation of normal cells to cancer cells. These hallmarks are: (1) sustained proliferative signaling; (2) evasion of growth suppressors; (3) resisting cell death (apoptosis); (4) stimulation of growth of blood vessels (angiogenesis); (5) unlimited growth (immortality); and (6) the capacity to invade surrounding tissue and spread to distant sites (metastasis) (Figure 1.4) [21]. Recently, two emerging hallmarks have been added to the list: reprogramming of the energy metabolism and evasion of the immune system together with two categories of enabling characteristics: genome instability and mutation and tumor-promoting inflammation [21].

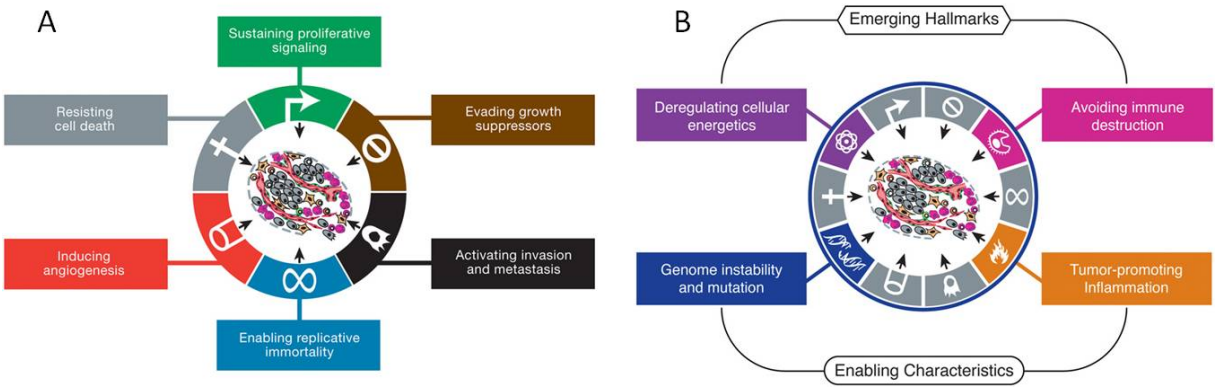

Figure 1.4: **The hallmarks of cancer.** Common alterations in cell physiology are essential for the development of cancer. A) The six essential hallmarks of cancer proposed by Hanahan and Weinberg in 2000. B) Two emerging hallmarks and characteristics of cancer proposed in 2011 by the same authors (Modified after [21]).

More emphasis has been put on the importance of studying the tumor microenvironment. Within a tumor there are specialized cell types that collectively contribute to tumorigenesis and progression of the tumor. Myoepithelial cells comprise a thin layer within the basement membrane in the duct wall of the breast and together with inflammatory cells constitute the intraluminal tumor microenvironment of preinvasive breast cancer. Fibroblasts and myofibroblasts, components of the connective tissue, inflammatory cells and endothelial cells constitute the stromal microenvironment of invasive breast cancer (Figure 1.5) [4].

Inflammatory cells, such as tumor associated macrophages (TAMs) can both suppress and promote tumor progression. By presenting tumor antigens to cytotoxic T-cells, the T-cells becomes cytotoxic to the cancer cells. On the other hand, TAMs can promote tumor growth by secreting breast tumor mitogens or stimulate tumor angiogenesis and metastasis. Recent studies have revealed the role of TAMs to be more promoting than suppressing in breast cancer [22].

Microenvironmental processes include loss of myoepithelial cells [23], epithelial-mesenchymal transition and angiogenesis [21]. The epithelial-mesenchymal transition is a hypothetical process where epithelial cells with low mobility loose their cell adhesion properties and acquire mesenchymal properties and higher mobility. Several lines of evidence suggest that expression of EMT-related genes correlate with invasive behavior and a poor prognosis of breast cancer [24].

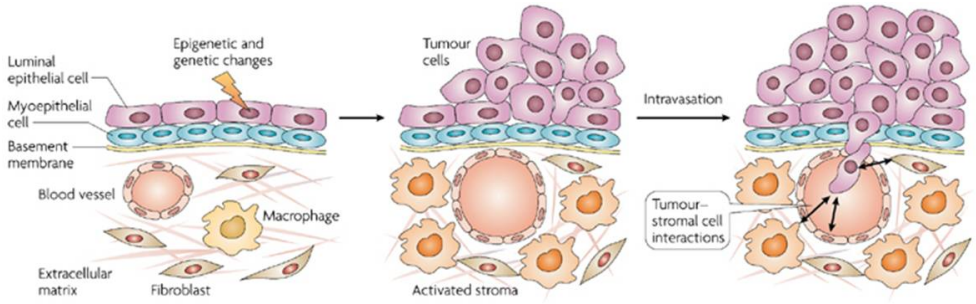

Figure 1.5: **Schematic illustration of the tumor-microenvironment interaction during invasion.** The influence and contribution of microenvironmental cells (fibroblasts, blood vessels and immune cells) have proven to be very important during the progress of invasion from *in situ* lesions to invasive breast cancer [25]. (Modified after [7])

1.2.2 Genomic alterations

Over the past decade remarkable progress have been made towards understanding the molecular mechanisms underlying each hallmark. Genome instability generates the genetic diversity that make up these hallmark functions. Genetic alterations are seen at both the nucleotide level and the chromosomal level and can be divided into four major categories: 1) Gene mutations 2) Alterations in chromosome number 3) Chromosome translocations and 4) Gene amplifications [26].

Gene mutations involve base substitutions/deletions/insertions of one or a few nucleotides, which alter the function of the protein expressed. This form of instability is not the most common. Alterations in chromosome number that involve losses or gains of whole chromosomes (aneuploidy) are changes found in nearly all major human tumor types. Chromosome translocations result in fusion of different chromosomes which can give rise

to fusion of two different genes. Gene amplification is frequently found in cancers and can be a result of chromosome translocations and aneuploidy.

Alterations in genes that are responsible for tumorigenesis are grouped into three types: oncogenes, tumor-suppressor genes and stability genes. Oncogenes promote cell growth, and if amplified or mutated, they are constitutionally expressed and promote cell proliferation. A frequently overexpressed oncogene in breast cancer is *HER2.* It causes amplified amounts of the HER2 protein, which leads to increased cell growth. Tumor-suppressor genes suppress cell growth and loss of function results in uncontrolled cell division. *TP53* is a commonly known tumor suppressor gene, also known as "the guardian of the genome" and is mutated in about 30 % of invasive breast cancers [27]. Stability genes are genes involved in the repair system of the DNA. Mutations in the repair machinery of the DNA increases the frequency of mutations in other genes during replication [26]. Two well known DNA repair genes involved in breast cancer are *BRCA1* and *BRCA2* (BReast-CAncer susceptibility gene 1 and 2). Mutations in these genes cause a rapid accumulation of mutations during carcinogenesis [28].

Alterations in cells are not always caused by changes in the DNA sequence. Significant alterations in gene expression pattern without obvious genetic alterations can be explained by epigenetics. Epigenetic modifications are changes in the gene expression pattern caused by DNA methylation, histone modification and RNA-associated silencing [29]. Epigenetic mechanisms have proven to play an important role in cancer behavior and an increased focus is directed towards this field of study.

1.3 Ductal carcinoma *in situ* (DCIS)

1.3.1 Histological grade

Several classifications of DCIS have been proposed based on nuclear grade and necrosis [30]. The Van Nuys classification [9] combines low and intermediate grades into a non-high grade category and the remaining into a high grade category. Group 1 characterizes non-high nuclear grade lesions without necrosis, group 2: non-high nuclear grade lesions with necrosis and group 3: all high nuclear grade lesions [31]. The most common classification of DCIS was published by The European Organization for Research and Treatment of Cancer (EORTC) in 1994 [32]. This classification defines three grades of differentiated DCIS termed poorly (high grade), intermediately, and well (low grade). Two criteria: 1) cytonuclear differentiation and 2) architectural differentiation constitute the classification and have been found to be more consistent than previously used criteria of architectural pattern and the presence/absence of necrosis [32].

1.3.2 Cellular receptor status

Positive ER status has been found in DCIS tumors to range from 60% to 78% [33]. ER expression seems to correlate with DCIS grade similarly to IDC tumors. HER2 is amplified and overexpressed in 32-55% of DCIS lesions and often correlated with an aggressive type [34]. So far, little evidence has supported the connection between ER, PgR and HER2 status and local recurrence of DCIS lesions [35].

1.3.3 Prognosis and treatment

DCIS shows a considerable level of variety in terms of histological features and aggressiveness making classification of DCIS challenging. DCIS was initially misdiagnosed as benign. Later studies have suggested that between 14 and 53% of DCIS may progress to invasive cancer over a period of 10 or more years [36]. Still, diagnosed with DCIS give a good prognosis. While scientists are trying to find biomarkers which can distinguish between those non-invasive lesions with a good prognosis and those transforming to a potentially life threatening invasive cancer, treatment of DCIS results in overtreatment of some patients [37].

Today, there is a move away from mastectomy (total removal of breast) to local excision of the tumor for DCIS patients [38]. Much debate is ongoing concerning whether adjuvant therapy is needed for all patients. Tamoxifen has shown to be effective in ER positive patients [33]. Lumpectomy (breast conserving surgery) followed by radiation therapy has proven more effective than lumpectomy alone, in preventing both invasive and noninvasive ipsilateral tumors. Still, the combination of lumpectomy and radiation therapy has not had an impact on the rate of regional or distant recurrence. Up to 15% of women will experience recurrence of cancer in the same breast (ipsilateral), with 50% of these being of an invasive type [33].

For patients with *in situ* cancer, the Van Nuys prognostic index (VNPI) is applied to aid in the complex treatment decision process [31]. Four independent predictors of local recurrence such as tumor size, surgical margin width, pathological classification and patient age are combined to give a total score ranging from 4 to 12. Patients with scores of 4, 5 or 6 can be considered for treatment with excision of the tumor only. Patients with intermediate scores 7, 8, or 9 are considered for additional treatment with radiation therapy and patients with high scores of 10, 11, or 12 often exhibit extremely high local recurrence rates, regardless of radiation therapy, and are considered for mastectomy [39].

1.3.4 Progression from *in situ* to invasive breast cancer

Enhancing the understanding of the mechanisms that underlie metastatic progression is crucial since metastasis is the principal cause of mortality. Some DCIS lesions rapidly transit to invasive ductal carcinoma (IDC), while others remain unchanged or disappear [40]. Good prognostic markers that could distinguish between aggressive and non-aggressive DCIS, remain to be identified. Advanced gene expression analysis over the past several years have increased the understanding of what happens at the molecular level during early preinvasive stages of breast cancer [4].

Studies have revealed that the most noticeable transcriptional changes occur at the transition from normal breast epithelium to atypical ductal hyperplasia and that the alterations are conserved throughout the later stages of progression through DCIS and IDC [41]. This finding suggests that the progressive potential of the lesion may be predicted from the gene-expression patterns expressed in the preinvasive stages (ADH and DCIS).

Several studies have demonstrated that the progression from DCIS to IDC is associated with subsets of genes that are found to be consistently overexpressed and linked to increased tumor grade and progression. Low-grade DCIS and ADH lesions have been found to possess distinct gene expression signatures associated with the ER phenotype and a better prognosis, while high-grade DCIS lesions possess a gene expression signature associated with increased cell proliferation and invasive growth behavior, promoting malignancy and metastasis [41].

1.3.5 Microenvironment and progression

Several molecular studies indicate that the tumor microenvironment plays an important role in both promoting and inhibiting the invasive process of breast cancer [4]. Both gene expression and epigenetic data suggest that the stromal and myoepithelial microenvironment in preinvasive breast cancer participates in the transition to invasive disease.

DCIS-associated myoepithelial cells show upregulation of genes encoding proteases and chemokines, when compared with normal myoepithelial cells. The proteases cathepsin F, K, L, matrix metalloproteinase 2 (MMP2) and chemokines CXCL12/SDF-1 and CXCL14 have been proposed regulators of cell growth, migration and invasion [42, 43, 25].

Hu *et al.* [23] used a cell line model for DCIS and demonstrated that the transition from DCIS til IDC was promoted by fibroblasts and inhibited by normal myoepithelial cells. They proved that myoepithelial cell differentiation required interaction between pathways involving TGFβ, Hedgehog, cell adhesion and p63. Loss of the myoepithelial cell differentiation accelerated the invasive process.

Chapter 2

Objectives of the thesis

Studies have revealed that the invasive phenotype of breast cancer is determined at the preinvasive stages of the tumor. Molecular studies of DCIS are therefore important in order to identify those lesions that have a greater risk of developing into invasive disease. All contributions to achieve increased knowledge about DCIS biology will in the future spare patients from unnecessary extensive treatment.

The overall objective of this thesis was to characterize *in situ* and invasive breast carcinomas by gene expression profiling. 38 DCIS and 24 small invasive ductal carcinomas were subjected to gene expression analysis. Differences in gene expression within DCIS and between DCIS and invasive breast carcinomas were examined to obtain insights about molecular mechanisms underlying tumor progression and to identify potential progression markers that could distinguish the aggressive DCIS from those of a more benign phenotype.

Chapter 3

Materials and methods

3.1 Patient material

38 fresh frozen DCIS tumor tissues were obtained from the Fresh Tissue Biobank, Department of Pathology, Uppsala University Hospital, Sweden, the Breast Cancer Tissue Bank, MR Cancer group, St.Olavs Hospital, Trondheim and Akershus University Hospital, Lørenskog. In addition, 24 cases of invasive ductal carcinoma (size<15mm) were selected from Uppsala and St.Olavs Hospital. Three samples from normal breast tissue were included as controls.

All samples were subjected to microarray gene expression analysis with 58 out of 62 samples being successfully analyzed. The resulting patterns were examined for differences among DCIS tumors and between DCIS and invasive breast carcinomas. Patient and tumor characteristics are summarized in table 3.1. Detailed patient data and additional information on the methods are provided as supplementary information (Appendix A and Appendix B).

Table 3.1: **Patient and tumor characteristics**

	DCIS (n=37)		**IDC (n=21)**	
Age, years (median)	56.6		60.7	n/a:15
Grade, I/II/III	2/8/19	n/a:8	5/10/6	
Receptor status	**Number (%)**		**Number (%)**	
ER+	15 (41%)	n/a:13 (35%)	15 (71%)	
PgR+	12 (32%)	n/a:13 (35%)	14 (67%)	
HER2+	8 (22%)	n/a:17 (46%)	6 (29%)	n/a:10 (48%)

n/a - not available

3.2 RNA isolation

Two procedures were performed to isolate total RNA from each sample. Both procedures involved the use of TRIzol® Reagent (Invitrogen, Life Technologies™, USA). The reagent is a monophasic solution of phenol and guanidine isothiocyanate suitable for isolating total RNA from cells and tissues [44]. During sample homogenization, TRIzol® Reagent maintains the integrity of the RNA, while disrupting cells and dissolving cell components. Addition of chloroform followed by centrifugation separates the solution into an aqueous and organic phase, where RNA remains in the aqueous phase. The modified method of RNA isolation involves the use of RNeasy mini columns [45], whereas the original method recovers RNA by precipitation with isopropanol [44].

Several precautions including use of disposable gloves, sterile plasticware and automatic pipettes reserved for RNA isolation were taken to prevent RNAse and microbial contamination. The work area and equipment were washed with ethanol and RNAse Away prior to isolation. Both procedures were performed in a chemical fume hood.

3.2.1 Total RNA isolation with TRIzol

Procedure

Homogenization Upon use, TRIzol® was stored 30 minutes at room temperature. Tissue samples stored at -80 ^{o}C were brought up on dry ice and cut into small pieces. Tumor tissue was homogenized in 500 µl of TRIzol® Reagent using one steel ball and Mixer Mill for 2 minutes at 30/s.

Phase separation Homogenized samples were incubated for 5-10 minutes at room temperature to permit complete dissociation of nucleoprotein complexes. 100 µl of chloroform per 500 µl of TRIzol® was added and tubes vigorously shaken for 15 seconds following 5 minutes of incubation at room temperature and centrifugation at 12000xg for 15 minutes at 4 ^{o}C.

RNA precipitation After centrifugation, the upper aqueous phase was transferred to a fresh RNAse free tube, while the remaining organic phase was stored for possible isolation of DNA or protein. RNA was precipitated by mixing with 250 µl isopropanol per 500 µl of TRIzol® used and incubated at 4 ^{o}C for approximately 30 minutes. The samples were centrifuged at 12000xg for 10 minutes at 4 ^{o}C.

RNA wash The supernatant was carefully removed and the pellet washed twice with 500 µl of 75% ethanol per 500 µl of TRIzol® used. The content was mixed by pipetting up and down prior centrifugation at 7500xg for 5 minutes at 4 ^{o}C.

Redissolving the RNA At the end of the procedure , the RNA pellet was briefly air-dried for 10 minutes at room temperature before dissolved in RNAse-free water. Amount of water was determined based on the amount of tissue and size of pellet. The samples were incubated for 5-10 minutes at 55 ^{o}C and kept on ice before stored at -80 ^{o}C.

3.2.2 Modified method for isolation of total RNA using TRIzol reagent and RNeasy mini columns

The RNeasy procedure combines the selective binding properties of a silica-based membrane with microspin technology. A specialized high-salt buffer system allows up to 100 µg of RNA to bind to the RNeasy silica membrane. Ethanol is added to provide appropriate binding conditions for the RNA and contaminants are efficiently washed away [45].

Procedure

Homogenization and Phase separation The homogenization and phase separation step was identical to the original method of RNA isolation except the amount of TRIzol® Reagent. 550 µl of TRIzol® Reagent was used for homogenization and another 550 µl added after use of the Mixer Mill.

RNA purification with RNeasy mini columns After addition of 220 µl chloroform, mixing and centrifugation, the upper aqueous phase (560 µl) was transferred to a new tube and added 840 µl of 100% ethanol. The content was mixed thoroughly by pipetting up and down several times. 700 µl of the sample was transferred to RNeasy Mini spin columns and centrifuged at 8000xg for 30 seconds at room temperature. The flow-through was discarded. The step was repeated using the remainder of the sample and the flow-through discarded. 350 µl of RW1 buffer was added into the RNeasy Mini spin column and the samples centrifuged for 30 seconds at 8000xg. The flow-through was discarded. The step was repeated. 500 µl of RPE buffer was added into the RNeasy Mini spin column and the samples centrifuged for 30 seconds at 8000xg. The flow-through was discarded. Another 500 µl of RPE buffer was added to the RNeasy Mini spin column and the samples centrifuged for 2 min at 8000xg to dry the RNeasy Mini spin column membrane. The long centrifugation dries the spin column membrane, ensuring that no ethanol is carried

over during RNA eluation. Residual ethanol may interfere with downstream reactions. The RNeasy Mini spin column was removed from the collection tube carefully to avoid carryover of ethanol.

RNA eluation The RNeasy Mini spin column was placed into a new 2 ml collection tube and centrifuged in a microcentrifuge at full speed for 1 minute. This step was performed to eliminate any possible carryover of RPE buffer. The RNeasy Mini spin column was placed in a new 1.5 ml collection tube. 30 µl of RNase-free water was pipetted directly onto the RNeasy Mini spin column membrane and the sample centrifuged for 1 minute at 8000xg to elute the RNA.

3.3 RNA quality control

3.3.1 RNA quantification

RNA quantification and quality were assessed using NanoDrop® ND-1000 Spectrophotometer version 3.7.1 (NanoDrop Technologies, Thermo Fisher Scientific Inc., USA) for measurement of RNA concentration and purity. Fiber optic technology and surface tension allow the sample to be held in place between two optical surfaces [46].

Procedure The NanoDrop software was turned on and the program Nucleic Acid was chosen. The upper and lower optical surfaces of the microspectrophotometer were cleaned with RNAse-free water by pipetting 1.5 µl onto the lower optical surface. The lever arm was closed to bathe the upper optical surface and both pedestals were wiped off with special lens-cleaning tissue before measurement. The instrument was initialized with 1.5 µl of nuclease-free water. After initialization, RNA-40 was chosen as sample type and a blank measurement of 1.5 µl of nuclease-free water recorded prior to sample measurement. 1.5 µl of each sample was pipetted directly onto the pedestal (Figure 3.1). The lever arm was closed, a sample column formed and recorded by clicking measure. Between each measurement, the surfaces were wiped with special lens-cleaning tissue to prevent sample carryover and residue buildup. A final cleaning with RNAse-free water was performed after last measurement. The readings were saved and printed.

Figure 3.1: **NanoDrop® ND-1000 Spectrophotometer.** The sample is applied directly to the lower optical surface for measurement of RNA concentration and purity (Modified after [47]).

3.3.2 RNA quality assessment by degradation

RNA integrity number (RIN), an algorithm for assigning integrity values to RNA measurements, is standard for RNA quality assessment. RNA quality check of integrity was determined using Bioanalyzer 2100 (Agilent Technologies, USA) [48].

Procedure

Preparation of gel matrix A gel matrix was taken out of the fridge at room temperature for 30 minutes, protected from light. A heating block was set to 70°C. 550 µL of gel matrix was added to a spin filter and spun for 10 minutes at 4000 rpm. 65 µL of filtered gel was aliquoted into 0.5 mL RNase free tubes. The aliquot was stored at 4 ^{o}C and used within one month.

Preparation of gel-dye matrix A RNA 6000 Nano dye concentrate and a RNA 6000 Nano marker were left in room temperature protected from light for 30 minutes before use. The RNA 6000 Nano dye concentrate was vortexed for 10 seconds, spun down and 1 µL added to an aliquoted filtered gel. The content was vortexed thoroughly for 10 seconds and centrifuged at 13000 rpm for 10 minutes at room temperature. The prepared gel-dye mix could be used for two chips within one day. The mix was re-spun for ten minutes before second time of use if more than one hour had passed.

Loading the chip with the gel matrix and RNA 6000 Nano Marker A new RNA Nano chip was placed on "chip priming station" and 9 µL of gel-dye matrix loaded into the well marked "G" with black background. A timer was set for 30 seconds and the plunger at the 1 mL mark. The chip priming station was closed and the plunger pressed until it was held by the syringe clip. After exactly 30 seconds the plunger was released. After 5 more seconds the plunger was slowly pulled back up to 1 mL. The chip priming station was reopened and additional 9 µL of the gel-dye matrix loaded in the two other wells marked "G". 5 µL of RNA 6000 nano marker was loaded into the well marked ladder and to each of the twelve RNA wells.

Loading the RNA sample and the RNA 6000 ladder The RNA samples were incubated in the 70° C heating block for 2 minutes and put on ice for 5 minutes. The samples were briefly centrifuged to clear any condensate from the tube's walls and cap. 1 µL of ladder (centrifuged) was loaded into the well marked "ladder" and 1 µL of RNA sample was loaded into each of the 12 wells. The loaded chip had to be used within 5 minutes due to evaporation of reagents.

Running the chip The Nano Chip was placed in the adapted vortex and mixed for 1 minute at 2400 rpm. Liquid spill at the top of the chip was carefully removed with a tissue. Two electrode cleaners were filled, one with 350 µL RNAse-free water and one with 350 µL RNase Zap. The electrode cleaner with RNase Zap was placed in the machine for 1 minute, removed and replaced with the electrode cleaner with water for 10 seconds. The machine was opened for 10 seconds for the electrodes to dry. The loaded Nano Chip was placed in the Bioanalyzer and the right assay selected. "Start" was pressed and the sample names entered.

Cleaning of the electrodes The Nano Chip was removed from the Bioanalyzer right after the run was finished. The electrodes were cleaned with 350 µL of RNAse-free water for 10 seconds and left opened to dry for 10 seconds.

3.4 Microarray gene expression analysis

Microarray gene expression analysis is a powerful tool to simultaneously measure the expression of thousand of genes from a single cell. Fluorescent cDNA is generated by reverse transcriptase from total RNA, with input range from 10-200 ng. With the use of T7 RNA polymerase, total RNA is amplified 100-fold and simultaneously labeled with cyanine 3-CTP for visualization (Figure 3.2) [49].

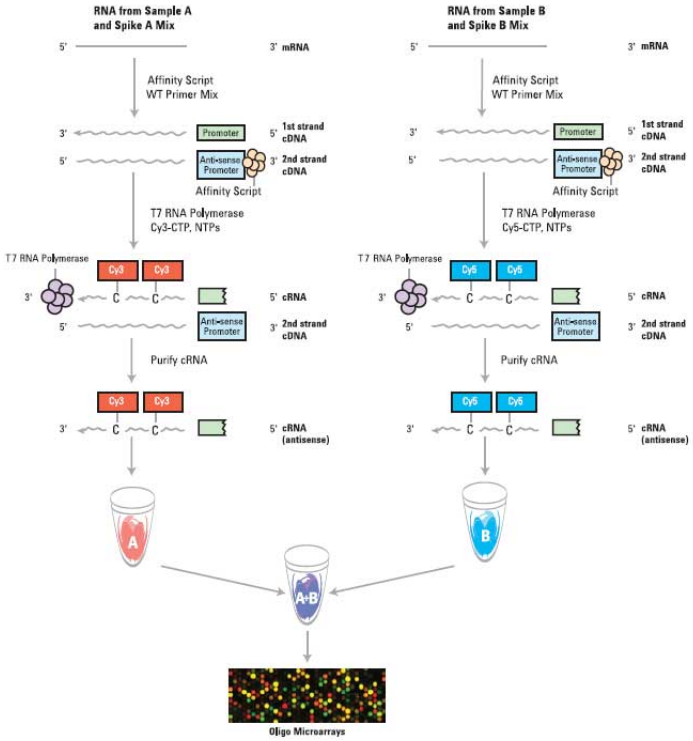

Figure 3.2: **Schematic of two color microarray gene expression procedure.** Single color (Cyanine-3-CTP) was used in this project. mRNA input is converted to cDNA by reverse transcriptase and a promoter incorporated at the poly-A tail. cRNA is amplified from the cDNA strand and cyanine-3-CTP simultaneously incorporated by RNA polymerase. cRNA is purified, hybridized to probes on the microarray and visualized by fluorescens (Modified after [49]).

Agilent manufacturing of microarrays is made possible with SurePrint technology [50]. The technology allows 8 arrays to reside on a single slide. The microarrays are manufactured using a proprietary non-contact industrial inkjet printing process. Oligo monomers are deposited uniformly onto specially-prepared glass slides. The *in situ* synthesis process prints 60-mer length oligonucleotide probes, base-by-base, from digital sequence files. The inkjet process enables the delivery of extremely small, accurate volumes (picoliters). The reactions involve standard phosphoramidite chemistry (Figure 3.3) and the process is completed without stops and contact with the slide surface.

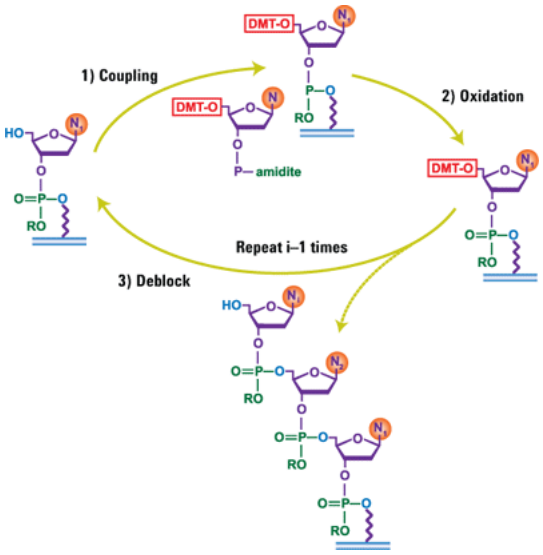

Figure 3.3: **SurePrint technology.** SurePrint uses phosphoramidite chemistry in the synthesis of the 60-mer length oligonucleotide probes. The process is repeated 60 times (Modified after [50]).

In this project, total RNA (100 ng) from tumor samples was amplified and labeled using Low RNA Input Quick Amp Labeling Kit (Agilent Technologies, USA). The procedure was followed by the protocol One-Color Microarray-Based gene Expression Analysis, Low Input Quick Amp Labeling, Version 6.5 (Agilent Technologies, USA).

Procedure

3.4.1 Sample preparation

Three heat-blocks were set to 37 ^{o}C, 65 ^{o}C and 80 ^{o}C. A water-bath was heated to 40 ^{o}C. First dilutions spike mixes (1:20) containing a volume of 3 µL was prepared in advance. The One-Color Spike Mix stock solution was vigorously mixed on a vortex mixer, heated for 37 ^{o}C for 5 minutes and vortexed once more before a brief centrifugation. The One-Color Spike Mix stock solution was diluted with Dilution buffer provided in the Spike-In kit. The first dilutions spike mixes were stored at -80 ^{o}C upon use.

Step 1 Preparation of Spike Mix First dilution of One-Color Spike Mix was thawed, heated at 37 ^{o}C for 5 minutes and vortexed before a brief centrifugation. Second dilution was created in a new tube with 2 µL input of First dilution and 48 µL of Dilution buffer. The content was mixed on a vortex mixer and spun down. Third dilution was created in a new tube with 2 µL input of Second dilution and 38 µL of Dilution buffer. The content was mixed on a vortex mixer and spun down. After use the dilutions were discarded.

Step 2 Preparation of labeling reaction A T7 Promoter Primer Mix was prepared by adding 8 µL of T7 Promoter Primer to a tube with 5 µL of RNAse-free water. The Primer Mix was vortexed, spun down and kept on ice. 100 ng of total RNA was added to a tube in a final volume of 2 µL. Concentrated samples were diluted with RNAse-free water. 2 µL of Third dilution Spike Mix was added to each tube together with 1.3 µL of T7 Promoter Primer Mix. The samples were incubated at 65 ^{o}C for 10 minutes to denature the primer and the template. Afterwords, the samples were put on ice and incubated for 5 minutes and spun quickly.

A cDNA Master Mix was prepared immediately prior to use by adding the components listed in table 3.2. The 5X First strand buffer had been prewarmed at 80 ^{o}C for 4 minutes to ensure adequate re-suspension of the buffer components and kept at room temperature before use. The AffinityScript RNase Block Mix is a blend of enzymes and was kept on ice prior to use.

Table 3.2: **cDNA Master Mix**

Components	Volume (µL)
5X first strand buffer	20 µL
0.1 M DTT	10 µL
10 mM dNTP mix	5 µL
AffinityScript RNase Block Mix	12 µL
Total volume	47 µL

4.7 µL of cDNA Master Mix was added to each sample tube and mixed by pipetting up and down. The samples were quickly spun down and then incubated at 40 ^{o}C for 2 hours in a circulating water bath to generate cDNA. The samples were moved to a heat-block of 70 ^{o}C for 15 minutes to inactivate the AffinityScript enzyme. The samples were placed on ice for 5 minutes and spun briefly.

A Transcription Master Mix was prepared in room temperature, immediately prior to use, by gently mixing the components as listed in table 3.3. The T7 RNA polymerase blend is a blend of enzymes and was kept on ice prior to use.

Table 3.3: **Transcription Master Mix**

Components	Volume (µL)
5X first Transcription buffer	32 µL
RNAse-free water	7.5 µL
NTP mix	10 µL
0.1 M DTT	6 µL
T7 RNA polymerase blend	2.1 µL
Cyanine-3-CTP	2.4 µL
Total volume	60 µL

6 µL of Transcription Master Mix was added to each sample tube, gently mixed by pipetting and spun down. The samples were incubated at 40 ^{o}C for 2 hours in a circulating water bath to amplify cRNA and incorporate cyanine-3-CTP. The samples were spun down and stored at -80 ^{o}C before purification.

Step 3 Purification of labeled/amplified RNA Purification of amplified cRNA samples were performed with Qiagens's RNeasy mini spin columns, where the total RNA was bound to a silica-based membrane and contaminants were efficiently washed away. The procedure was performed at room temperature and as quick and dark as possible. A centrifuge was pre-cooled to 4 ^{o}C. 84 µL of nuclease-free water was added to all cRNA samples. 350 µL of RLT buffer was added and mixed by pipetting. 250 µL of EtOH (100%) was added to give appropriate binding conditions and mixed thoroughly by pipetting. Total volume of 700 µL for each sample was transferred to a RNeasy mini column in a 2 mL collection tube. The samples were centrifuged at 13,000 rpm at 4 ^{o}C for 30 seconds. The flow-through was discarded. The RNeasy mini columns were added 500 µL of RPE buffer (containing EtOH). The samples were centrifuged at 13,000 rpm at 4 ^{o}C for 30 seconds and the flow-through discarded. The process was repeated, and the samples were centrifuged at 13,000 rpm at 4 ^{o}C for 60 seconds. The RNeasy mini columns were centrifuged at 13,000 rpm at 4 ^{o}C for additionally 30 seconds to remove any remaining traces of RPE buffer. Cleaned cRNA was eluted by transferring the RNeasy columns to new collection tubes and added 30 µL of nuclease-free water directly onto the RNeasy filter membrane. The samples were incubated for 1 minute, then centrifuged at 13,000 rpm at 4 ^{o}C for 30 seconds. The final flow-through contained the cRNA and was kept on ice. The RNeasy mini spins columns were discarded.

Step 4 Quantification of cRNA Quantification of cRNA was performed using NanoDrop® ND-1000 Spectrophotometer version 3.7.1 (NanoDrop Technologies) as described in section 3.3.1. Microarray Measurement was chosen instead of Nucleic Acid.

3.4.2 Hybridization

Step 1 Preparation of 10X Blocking Agent The 10X Blocking agent was prepared in advance. 500 µL of nuclease-free water was added to the vial containing lyophilized 10X Blocking agent. The content was mixed by gently vortexing and heated for 4-5 minutes at 37 ^{o}C prior to centrifugation for 5-10 seconds. The 10X Blocking agent was stored at -80 ^{o}C.

Step 2 Preparation of hybridization samples A heat block was set to 60 ^{o}C and the hybridization oven turned on 65 ^{o}C. The 10X Blocking agent was thawed, mixed by gently vortexing and heated for 4-5 minutes at 37 ^{o}C prior centrifugation for 5-10 seconds. For each microarray (8-pack), each of the components as indicated in table 3.4 were added to a new tube.

Table 3.4: **Fragmentation mix for 8-pack microarray formats**

Components	Volume/Mass 8-pack microarray
Cyanine 3-labeled, linearly amplified cRNA	600 ng (counted in µL)
10X Blocking agent	5 µL
Nuclease-free water	x µL
25X Fragmentation buffer	1 µL
Total volume	25 µL

The samples were incubated for exactly 30 minutes at 60 ^{o}C to fragment RNA. After incubation the samples were immediately cooled on ice for one minute. 25 µL of 2xGEx Hybridization buffer HI-RPM were added to each sample tube to stop the fragmentation reaction. The content was mixed by careful pipetting to avoid introducing bubbles. The samples were centrifuged at 13,000 rpm at room temperature for 60 seconds to drive the sample content off the walls and lid and to aid in bubble reduction. The centrifugation was repeated if necessary. The samples were placed on ice and loaded onto the array as soon as possible.

Step 3 Preparation of hybridization assembly A clean gasket slide was loaded into the Agilent SureHyb chamber base with the label facing up. 45 µL of each sample was dispensed slowly onto the gasket well without letting the pipette tip or the sample solution touch the gasket walls. An array with the Agilent-labeled barcode facing down was slowly placed down onto the SureHyb gasket slide keeping the two slides parallel at all times. The SureHyb chamber cover was placed onto the sandwiched slides and the clamp hand-tighten onto the chamber. The assembled chamber was inspected for stationary bubbles and knocked if necessary. Low volume content was noted. The chamber was placed in a hybridization oven at 65 ^{o}C and 10 rpm for 17 hours.

3.4.3 Washing

2 mL of Triton X-102 was added to Gene expression wash buffer 1 and 2 before use. A dish was filled with Gene expression wash buffer 2 and added a magnetic stir bar and prewarmed to 38 ^{o}C. Two other dishes were filled with Gene expression wash buffer 1, a magnetic stir bar and kept at room temperature. A slide rack was placed into

dish number 2. The hybridization chamber was removed from the oven and inspected for volume content and stationary bubbles. The hybridization chamber was disassembled and the array-gasket sandwich quickly transferred to dish number 1 containing wash buffer 1. The slides were only touched at the ends and separated from the barcode end with the use of a forceps. The gasket slide was left in dish number 1, while the microarray slide was moved to dish number 2 and placed into the slide rack with wash buffer 1. After all slides were placed in the slide rack, the stirring was started. The slides were washed 1 minute in wash buffer 1 at 6 rpm, then the slide rack was quickly moved to wash buffer two and washed for 1 minute at 6 rpm. The slide rack was slowly removed (5 to 10 seconds) to minimize droplets on the slides. The slides were put in a slide holder with the Agilent barcode facing up and immediately scanned to minimize environmental oxidants influence on the signal intensities.

3.4.4 Scanning and Feature Extraction

The assembled slide holders were put into the scanner carousel. The slot number of the first slide was chosen "start slot" and the slot number for the last slide was chosen "end slot". The profile "AgilentG3_GX-1Color" was selected. The scan settings for one-color scans were set as informed by table 3.5. The scanning was started by clicking "Scan Slot" in the Scan Control main window.

Table 3.5: **Scanner settings**

	For 8×60K G3 Microarray format
Dye channel	Green
Scan region	Scan area (61×21.6 mm)
Scan resolution	3
Tiff	20 bit

After generation of microarray scan .tif images, the data were extracted with the use of Agilent Feature Extraction Software to obtain information about the probe features. The data were added to the FE Project by clicking "Add New Extraction Set(s)". After successful extraction, the QC (Quality Control) reports for all samples were available for inspection.

3.5 Data analysis

3.5.1 Quality control

GeneSpring GX software version 12 (Agilent Technologies, USA) was used in the quality control of the gene expression microarray data. The software is designed to give accessible statistical tools for fast visualization and understanding of the microarray data within a biological context [51]. All successfully analyzed samples (n=58) were used as input for analysis in GeneSpring GX. The analysis was performed in an Advanced workflow to be able to choose different normalization methods.

Normalization Normalization is necessary to adjust microarray data for effects which arise from variation in the technology rather than from biological differences [52]. The goal is to minimize the systematic non-biological differences and reveal true biological differences. Sources of technical variation include unequal quantities of starting RNA and differences in hybridization and manufacturing between chips. Step 1 involves transforming the signal values to the log base 2. After that two normalization options are possible for one-color data: Percentile shift and Quantile normalization. Both normalization procedures were performed on the dataset to investigate how the two types of normalization would impact downstream analyzes.

Percentile shift normalization arrange the log transformed signal values in increasing order. The rank of the 75th percentile is computed and the samples centered thereafter. A 50th percentile can also be chosen, but some genes tend to be lower expressed (values close to zero) and a 75th percentile ensures a representative distribution of signals.

Quantile normalization rank the data within a sample, calculates the median intensity for each quantile and replaces the raw data with the mean intensity. The method assumes that there is an underlying common distribution of intensities across arrays and that the distribution of gene abundance is the same for all samples. The data are therefore normalized so the distribution of probe intensities for each array in the set is the same [51].

Filter probeset by expression The probeset was filtered based on the probes signal intensity values. Very low signal values or saturated probes were removed. The range of intensity value was chosen to a upper percentile cutoff of 100% and a lower percentile cutoff of 20%, excluding some of the probes with very low signal values.

Filter probeset by flags Flags are used to denote the quality of the probes [51]. The flags are specific for the array type used and hence the flag notation was determined by

Agilent Technologies. The probes were filtered based on the flag values Detected and Not detected.

Principal component analysis Principal component analysis (PCA) has proven to be a powerful tool in detection of trends, patterns and grouping among samples as well as detection of outliers. It is an unsupervised dimension reduction technique, where patterns in a complex dataset can be visualized by projecting a large set of variables to a smaller set of variables. This reduction of dimensionality makes it possible to visualize complex data in a three dimensional plot. The new set of variables are termed principal components (PCs) and they represent a linear combination of the original variables that are independent of each other. The projections of the samples onto the PCs are defined as scores and similar samples will group together in clusters in a score plot. The first and second PC capture the maximum variation of the data [51]. The dataset was class labeled and colored to visualize separation of samples in the dataset based on different parameters (diagnosis, RIN, array batch number and subtype). No unsatisfactory samples were detected as outliers or removed from further analysis.

3.5.2 Significance analysis of microarrays (SAM)

Significance analysis of microarrays (SAM) (Stanford University, USA) is a statistical tool for finding significant genes in the gene expression data. The gene expression measurements from all microarray experiments together with class grouping of diagnosis were used as input to Excel, where SAM operates. SAM computed a statistic deviation for each gene, reflecting the strength of the relationship between gene expression and the class variable. Repeated permutations (100) of the data were used to determine if the expression of any genes was significantly related to the class grouping. The cutoff for significance was determined by a tuning parameter delta. The value of delta was chosen to give a false discovery rate (FDR) less than 5% [53]. Gene information such as functions and annotations was found from SOURCE, a scientific database developed at Stanford University, USA [54].

3.5.3 Hierarchical clustering

Hierarchical clustering was performed in Cluster version 3.0 (University of Tokyo, Japan) and Java TreeView version 1.1.6 (Alok, USA) to identify and visualize patterns within the dataset. The program Cluster organize and analyze the data, while TreeView allows the data to be visualized. With the use of clustering algorithms the samples and probes were grouped based on similarities in the expression profile [55]. Two-dimensional hierarchical

clustering was performed using Pearson correlation uncentered as distance function which involve: (1) no mean-centering of probes, (2) positive and negative correlated probes are close to 1 and -1, respectively and (3) unrelated probes are close to zero [56]. The probeset was filtered based on standard deviation to exclude the probes of least variance. Average linkage and median centering were chosen parameters. Hierarchical clustering was performed on the dataset including all samples (IDC and DCIS) and on the dataset restricted to pure DCIS samples. Both unsupervised and supervised clustering were used. In the unsupervised method all genes were included, while supervised clustering involved input of significant genes derived from SAM.

3.5.4 Molecular subclassification with PAM50

All samples were assigned to one of the five molecular subtypes by correlation to the expression centroids using PAM50. This work was performed by a fellow collaborator in the department, Robert Lesurf.

3.5.5 Ingenuity pathway analysis (IPA)

Ingenuity pathway analysis (IPA) (Ingenuity Systems, USA) is a web-based software with a large database that contains information about biological and chemical relationships extracted from scientific literature [57]. The software was used to gain insight about the cellular processes the upregulated and downregulated genes would relate to. The most significant cellular and molecular functions were further explored and the significant molecules viewed in a pathway context.

3.5.6 Gene Ontology analysis

DAVID (Database for Annotation, Visualization and Integrated Discovery) was used to understand the functionality of the genes derived from SAM in terms of gene ontology (GO). Gene ontology provides a controlled vocabulary of defined terms describing gene product properties independent of species. The ontology covers: (1) Cellular component, (2) molecular function and (3) biological process. Each GO term has a name indicating the domain/category it belongs to. The p-value calculated for each GO term indicates the likelihood of genes of interest to fall into a category by chance [51]. Each p-value was adjusted by the Benjamini-Hochberg procedure to correct for multiple testing.

Chapter 4

Results

4.1 Quality control

37 *in situ* and 21 invasive breast carcinomas, 58 out of 62 tumor tissues, were successfully analyzed by gene expression microarrays. Four samples (one DCIS, three IDC) failed the analysis and were excluded from the study due to low quality of RNA and subsequently poor amplification of cRNA. Bioanalyzer results are provided in Appendix C for closer examination. The successfully analyzed samples passed all quality control criteria and were used for further analysis. Two normalization procedures were tested on the data; 75th percentile and quantile normalization.The results of normalization and filtering of probeset can be visualized in Figure 4.1 and Figure 4.2.

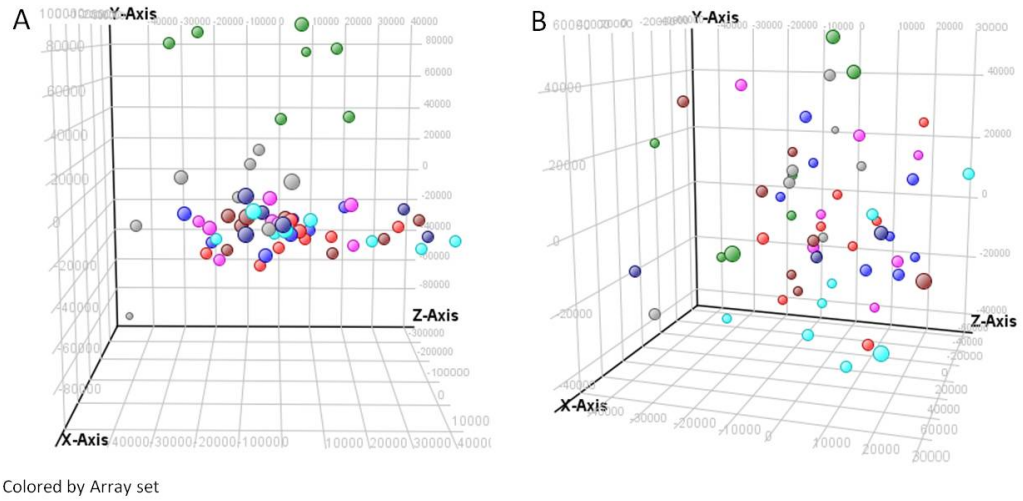

Figure 4.1: **Effects of normalization and filtering on array batch differences.** (A) After normalization (75th percentile), two array batches, colored by green and grey, respectively, separated from the remaining six. (B) After filtering the probeset by expression values and flags, the separation was eliminated.

Coloring the samples based on array batch (8 samples) after normalization helped to identify two distinct arrays that separated from the remaining six. After filtering the data on low expression values and flags, the differences between array batches were eliminated (Figure 4.1). Coloring the samples based on RNA integrity number (RIN) after normalization revealed a clear grouping of samples with low RIN. This grouping could not be eliminated after filtering the data on low expression values and flags (Figure 4.2), indicating a clear effect of degraded RNA.

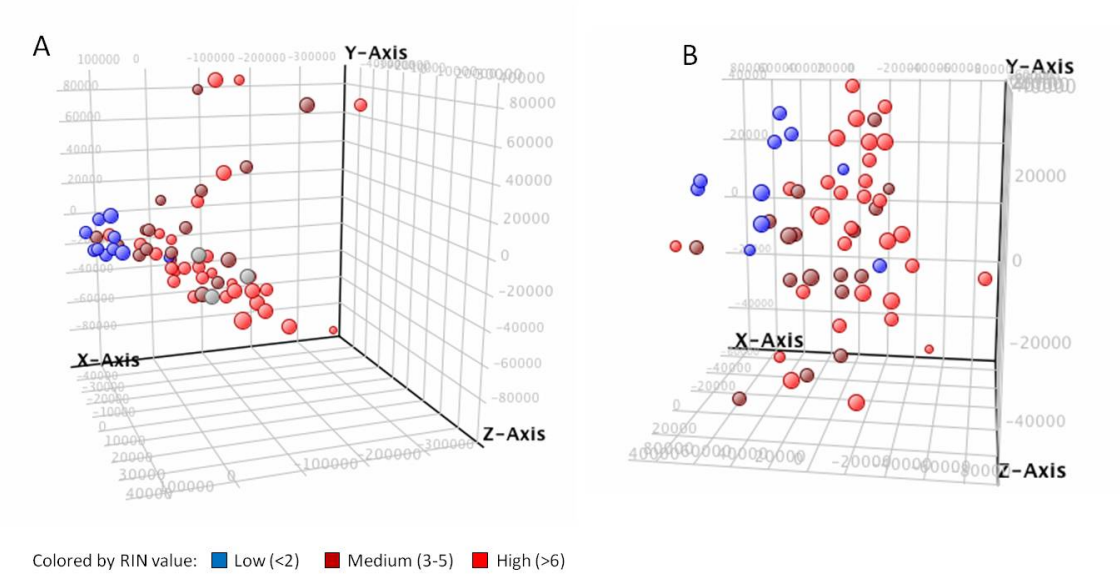

Figure 4.2: **Effects of normalization and filtering on differences caused by low RNA integrity number.** (A) After normalization (75th percentile), samples with low RIN (blue) clearly clustered together. (B) After filtering the probeset by expression values and flags, the grouping was still visible.

4.2 Molecular subclassification with PAM50

The PAM50 "intrinsic" gene classifier was used to assign all tumors to one of the five molecular subtypes. Quantile normalized data were used for the classification. Both groups of diagnosis were found to contain all five subtypes. The distribution of subtypes within DCIS versus IDC was as follows: Luminal A (22% vs. 19%), luminal B (22% vs. 38%), HER2+ (27% vs. 10%), basal-like (13% vs. 14%) and normal-like (16% vs. 19%). PCA plot labeling the samples according to molecular subtype revealed a tendency of the samples to cluster together based on subtype (Figure 4.3).

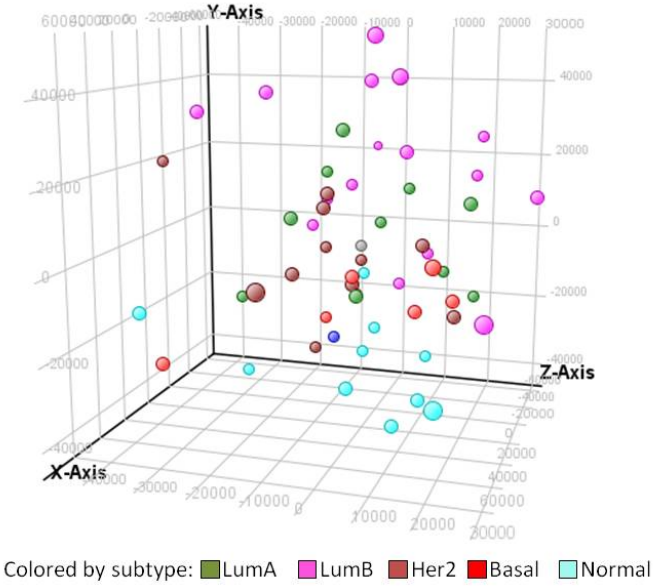

Figure 4.3: **Result of subtype classification.** PCA score plot, where the samples were colored after molecular subclassification using PAM50 indicated that the samples can be distinguished based on subtype.

All tumors, with two exceptions, corresponded well with established characteristics of the "intrinsic" subtypes and the translation to classical immunohistochemistry (IHC). Two samples were not in agreement with data obtained from IHC. Luminal A tumors are predominantly ER+ and basal-like tumors are predominantly ER-, but the opposite was observed for two samples (Figure 4.4).

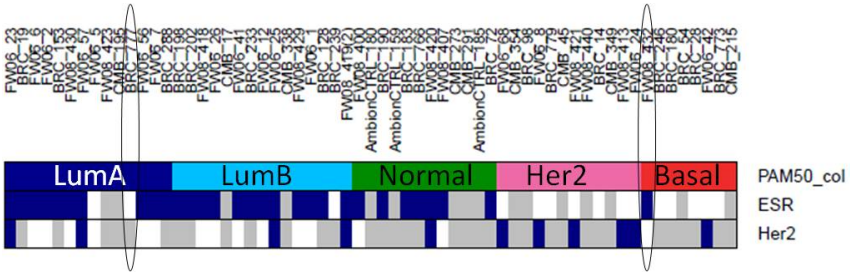

Figure 4.4: **Subclassification of the tumors with corresponding ER and HER2 status**. Blue bars denote positive status, white bars denote negative status and grey bars represents n/a. Two samples (circled) deviated from the typical characteristics of the "intrinsic" subtypes. BRC-777 is ER- and luminal A, while FW08-432 is ER+ and basal-like.

4.3 Differential expression between IDC and DCIS

Differences in gene expression between *in situ* and invasive breast carcinoma were investigated. The whole dataset was used and the samples assigned to 1 and 2, representing DCIS and IDC, respectively. PCA of the dataset could not separate the samples based on patient diagnosis (Figure 4.5).

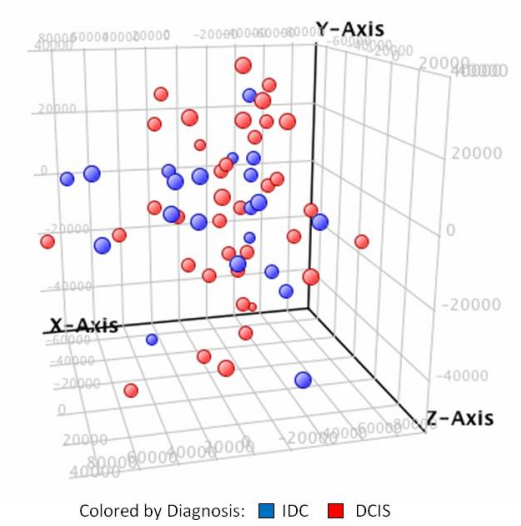

Figure 4.5: **PCA score plot by diagnosis.** Unsupervised PCA plot of the samples colored by diagnosis, IDC (blue) vs. DCIS (red), did not reveal a distinct separation of classes.

Analysis of class separation using SAM (100 permutations) identified 207 and 167 genes differentially expressed between IDC and DCIS (FDR<5%), with use of 75th percentile and quantile normalization, respectively (Figure 4.6).

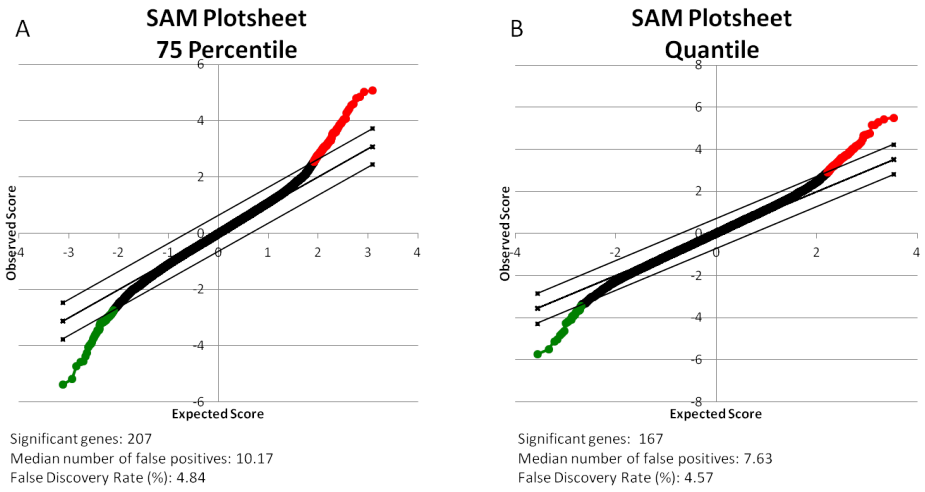

Figure 4.6: **SAM analysis of DCIS and IDC using two different normalization methods.** (A) 75th percentile (B) Quantile normalization. Red color represents upregulated genes in IDC tumors, while green color represents upregulated genes in DCIS tumors.

The number of upregulated genes in IDC was 120 vs. 115 comparing quantile and 75th percentile normalization methods, respectively. Upregulated genes in DCIS were 21 vs. 41, for quantile and 75th percentile normalization method, respectively. Number of overlapping genes for both normalization methods are illustrated in Figure 4.7.

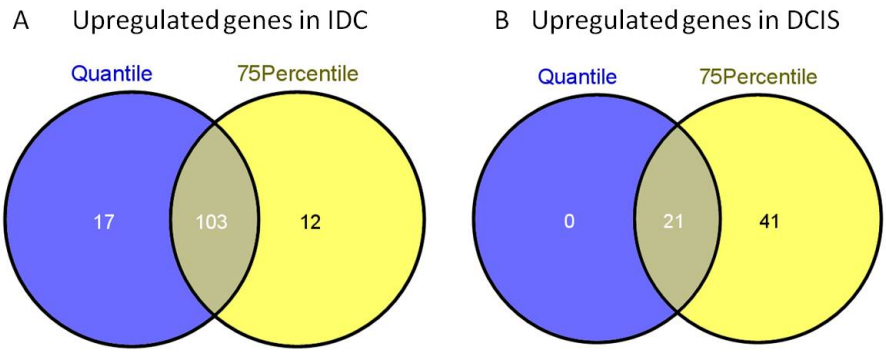

Figure 4.7: **Venn diagrams illustrating overlapping genes derived from SAM using two different normalized datasets.** Comparing SAM performed on the dataset normalized with quantile vs. 75th percentile identified (A) 103 overlapping genes upregulated in IDC tumors and (B) 21 overlapping genes upregulated in DCIS tumors.

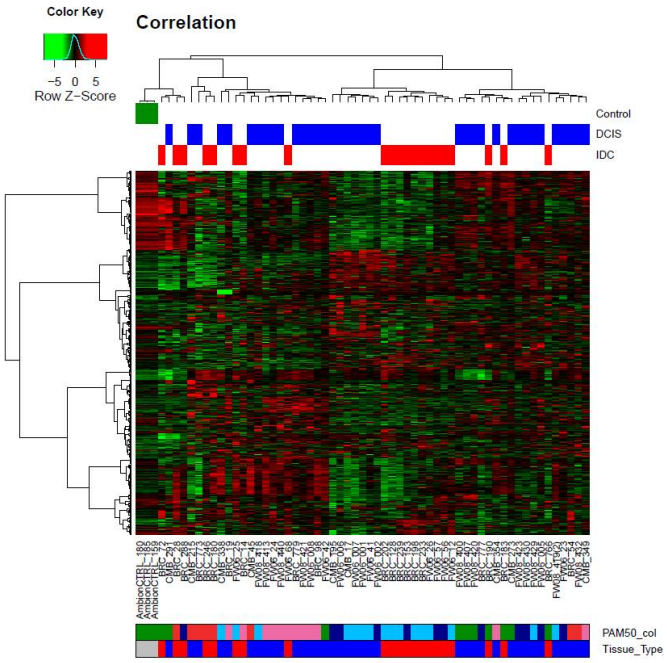

Figure 4.8: **Unsupervised hierarchical clustering of IDC and DCIS across the whole dataset.** No distinct separation of patient diagnosis could be revealed, but rather a tendency of the tumors to cluster by subtype. The columns represent the tumors, while the rows represent the most variable genes. Color coding of diagnosis: Blue=DCIS, red=IDC and green=normal controls. Color coding of subtype: Green= normal-like, pink=HER2+, red=basal-like, dark blue=luminal A and light blue=luminal B. The heat map represents upregulated gene expression in red, downregulated gene expression in green and neutral gene expression in black.

Unsupervised hierarchical clustering using the most variable genes across all samples (quantile normalized) did not separate the patients by diagnosis (Figure 4.8), but rather indicated the clustering to be driven by others factors. A tendency of the tumors to group after subtype as seen in the PCA score plot (Figure 4.3), was confirmed.

After performing SAM, to identify genes differentially expressed between the two diagnosis groups, the genelists from both normalization methods were used to perform supervised hierarchical clustering (Figure 4.9). As expected, the group of DCIS and invasive tumors separated in two clusters, but some DCIS tumors were found to have a gene expression profile more similar to invasive than *in situ* tumors. Especially, a subgroup of eight DCIS tumors revealed upregulation of genes characteristic of the invasive tumors (emphasized in Figure 4.9).

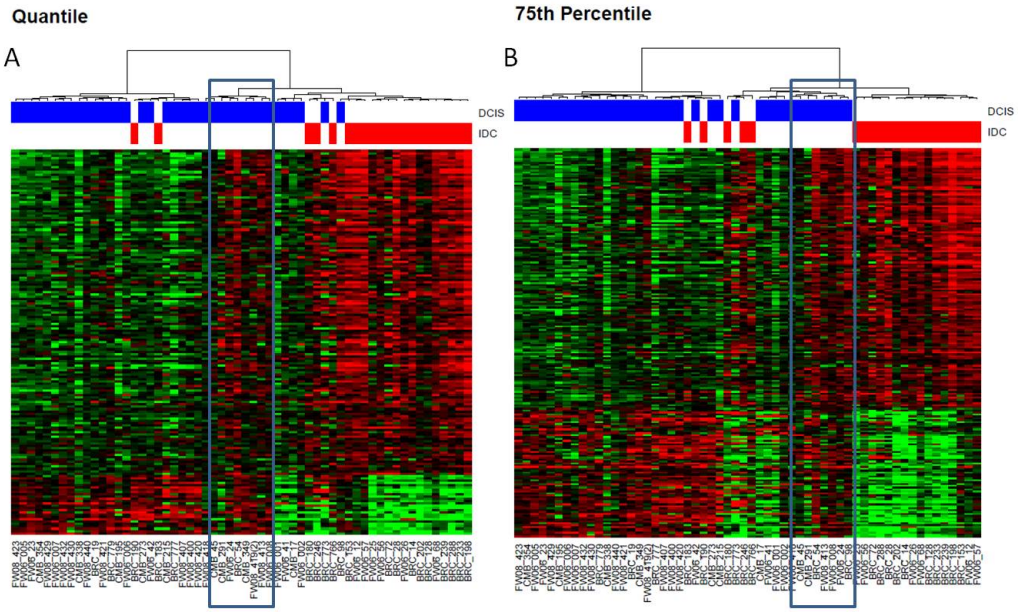

Figure 4.9: **Supervised hierarchical clustering of IDC and DCIS using two different normalized datasets.** (A) Quantile normalized and (B) 75th percentile normalized. Supervised hierarchical clustering of the significant genes derived from SAM separated the patients in two groups. One group almost exclusive for DCIS (blue) and a second group that branched off in two groups; one containing DCIS (blue) and one IDC (red) patients. Eight DCIS tumors revealed upregulation of genes characteristic of the invasive tumors (marked by blue squares).

For the downstream analyzes, the genelists derived from SAM using the 75th percentile normalized dataset were used. Two separate genelists; upregulated genes in IDC tumors and upregulated genes in DCIS tumors were used as input to IPA. The top 20 genes, with highest fold change, able to discriminate between IDC and DCIS tumors are listed in Table 4.1.

Table 4.1: **Top 20 genes of highest fold change that discriminate between DCIS and IDC**

Genes upregulated in DCIS		
Gene name	**Gene Symbol**	**Fold change**
dystonin	DST	0,585
solute carrier family 16, member 9	SLC16A9	0,534
chromosome 12 open reading frame 54	C12orf54	0,517
acyloxyacyl hydrolase (neutrophil)	AOAH	0,502
CD163 molecule-like 1	CD163L1	0,501
BCL2-like 10 (apoptosis facilitator)	BCL2L10	0,482
myosin light chain kinase	MYLK	0,455
discs, large homolog 2 (Drosophila)	DLG2	0,455
adenosine A2b receptor	ADORA2B	0,448
adherens junctions associated protein 1	AJAP1	0,445
Genes upregulated in IDC		
Gene name	**Gene Symbol**	**Fold change**
phosphatidic acid phosphatase type 2 domain 1A	PPAPDC1A	5,738
collagen, type I, alpha 1	COL1A1	5,016
synapse differentiation inducing 1	SYNDIG1	4,539
asporin	ASPN	4,525
leucine rich repeat containing 15	LRRC15	4,277
collagen, type III, alpha 1	COL3A1	4,095
matrix metallopeptidase 11 (stromelysin 3)	MMP11	4,068
fibronectin type III domain containing 1	FNDC1	4,004
collagen, type X, alpha 1	COL10A1	3,992
hypothetical protein LOC651721	LOC651721	3,728

Ten molecular and cellular functions were found to be significantly upregulated in IDC tumors compared with DCIS with a threshold of 0.01 (Figure 4.10). These functions involved cellular assembly and organization, cellular function and maintenance, cell morphology, cellular movement, cell-to-cell signaling and cellular growth and proliferation among others.

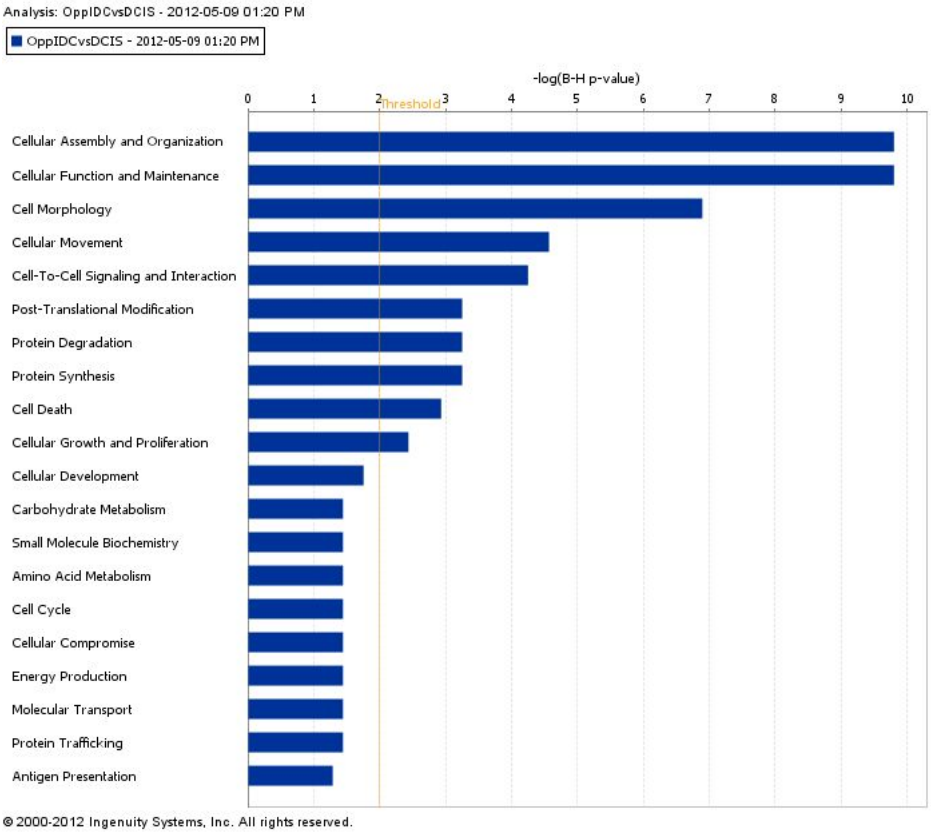

Figure 4.10: **Molecular and cellular functions upregulated in IDC compared to DCIS.** Input to IPA could reveal ten significantly molecular and cellular functions upregulated in IDC tumors compared to DCIS. Adjusted p-value (Benjamini-Hochberg) was set to 0.01.

A selection of the most significant molecular and cellular functions were further explored and could reveal organization of collagen fibrils, shape change of fibroblasts , morphology of connective tissue cells , migration of cells and cell proliferation among others. The functional annotation and the associated genes involved is listed in Table 4.2. Gene ontology categories enriched in IDC tumors as identified from DAVID analysis are listed in Table 4.3. Many of the significant terms involved extracellular matrix, signaling and collagen synthesis.

Table 4.2: **Selected significant functions and the associated genes upregulated in IDC compared with DCIS from IPA**

Category	Functions Annotation	p-Value	Genes
Assembly and Organization	organization of collagen fibrils	3,07E-13	ADAMTS14,COL1A1, LOX
	formation of collagen fibrils	1,24E-08	COL5A1,MMP2,POSTN
	abnormal morphology of collagen fibrils	2,12E-08	COL3A1,COL5A1,COL5A2
	organization of filaments	1,99E-07	ADAMTS14,BGN, FN1
Cell Morphology	shape change of fibroblasts	7,16E-06	FBN1,FN1,SPARC,TSC2
	morphology of fibrils	5,56E-10	COL3A1,COL5A1,COL5A2
	shape change of tumor cell lines	1,53E-03	CDH11,FN1, PITX2,SRPX2
	morphology of connective tissue cells	1,92E-03	POSTN, SPARC,MMP14
Cellular Movement	migration of cells	2,75E-07	MMP11,MMP14,MMP2
	migration of tumor cells	5,27E-04	POSTN,FN1,MMP14,MMP2,
	invasion of cells	1,83E-04	FN1,MMP14,MMP2,POSTN,
Signaling and Interaction	fusion of muscle cells	1,05E-05	ADAM12,MMP14,MMP2,
	attachment of cells	5,74E-05	FN1,LOX,MMP2,POSTN,
	fusion of myoblasts	7,71E-05	ADAM12,MMP14,MMP2
	binding of cell surface	9,28E-04	MMP14,MSR1,SPARC
	binding of connective tissue cells	9,28E-04	BGN,SPARC,THBS2
	adhesion of cell-associated matrix	1,36E-03	COL3A1,COL5A3,FN1,NID2
	adhesion of connective tissue cells	1,71E-03	COL1A1,FN1,MMP2,POSTN
Growth and Proliferation	proliferation of connective tissue cells	3,65E-03	LOX,LUM,WNT2,MMP2

Table 4.3: **GO classes enriched in IDC from DAVID analysis**

Category	**Term**	**Genes**	**Adjusted p-value (Benjamini-Hochberg)**
SP_PIR_KEYWORDS	extracellular matrix	31	6.8E-30
GOTERM_CC_FAT	proteinaceous extracellular matrix	35	1.5E-28
GOTERM_CC_FAT	extracellular matrix	35	9.8E-28
SP_PIR_KEYWORDS	signal	70	2.0E-25
UP_SEQ_FEATURE	signal peptide	70	1.7E-24
SP_PIR_KEYWORDS	Secreted	51	8.9E-23
GOTERM_CC_FAT	extracellular region	59	3.0E-22
GOTERM_CC_FAT	extracellular region part	44	3.9E-22
SP_PIR_KEYWORDS	glycoprotein	73	9.7E-21
GOTERM_CC_FAT	extracellular matrix part	20	2.9E-19
SP_PIR_KEYWORDS	collagen	16	2.2E-16
GOTERM_CC_FAT	collagen	13	2.4E-16
INTERPRO	Collagen triple helix repeat	16	5.1E-16
UP_SEQ_FEATURE	glycosylation site:N-linked	64	6.2E-14
GOTERM_BP_FAT	collagen fibril organization	11	5.8E-13

4.4 Differential expression among DCIS

Differences in gene expression among *in situ* lesions were studied by excluding the invasive breast tumors from the dataset. Unsupervised hierarchical clustering including the most variable genes (SD=1.00, Probes=5207) across the dataset restricted to pure DCIS revealed two distinct clusters. One group contained 13 tumors, all classified as either luminal A or luminal B and with the majority of samples of intermediate grade. The second group contained 24 tumors, representing ten HER2+, four basal-like, seven normal-like and three luminal B tumors. Of these tumors, 16 were high grade, two low or intermediate grade and six unknown (Figure 4.11).

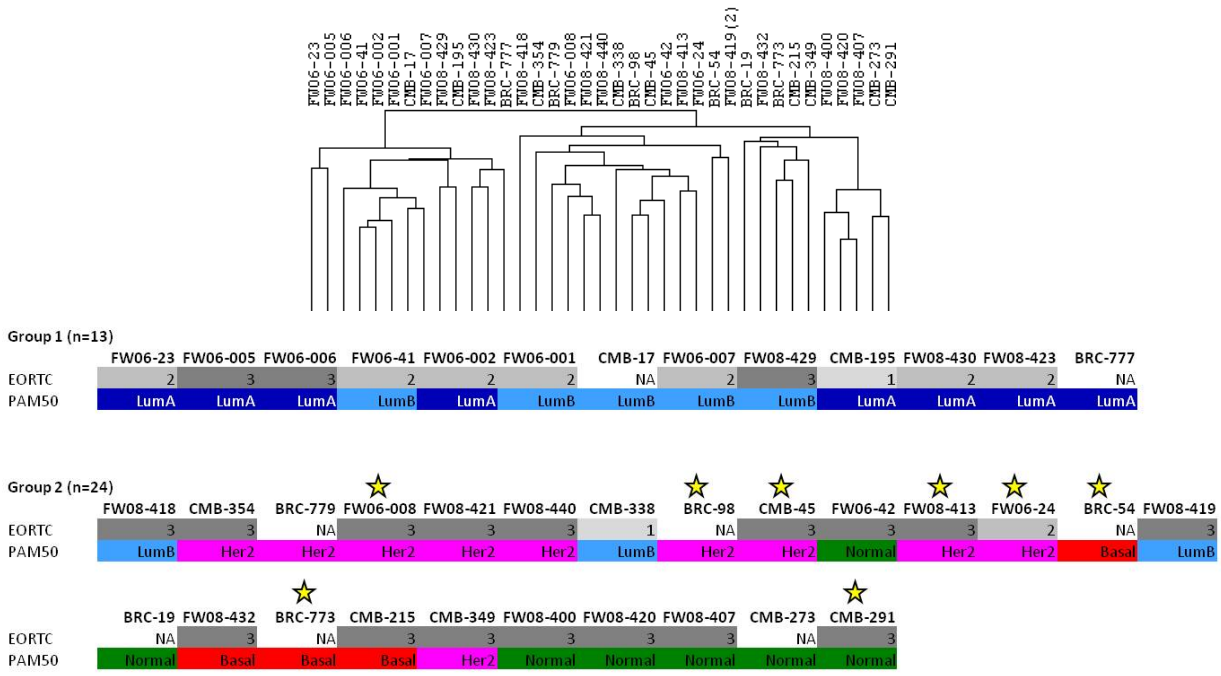

Figure 4.11: **Unsupervised hierarchical clustering of pure DCIS**. Unsupervised hierarchical clustering separated the *in situ* tumors into two groups. Group 1 (n=13) consisted of luminal A and luminal B tumors of intermediate grade, with the exception of three high grade and one low grade sample. Group 2 (n=24) consisted of mostly HER2+, basal-like and normal-like tumors and three luminal B samples. In the second group, all tumors were of high grade, with the exception of two samples, one low grade and one intermediate. Six tumors were of unknown grade. The same subgroup of eight tumors with gene expression features of invasive carcinomas, from Figure 4.9 were found in group 2 of DCIS and are highlighted with stars.

In an unsupervised principal component analysis, where samples were labeled according to cluster group (group 1=red, group 2=blue), a clear separation of the two groups was observed (Figure 4.12). This result supported the finding from unsupervised hierarchical clustering of two groups with distinctly different expression profiles among DCIS tumors.

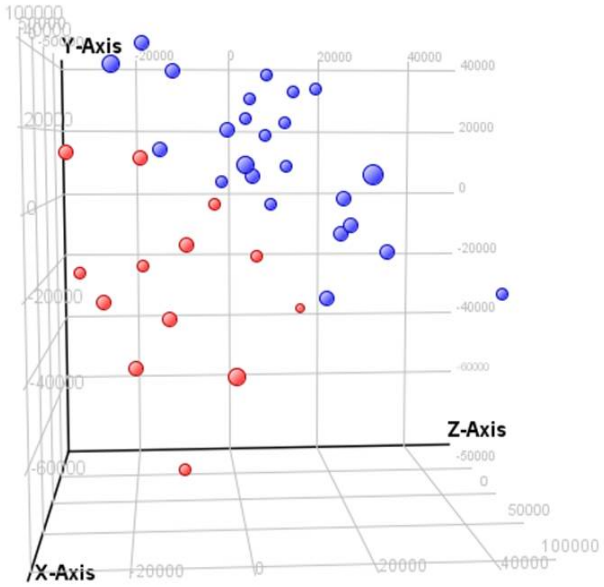

Figure 4.12: **PCA score plot of DCIS tumors.** Unsupervised PCA score plot illustrates the clear separation of DCIS tumors by cluster group: group 1 in red and group 2 in blue.

Two class SAM analysis (500 permutations) identified more than 6000 genes (FDR<5%) differentially expressed between the two groups of DCIS. This result confirmed the clear and significant differences in gene expression between the groups. The most significant genes (FDR≈0%) upregulated in each group, group 1 (562 genes) and group 2 (388 genes) were ranked by fold change and fed into IPA. Twenty of the genes with highest fold change and able to discriminate between the two groups of DCIS are listed in Table 4.4. Three calcium binding proteins of the S100A family (S100A7, S100A8, S100A9) outclassed the rest with fold changes of 20, 17 and 14 respectively. Two chemokines, CXCL1 and CXCL2 were also found to be significantly upregulated, with fold changes of 9 and 6, respectively.

Table 4.4: **Top 20 genes of highest fold change that discriminate between the two groups of DCIS**

DCIS group 1

Gene name	Gene Symbol	Fold change
THAP domain containing, apoptosis associated protein 3	THAP3	0,650
chromosome 12 open reading frame 10	C12orf10	0,644
chromosome 14 open reading frame 93	C14orf93	0,638
methyltransferase 5 domain containing 1	METTL15	0,634
cyclin M3	CNNM3	0,626
tumor suppressor candidate 2	TUSC2	0,623
aftiphilin	AFTPH	0,623
leucine zipper and CTNNBIP1 domain containing	LZIC	0,622
CSRP2 binding protein	CSRP2BP	0,621
HIRA interacting protein 3	HIRIP3	0,617

DCIS group 2

Gene name	Gene Symbol	Fold change
S100 calcium binding protein A7	S100A7	20,6
S100 calcium binding protein A8	S100A8	17,2
S100 calcium binding protein A9	S100A9	14,1
LOC645638	LOC645638	10,6
chemokine (C-X-C motif) ligand 1	CXCL1	9,7
lipocalin 2	LCN2	7,4
Kruppel-like factor 5	KLF5	6,9
phospholipase A2, group IIA	PLA2G2A	6,7
C-type lectin domain family 4	CLEC4G	6,4
chemokine (C-X-C motif) ligand 2	CXCL2	6,3

Genes highly expressed in DCIS group 2 were found to be involved in diverse cellular functions related to the immune system. A selection of the most significant molecular and cellular functions and the associated genes upregulated in DCIS group 2 compared with DCIS group 1, are listed in table 4.5. Different surface antigens expressed on leukocytes were identified: CD28 (naive T cells), CD209 (dendritic cells and macrophages), CD40 (macrophages), CD97 (all leukocytes) and CD99 (all leukocytes). Several receptors of various cytokines, including interleukin 1, 2, 3, 4, 7, 18 and 22 as well as the cytokines CD70 and IL-16 were found to be highly expressed. Furthermore, IPA identified pathways connected to the immune system that were highly enriched in DCIS group 2: T helper cell differentiation, primary immunodeficiency signaling and crosstalk between dendritic cells and natural killer cells (Figure 4.13).

Table 4.5: **Selected significant functions and the associated genes upregulated in DCIS group 2 compared with group 1 from IPA**

Functions Annotation	p-Value	Genes
Function and Maintenance		
T cell development	2,78E-10	ADA,CD28,CD70
function of blood cells	7,70E-10	CD209,CD28,CD40
function of leukocytes	1,04E-09	CD209,CD28,CD40
differentiation of T lymphocytes	2,37E-08	CD28,CD70,CIITA,M1,PTPN22
function of T lymphocytes	1,93E-07	CD28,IL16,IL18R1,IL4R
function of phagocytes	1,26E-05	CD209,CD40,IL1R1,IL2RG,IL4R,JAK3
function of antigen presenting cells	4,42E-05	CD209,CD40,IL2RG,IL4R
cellular homeostasis	8,19E-05	CCL14,CD28,CD40,CD70,CXCL1
function of dendritic cells	2,55E-04	ADA,BATF3,CD209,CD40,CIITA,
Growth and Proliferation		
proliferation of blood cells	1,92E-12	CD209,CD28,CD40,CD70,CD99 IL1R1,IL2RG,IL4R,IL7R,JAK3
proliferation of lymphocytes	2,15E-12	CD209,CD28,CD40,IL1R1, CD70,CD99,CDK6,IL2RG,IL4R
proliferation of T lymphocytes	9,86E-12	CD209,CD28,CD40,CD70,
Cellular Development		
differentiation of antigen presenting cells	3,54E-03	BATF3,BMP2,CD40
developmental process of helper T lymphocytes	3,93E-03	CD28,IL18R1,IL1R1,IL4R,JAK3
Cell Death		
cell death of immune cells	7,60E-09	CD28,CXCL2,IL2RG,IL3RA,IL7R,
apoptosis of leukocytes	5,83E-07	CD28,CD40,CD70,CXCL1,CXCL2
cell death of lymphocytes	1,48E-06	CD28,CD40,CD70,
cell death of T lymphocytes	3,26E-06	CD28,CD40,CD99,IL2RG,IL7R,
survival of T lymphocytes	1,36E-05	CD28,CD40,FAS,IL2RG,IL7R,JAK3,
Cell Morphology		
morphology of lymphocytes	9,55E-07	BACH2,BCL11A,BCL11B,CD28
morphology of leukocytes	1,41E-06	CD28,IL1R1,IL2RG,IL4R,IL7R
morphology of blood cells	2,35E-05	IL1R1,IL2RG,IL4R,IL7R,
Cellular Movement		
migration of cells	2,50E-11	CCL14,CD209,CD28,CD40,CD97,CD99,CXCL1
leukocyte migration	2,36E-10	CCL14,CD209,CD28,CD40,CD99, CXCL1,CXCL2
cell movement	1,27E-09	CCDC88A,CCL14,CD209,CD28, CD40,CD97,CD99,CXCL1,CXCL2
cell movement of leukocytes	1,89E-08	CCL14,CD209,CD28,CD40,CD99, CXCL1,CXCL2,IL16,IL18BP
chemotaxis of cells	1,27E-07	CCDC88A,CCL14,CD28,CD40,CXCL1,CXCL2, IL16,IL1R1,LIF,LOX,LSP1,NCKAP1L

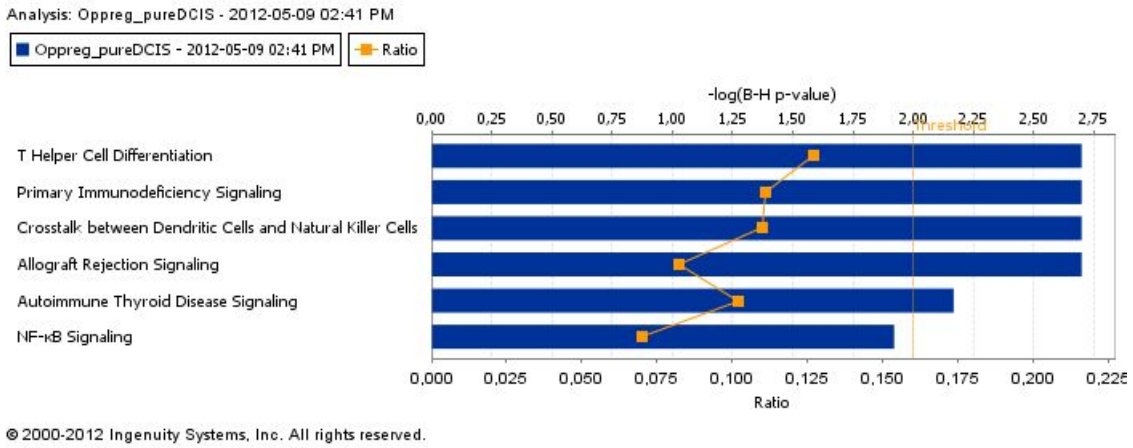

Figure 4.13: **Significant pathways enriched in DCIS group 2 from IPA.** Genes upregulated in DCIS group 2 belong to pathways related to the immune system: T helper cell differentiation, primary immunodeficiency signaling and crosstalk between dendritic cells and natural killer cells.

By subjecting the same genelists to DAVID, GO classes including immune response, defense response and regulation of lymphocyte activation among others were identified in DCIS group 2 (Table 4.13). These results confirm that the difference between the two groups of DCIS is partly explained by the role of the immune system.

Table 4.6: **GO classes enriched in Group 2 of DCIS from DAVID analysis**

Category	Term	Genes	Adjusted p-value (Benjamini-Hochberg)
GOTERM_BP_FAT	immune response	49	5.3E-11
SP_PIR_KEYWORDS	signal	102	9.1E-7
UP_SEQ_FEATURE	signal peptide	102	4.2E-6
SP_PIR_KEYWORDS	disulfide bond	94	8.1E-7
SP_PIR_KEYWORDS	glycoprotein	124	5.5E-7
UP_SEQ_FEATURE	disulfide bond	92	2.7E-6
UP_SEQ_FEATURE	glycosylation site:N-linked	116	4.0E-5
KEGG_PATHWAY	Cytokine-cytokine receptor interaction	24	1.7E-5
GOTERM_BP_FAT	defense response	34	2.4E-4
GOTERM_CC_FAT	extracellular region	68	8.6E-5
SP_PIR_KEYWORDS	Secreted	59	5.8E-5
GOTERM_BP_FAT	regulation of lymphocyte activation	15	1.0E-3
GOTERM_BP_FAT	regulation of cell activation	16	1.2E-3
SP_PIR_KEYWORDS	immune response	17	1.8E-4
GOTERM_BP_FAT	lymphocyte activation	17	9.9E-4

Chapter 5

Discussion

5.1 Methodological considerations

5.1.1 Patient material

This study included fresh frozen tumor tissue obtained from three different hospitals with separate routines of tissue collection and preservation. Ideally, all tumors should have been collected at the same hospital to avoid variation between such protocols. DCIS tissue material is often scarce and some variation due to differences in origin must be accepted in order to obtain a reasonable number of samples for analysis. A relatively high number of DCIS (n=38) was obtained. However, the number of DCIS cases was not sufficient to validate the findings in this study. By EORTC grading, they were assigned to the following goups: low grade (n=2), intermediate grade (n=8) and high grade (19). Unfortunately, in 8 of the cases grading was not available.

To compare expression profiles of DCIS with those of invasive breast cancer, 24 cases of small invasive breast tumors were studied. Five of these tumors were histological grade 1, 10 tumors were grade 2 and 6 tumors were grade 3. Usually, in previous studies comparing *in situ* and invasive breast cancer, IDC cases have been in excess. This uneven distribution could impact the results and should be considered when comparing these data with previous studies.

5.1.2 RNA isolation and integrity

RNA from each sample was randomly isolated based on diagnosis and origin to balance potential differences of day-to-day variations. Due to variation in quantity of tissue, small adjustments had to be made during the isolation process to maximize yield. These

adjustments involved additional homogenization and washing steps, longer incubation and precipitation times and addition of carrier. Larger quantities of tissue (>10 mg) were better homogenized than smaller quantities resulting in higher yield and purity. RNA was isolated from each sample with some variations concerning purity and yield, which is provided as supplementary information in Appendix C.

Three samples with high content of fat and five samples of small tissue quantity were re-isolated using a modified method of the TRIzol extraction protocol, followed by RNeasy mini columns to obtain optimal RNA yield and purity. Ideally, the same protocol should have been used for all samples, but the small adjustments were considered negligible when comparing biological differences.

Precautions were taken to prevent RNase contamination during all steps of the procedure by wearing gloves, using sterile equipment and proper technique described by protocol. Still, the integrity of RNA was found to range from very low to high, as measured with Agilent 2100 Bioanalyzer. Whether the degradation occurred before or during RNA isolation can not be determined. Samples from same isolation batch were found to differ in RIN, suspecting the degradation to have happened prior to isolation.

The RNA integrity number varied from 1.6 to 8.6. No cutoff was set for samples with low RIN, however four samples with low RIN failed to be amplified and were not used for microarray profiling. The remaining samples with low RIN were included, and used for hybridization as long as the remaining quality criteria were met. How variations in integrity of RNA will impact the biological profile of the samples are not known. PCA score plot of all samples revealed a grouping of samples with low RIN. This grouping was less clear after filtering the probes based on low expression values and flags, indicating some of the common features of these samples to relate to low signal intensities. Even though PCA could reveal assembling of samples with low RIN, the same samples did not cluster in an unsupervised hierarchical clustering analysis, providing evidence that these samples express different gene expression patterns explained by other factors than RNA integrity.

5.1.3 Preprocessing of microarray data and the effects of normalization

To reliably compare microarray data, non-biological differences due to technical variation need to be minimized. All experiments were performed within the same laboratory facility using the same protocol for all samples. Multiple production batches of reagents and arrays as well as day-to-day variation were inevitable.

Two methods of normalization,75th percentile and quantile normalization, were used to

adjust for systematic errors in the data making comparisons across multiple samples feasible. Both methods led to the same results by PCA, both before and after filtering the probeset based on expression values and flags. After SAM analysis between IDC and DCIS, more genes were found to be significantly differentially expressed in the dataset which had been normalized with 75th percentile (207 vs. 167). This indicates that quantile normalization is more strict and excludes more variance than the 75th percentile method. To allow more variance, 75th percentile normalization was used for downstream analysis to ensure inclusion of more biological differences.

PCA allowed batch differences to be identified in the data after normalization (both 75th percentile and quantile). However, after filtering the data on low expression values and flags, the batch effects disappeared, highlighting the importance of minimizing technological differences prior to investigation of biological differences to reveal reliable patterns of expression.

5.2 Biological considerations

5.2.1 Tumor heterogeneity

Breast cancer is a complex and heterogeneous disease. DCIS is considered the precursor lesion of invasive breast cancer, but the progression might occur through multiple pathways, due to diverse histology, genomic alterations and variation in gene expression found both between patients (intertumor heterogeneity) and within cancer cells of the same patient (intratumor heterogeneity). Because of the intratumor heterogeneity, an individual tumor-tissue sample may not be representative of the whole tumor and its metastatic potential [58]. In this study, the aim was to compare groups of tumors from different patients assuming the global gene expression profile to be representative for each tumor.

Molecular subtyping of breast tumors based on gene expression profiles combined with histopathological features have proven valuable in explaining the heterogeneity of breast tumors. Each of the "intrinsic" subtypes have been found to differ in progression and therapeutic response [59]. All five "intrinsic" subtypes were found present in both DCIS and IDC tumors and the progression of *in situ* to invasive cancer is likely to occur by distinct mechanisms dependent on subtype.

5.2.2 Differential expression between IDC and DCIS

A previous gene expression profiling study by Ma *et al.* [41] reported that the most noticeable transcriptional changes occur at the transition from normal breast epithelium

to atypical ductal hyperplasia and that the alterations are conserved throughout the later stages of progression through DCIS and IDC. Porter *et al.* [60] also found the most dramatic transcriptional change to occur at the normal to *in situ* transition. However, other studies using gene expression profiling have demonstrated important differences between *in situ* and invasive tumors (Hannemann *et al.* [61], Schuetz *et al.* [62], and Muggerud *et al.* [40]). This study could confirm important differences between *in situ* lesions and invasive carcinomas based on their expression profiles.

Unsupervised hierarchical clustering using the most variable genes across all samples did not separate the patients by diagnosis, but revealed a trend of the tumors to cluster by subtype. This suggests that the global expressions patterns contain features that correlate stronger to molecular subtype than diagnosis, and that the diversity exists at both the preinvasive and invasive stage. The clustering could not be related to common markers such as grade or ER status, indicating the potential to identify markers not yet discovered.

5.2.2.1 Genes upregulated in invasive tumors point to stromal interactions and invasion

Approximately 200 genes were found to discriminate IDC from DCIS tumors. Genes highly expressed in the invasive tumors included those genes related to organization of collagen fibrils (MMP2, POSTN, ADAMTS14), shape change of fibroblasts (FN1, SPARC), morphology of connective tissue cells (MMP14, POSTN, SPARC), migration of cells (ADAM12, MMP11, MMP14, MMP2, SPARC, FN1, POSTN) and cell proliferation (BGN, SPARC, LOX, MMP2, VCAN, MMPs) among others. MMP11, MMP2, MMP14 are proteins of the matrix metalloproteinase family, involved in the breakdown of extracellular matrix and hence an indicator of invasion. SPARC encodes a cysteine-rich acidic matrix-associated protein involved in extracellular matrix synthesis, promotion of cell shape change and tumor cell invasion. FN1 encodes fibronectin which is involved in cell adhesion and migration processes like metastasis and POSTN is known to induce cell attachment and spreading [54].

The fact that genes discriminating between *in situ* and invasive tumors are encoding extracellular matrix (ECM)-related proteins is expected since the most prominent difference between *in situ* and invasive cancer is the degradation of the basement membrane and the interaction with stromal components. Hence, expression of genes associated with ECM indicates the breakage of the basement membrane and tumor-stromal interactions.

Some genes highly expressed in the invasive tumors were found to be involved in the epithelial-to-mesenchymal transition (EMT). These genes involved collagen, type VI, alpha 1 (COL6A1) and 2 (COL6A2), cathepsin B (CTSB) and matrix metallopeptidase 2

(MMP2). COL6A1 and COL6A2 have been proposed as mesenchymal markers by Jechlinger *et al.* [24]. The expression of EMT-related genes correlate with invasive behavior and is thus a second confirmation of a broken basement membrane.

BCL2 was found upregulated in DCIS. This anti-apoptotic protein has been associated with a poor prognosis of breast cancer if downregulated. Its upregulation in DCIS correspond to a downregulation in IDC, which is consistent with a poorer prognosis for IDC patients and previous findings from Park *et al.* [63].

5.2.2.2 DCIS with upregulated genes characteristic of invasive tumors

Interestingly, eight of the DCIS tumors were found to have an expression profile more similar to invasive tumors. These eight tumors represented three subtypes (five HER2+, two basal-like and one normal-like) and were mostly high-grade (one intermediate grade, three unknown). The subgroup showed elevated expression levels of ECM-related genes, indicating some tumor-stromal signaling to occur prior to degradation of the basement membrane in the *in situ* lesions. If this signaling is only present in the most aggressive preinvasive lesions, these expression features could have the potential of predicting which of the DCIS lesions that will transit to IDC.

Other high-grade DCIS clustered with low and intermediate grade DCIS and expressed low levels of MMPs and ECM-related genes. This result demonstrates differences between DCIS lesions of the same histopathological grade also found by Muggerud *et al* [40]. Precautions should be taken when assuming all high grade lesions to develop into invasive tumors. These results indicate others factors beyond grade to play a role in the process of invasion.

5.2.3 Differential expression among DCIS

To explore differences within the DCIS cases, the IDC tumors were excluded and unsupervised hierarchical clustering performed on the dataset exclusive for DCIS. The tumors separated into two groups: Group 1 consisted of luminal A and luminal B subtypes only, and were predominantly of intermediate and low grade. Group 2 consisted of HER2+, basal-like and normal-like subtypes predominantly, and were almost exclusively of high grade. The same subgroup of eight tumors with gene expression features of invasive carcinomas were found in group 2 of DCIS.

Previous studies have demonstrated how ER+ and ER- tumors display distinctly different gene expression profiles, indicating a major role of the estrogen receptor in breast cancer [64]. However, in this study both expression based groups of DCIS were found to contain

ER+ tumors, indicating other underlying biological properties separating the two groups of DCIS independent of ER status.

Greater diversity was observed among *in situ* tumors than between *in situ* and invasive tumors. By SAM, more than 6000 genes discriminated the two groups of DCIS, compared to the 207 genes which discriminated IDC and DCIS. The differences in the expression profiles among DCIS seemed to be driven by other molecular mechanisms than those involved in progression from *in situ* to invasive breast cancer, also confirmed in a previous study on gene expression profiling of DCIS performed by Hannemann *et al.* [61]. Important differences between *in situ* lesions and invasive carcinomas involved high expression of ECM-related genes, whereas the genes discriminating the two groups of DCIS were related to immune response. As mentioned, previous results have reported less dramatic transcriptional changes to occur from the *in situ* to invasive transition, highlighting the importance of investigating differences within DCIS and not solely compare *in situ* tumors against invasive tumors.

5.2.3.1 Invasive-like subgroup of DCIS

Some genes highly expressed in DCIS group 2 were related to invasion and metastasis: S100A7, MMP23B and SNAI1 (involved in EMT). These genes were also found to be highly expressed in a high grade DCIS invasive-like subgroup identified by Muggerud *et al.* [40]. Clearly, some DCIS lesions exhibit invasive properties that may be detected prior to invasion.

Other consistent results from this study with Muggerud *et al.* were the genes S100A8 and CXCL1. Both genes were found to be significantly upregulated with a fold change of 17 and 9, respectively and related to the immune system. S100A8 encodes a calcium-binding protein which promotes phagocyte migration and infiltration of granulocytes at sites of wounding. It has also been found to play a role in acute and chronic inflammations and inducing the release of IL8. CXCL1 encodes a member of the CXC subfamily of chemokines and has chemotactic activity for neutrophils and may play a role in inflammation [54]. The closely related chemokine CXCL2 was also found to be significantly upregulated and it is known to be produced by activated monocytes and neutrophils and expressed at sites of inflammation [54]. Signs of inflammation have been associated with poor prognosis in breast cancer [65] and indicate the tumors in the DCIS group 2 to be of a more aggressive type.

High expression of receptors of various cytokines, including interleukin 1, 2, 3, 4, 7, 18 and 22 as well as the cytokines CD70 and IL-16 were found in DCIS group 2. Cytokines play important roles in cell-to-cell communication in the immune system and different amounts

of cytokine signaling have been found to affect the role of the immune response in cancer cells [65]. In agreement with this, a strong immune response and interleukin signaling was identified in DCIS cases explored by Kristensen *et al.* [66] and involved signaling of interleukin 2, 4, 6, 12 and 23. IL-4 is a cytokine that stimulates B-cell differentiation and chronic inflammation in cancer cells. It has also shown to promote mast cell activation and tumor growth [65]. This result indicates the presence of cytokine signaling in *in situ* tumors before the process of invasion and might be a second indicator of the tumors in group 2 of DCIS to be at the edge of invasion.

The tumors within group 2 were mostly of high grade (16 tumors of grade 3, one tumor of grade 1, one tumor of grade 2 and six unknown), revealing some correlation between grade and invasion. However, three high grade DCIS lesions were also present in DCIS group 1, emphasizing the need for more investigation into features beyond histological grade and ER status.

5.2.3.2 Upregulated immune response in DCIS group 2

A majority of the genes discriminating the two groups of DCIS were related to immune response. DCIS from group 2, with invasive characteristics, seemed to have high expression of genes related to immune response. Genes encoding surface markers of both the innate (macrophages, dendritic and natural killer cells) and the adaptive (T and B cells) immune system were identified. The immune system have been found to both suppress and promote tumorigenesis. In approximately 50% of all breast cancer patients, antibodies against tumor-specific antigens are active [67]. Usually, a high infiltration of macrophages, mast cells and neutrophils correlates with increased angiogenesis and tissue remodeling by the production of growth factors, cytokines, chemokines and matrix metalloproteinases, while an abundance of infiltrating lymphocytes indicates a cytotoxic attack from T and B cells [65].

Presence of cell types of both the innate and the adaptive immune system, suggests the role of the immune response to be both protumorigenic with an inflammatory response and antitumorigenic with infiltration of lymphocytes. Whether the tumor promoting or tumor suppressing role is more dominant remain to be identified and follow-up studies are needed to investigate the paradoxical role of innate and adaptive leukocytes. In the future, the tumor promoting properties of the immune cells as well as the antitumor response could potentially help distinguishing between the most aggressive forms of DCIS and those of a more benign phenotype.

The upregulated immune response was found in HER2+, luminal B, normal-like and basal-like subtypes, indicating a difference in immune response related to subtype. Especially, luminal A tumors seemed to separate from the remaining subtypes, with lower

expression of ECM-related genes and immune signaling. An upregulation of the immune response in the HER2+ subgroup have previously been reported by Kristensen *et al.* [66]. Macrophage infiltration was found to distinguish basal-like and HER2+ subtypes from the other subtypes. This upregulation was found at both the *in situ* and invasive stage, indicating the immune system to play a role in the early development of the disease.

Loo *et al.* [68] discovered a high fraction of non-tumor cells in the HER2+ and normal-like subtypes. One explanation for the upregulated immune response in HER2+ subtypes could be the function of HER2 as an antigen and hence the attraction of T-cells. However, presence of specific antibodies and T-cells is not consistent with being HER2+, but HER2 is investigated as a promising molecule for immunotherapy by generating specific T and B cell responses with cytotoxic potential [69].

The role of HER2 in progression of breast cancer is unclear. In DCIS the HER2 overexpression seem to be prevalent and reported to be up to 50% [70]. 10 of 37 DCIS cases in this study were classified as HER2+ and found to belong to group 2 of DCIS with an upregulated immune response. In addition, five of these HER2+ tumors co-clustered with the invasive tumors and were considered potentially more aggressive. Park *et al.* [70] suggested that biological and genetic alterations other than HER2 amplification played a critical role in the progression of DCIS to IDC. Further investigation on the role of HER2 amplification in DCIS might help to better understand these tumors in terms of progression.

Chapter 6

Conclusion and future perspectives

This study demonstrates that gene expression profiling can distinguish *in situ* and invasive breast cancer as well as reveal differences within cases of DCIS. Important differences between *in situ* lesions and invasive carcinomas involved high expression of ECM-related genes and stromal signaling, indicating an active tumor-stromal interaction in IDC tumors.

There appears to be a group of DCIS with a gene expression profile of invasive characteristics. The profile seemed to correlate with subtype and grade to some extent. HER2+, basal-like and normal-like subtypes of DCIS of high grade showed elevated levels of ECM-related genes and a more active stroma, whereas luminal A tumors were found to separate from the above with lower expression of such genes. Tumor-stromal signaling could therefore be detected in DCIS lesions prior to invasion, and should be explored as potential markers, even though no distinct biomarker was identified in this study. Breast tumors are known to behave differently based on subtype and in the future, focus could be set on progression markers related to subtype instead of patient diagnosis.

Greater diversity was found in the gene expression profiles within DCIS tumors, than between *in situ* and invasive carcinomas. Over 6000 genes were found to discriminate two groups of DCIS and involved high expression of genes related to immune response. Also here, the expression profiles seemed to correlate with subtype to some extent. The elevated levels of immune signaling in DCIS group 2 were found in HER2+, basal-like, normal-like and luminal B subtypes, but were completely absent in luminal A tumors. The suppressing role of the immune system compared with the promoting role needs to be further investigated, and could potentially increase our knowledge concerning the progression of *in situ* lesions to invasive breast cancer.

Bibliography

[1] Cancer Registry of Norway 2009 (web): Breast Cancer facts, http://www.kreftregisteret.no/no/Generelt/ Fakta-om-kreft-test/Brystkreft/ [Accessed 07.03.2012].

[2] Oncolex 2011 (web): Breast Cancer facts, http://www.oncolex.no/en/Bryst/ [Accessed 09.03.2012].

[3] G. Dimri, H. Band, V. Band, et al. Mammary epithelial cell transformation: insights from cell culture and mouse models. *Breast Cancer Res*, 7(4):171–9, 2005.

[4] D.C. Sgroi. Preinvasive breast cancer. *Annual Review of Pathological Mechanical Disease*, 5:193–221, 2010.

[5] L.C. Collins, R.M. Tamimi, H.J. Baer, J.L. Connolly, G.A. Colditz, and S.J. Schnitt. Outcome of patients with ductal carcinoma in situ untreated after diagnostic biopsy. *Cancer*, 103(9):1778–1784, 2005.

[6] UpToDate 2012 (web): Breast Cancer Development, http://www.uptodate.com/contents/image.do? imageKey=PI02.03.2012].

[7] T. Vargo-Gogola and J.M. Rosen. Modelling breast cancer: one size does not fit all. *Nature Reviews Cancer*, 7(9):659–672, 2007.

[8] M. Ignatiadis and C. Sotiriou. Understanding the molecular basis of histologic grade. *Pathobiology*, 75(2):104–111, 2008.

[9] Norwegian Breast Cancer Group 2011 (web):Histopatologisk diagnostikk, http://www.nbcg.no/nbcg.blaaboka.html [Accessed 09.03.2012].

[10] S.E. Singletary and J.L. Connolly. Breast cancer staging: working with the sixth edition of the ajcc cancer staging manual. *CA: a cancer journal for clinicians*, 56(1):37–47, 2006.

[11] L.K. Dunnwald, M.A. Rossing, C.I. Li, et al. Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. *Breast Cancer Res*, 9(1):R6, 2007.

[12] D.J. Slamon, W. Godolphin, L.A. Jones, J.A. Holt, S.G. Wong, D.E. Keith, W.J. Levin, S.G. Stuart, J. Udove, A. Ullrich, et al. Studies of the her-2/neu proto-oncogene in human breast and ovarian cancer. *Science*, 244(4905):707–712, 1989.

[13] D.J. Slamon, B. Leyland-Jones, S. Shak, H. Fuchs, V. Paton, A. Bajamonde, T. Fleming, W. Eiermann, J. Wolter, M. Pegram, et al. Use of chemotherapy plus a monoclonal antibody against her2 for metastatic breast cancer that overexpresses her2. *New England Journal of Medicine*, 344(11):783–792, 2001.

[14] Perou C M, Sorlie T, Eisen M B, van de Rijn M, Jeffrey S S, Rees C A, Pollack J R, Ross D T, Johnsen H, Akslen L A, Fluge O, Pergamenschikov A, Williams C, Zhu S X, Lonning P E, Borresen-Dale A L, Brown P O, and Botstein D. Molecular portraits of human breast tumours. *Nature*, 406(6797):747–52.

[15] Therese Sorlie, Charles M. Perou, Robert Tibshirani, Turid Aas, Stephanie Geisler, Hilde Johnsen, Trevor Hastie, Michael B. Eisen, Matt Van de Rijn, Stefanie S. Jeffrey, Thor Thorsen, Hanne Quist, John C. Matese, Patrick O. Brown, David Botstein, Per Eystein Lonning, and Anne-Lise Borresen-Dale. 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*, 98(19):10869–10874, 2001.

[16] R. Sandhu, J.S. Parker, W.D. Jones, C.A. Livasy, and W.B. Coleman. Microarray-based gene expression profiling for molecular classification of breast cancer and identification of new targets for therapy. *Lab Medicine*, 41(6):364–372, 2010.

[17] J.S. Parker, M. Mullins, M.C.U. Cheang, S. Leung, D. Voduc, T. Vickery, S. Davies, C. Fauron, X. He, Z. Hu, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology*, 27(8):1160–1167, 2009.

[18] T. Sorlie, R. Tibshirani, J. Parker, T. Hastie, JS Marron, A. Nobel, S. Deng, H. Johnsen, R. Pesich, S. Geisler, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proceedings of the National Academy of Sciences*, 100(14):8418, 2003.

[19] E. Espinosa, J.Á.F. Vara, I.S. Navarro, A. Gámez-Pozo, Á. Pinto, P. Zamora, and A. Redondo. Gene profiling in breast cancer: Time to move forward. *Cancer treatment reviews*, 2011.

[20] C. Sotiriou and L. Pusztai. Gene-expression signatures in breast cancer. *New England Journal of Medicine*, 360(8):790–800, 2009.

[21] D. Hanahan and R.A. Weinberg. Hallmarks of cancer: the next generation. *Cell*, 144(5):646–674, 2011.

[22] R.D. Leek and A.L. Harris. Tumor-associated macrophages in breast cancer. *Journal of mammary gland biology and neoplasia*, 7(2):177–189, 2002.

[23] M. Hu, J. Yao, D.K. Carroll, S. Weremowicz, H. Chen, D. Carrasco, A. Richardson, S. Violette, T. Nikolskaya, Y. Nikolsky, et al. Regulation of in situ to invasive breast carcinoma transition. *Cancer Cell*, 13(5):394–406, 2008.

[24] M. Jechlinger, S. Grunert, I.H. Tamir, E. Janda, S. Lüdemann, T. Waerner, P. Seither, A. Weith, H. Beug, and N. Kraut. Expression profiling of epithelial plasticity in tumor progression. *Oncogene*, 22(46):7155–7169, 2003.

[25] M. Allinen, R. Beroukhim, L. Cai, C. Brennan, J. Lahti-Domenici, H. Huang, D. Porter, M. Hu, L. Chin, A. Richardson, et al. Molecular characterization of the tumor microenvironment in breast cancer. *Cancer cell*, 6(1):17–32, 2004.

[26] C. Lengauer, K.W. Kinzler, B. Vogelstein, et al. Genetic instabilities in human cancers. *Nature*, 396(6712):643–643, 1998.

[27] A. Langerod, H. Zhao, O. Borgan, J.M. Nesland, IR Bukholm, T. Ikdahl, R. Karesen, AL Borresen-Dale, and S.S. Jeffrey. Tp53 mutation status and gene expression profiles are powerful prognostic markers of breast cancer. *Breast Cancer Res*, 9(3):R30, 2007.

[28] K. Yoshida and Y. Miki. Role of brca1 and brca2 as regulators of dna repair, transcription, and cell cycle in response to dna damage. *Cancer science*, 95(11):866–871, 2004.

[29] G. Egger, G. Liang, A. Aparicio, and P.A. Jones. Epigenetics in human disease and prospects for epigenetic therapy. *Nature*, 429(6990):457–463, 2004.

[30] M.D. Lagios. Ductal carcinoma in situ: controversies in diagnosis, biology, and treatment. *The Breast Journal*, 1(2):68–78, 1995.

[31] M.J. Silverstein, M.D. Lagios, P.H. Craig, J.R. Waisman, B.S. Lewinsky, W.J. Colburn, and D.N. Poller. A prognostic index for ductal carcinoma in situ of the breast. *Cancer*, 77(11):2267–2274, 1996.

[32] R. Holland, J.L. Peterse, R.R. Millis, V. Eusebi, D. Faverly, MJ Van de Vijver, B. Zafrani, et al. Ductal carcinoma in situ: a proposal for a new classification. In *Seminars in diagnostic pathology*, volume 11, page 167, 1994.

[33] M.B. Daly. Tamoxifen in ductal carcinoma in situ. In *Seminars in oncology*, volume 33, pages 647–649. Elsevier, 2006.

[34] M.J. van de Vijver, J.L. Peterse, W.J. Mooi, P. Wisman, J. Lomans, O. Dalesio, and R. Nusse. Neu-protein overexpression in breast cancer. *New England Journal of Medicine*, 319(19):1239–1245, 1988.

[35] D.B. Cornfield, J.P. Palazzo, G.F. Schwartz, S.A. Goonewardene, A.J. Kovatich, I. Chervoneva, T. Hyslop, and R. Schwarting. The prognostic significance of multiple morphologic features and biologic markers in ductal carcinoma in situ of the breast. *Cancer*, 100(11):2317–2327, 2004.

[36] B. Erbas, E. Provenzano, J. Armes, and D. Gertig. The natural history of ductal carcinoma in situ of the breast: a review. *Breast cancer research and treatment*, 97(2):135–144, 2006.

[37] J.A. Zujewski, L.C. Harlan, D.M. Morrell, and J.L. Stevens. Ductal carcinoma in situ: trends in treatment over time in the us. *Breast cancer research and treatment*, pages 1–7, 2011.

[38] J. Cuzick. Treatment of dcis- results from clinical trials. *Surgical oncology*, 12(4):213–219, 2003.

[39] M.J. Silverstein. The university of southern california/van nuys prognostic index for ductal carcinoma in situ of the breast. *The American journal of surgery*, 186(4):337–343, 2003.

[40] Aslaug Aamodt Muggerud, Michael Hallett, Hilde Johnsen, Kristine Kleivi, Wenjing Zhou, Simin Tahmasebpoor, Rose-Marie Amini, Johan Botling, Anne-Lise Boerresen-Dale, Therese Soerlie, and Fredrik Waernberg. Molecular diversity in ductal carcinoma in situ (dcis) and early invasive breast cancer. *Molecular Oncology*, 4(4):357–368, 2010.

[41] X.J. Ma, R. Salunga, J.T. Tuggle, J. Gaudet, E. Enright, P. McQuary, T. Payette, M. Pistone, K. Stecker, B.M. Zhang, et al. Gene expression profiles of human breast cancer progression. *Proceedings of the National Academy of Sciences*, 100(10):5974, 2003.

[42] S. Barbero, A. Bajetto, R. Bonavia, C. Porcile, P. Piccioli, P. Pirani, J.L. Ravetti, G. Zona, R. Spaziante, T. Florio, et al. Expression of the chemokine receptor cxcr4 and its ligand stromal cell-derived factor 1 in human brain tumors and their involvement in glial proliferation in vitro. *Annals of the New York Academy of Sciences*, 973(1):60–69, 2002.

[43] A. Muller, B. Homey, H. Soto, N. Ge, D. Catron, M.E. Buchanan, T. McClanahan, E. Murphy, W. Yuan, S.N. Wagner, et al. Involvement of chemokine receptors in breast cancer metastasis. *nature*, 410(6824):50–56, 2001.

[44] Total rna isolation with trizol reagent. Protocol, Invitrogen.

[45] Rneasy mini handbook. Protocol Fourth edition, Qiagen, 2010.

[46] Quantification of nucleic acids using a microvolume spectrophotometer. Protocol, NanoDrop Technology, 2009.

[47] NanoDrop Technologies 2009 (web): Micro-Volume UV-Vis Spectrophotometer, http://www.nanodrop.com/ Productnd2000overview.aspx [Accessed 02.03.2012].

[48] Bioanalyzer 2100. Protocol, Agilent Technologies.

[49] One-color microarray-based gene expression analysis. Protocol version 6.5, Agilent Technologies, May 2010.

[50] Agilent 2010 (web): Agilent Microarray Technologies, http://www.genomics.agilent.com/ GenericB.aspx?PageType =Custom&SubPageType=Custom&PageID=2011 [Accessed 02.03.2012].

[51] Genespring gx manual. Manual, Agilent Technologies.

[52] G.K. Smyth and T. Speed. Normalization of cdna microarray data. *Methods*, 31(4):265–273, 2003.

[53] Balasubramanian Narasimhan Robert Tibshirani Virginia Tusher Gil Chu, Jun Li. Sam users guide. Users guide and technical document, Stanford University.

[54] Source. Database (http://source.stanford.edu), Stanford University, 2000.

[55] M.B. Eisen, P.T. Spellman, P.O. Brown, and D. Botstein. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*, 95(25):14863, 1998.

[56] Michael B. Eisen and Michiel de Hoon. Cluster 3.0 manual. Manual, Stanford University, 2002.

[57] Ingenuity Systems 2012 (web): IPA Library, http://www.ingenuity.com/library/index.html [Accessed 23.04.2012].

[58] F. Michor and K. Polyak. The origins and implications of intratumor heterogeneity. *Cancer Prevention Research*, 3(11):1361–1364, 2010.

[59] C. Desmedt, B. Haibe-Kains, P. Wirapati, M. Buyse, D. Larsimont, G. Bontempi, M. Delorenzi, M. Piccart, and C. Sotiriou. Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clinical Cancer Research*, 14(16):5158–5165, 2008.

[60] D. Porter, J. Lahti-Domenici, A. Keshaviah, Y.K. Bae, P. Argani, J. Marks, A. Richardson, A. Cooper, R. Strausberg, G.J. Riggins, et al. Molecular markers in ductal carcinoma in situ of the breast1 1 national cancer institute cancer genome anatomy project and specialized program in research excellence in breast cancer at dana-farber/harvard cancer center (ca89393) and johns hopkins university (ca88843); department of defense breast cancer center of excellence grants. *Molecular cancer research*, 1(5):362, 2003.

[61] J. Hannemann, A. Velds, J.B.G. Halfwerk, B. Kreike, J.L. Peterse, and M.J. Van De Vijver. Classification of ductal carcinoma in situ by gene expression profiling. *Breast Cancer Research*, 8(5):R61, 2006.

[62] C.S. Schuetz, M. Bonin, S.E. Clare, K. Nieselt, K. Sotlar, M. Walter, T. Fehm, E. Solomayer, O. Riess, D. Wallwiener, et al. Progression-specific genes identified by expression profiling of matched ductal carcinomas in situ and invasive breast tumors, combining laser capture microdissection and oligonucleotide microarray analysis. *Cancer research*, 66(10):5278, 2006.

[63] S.H. Park, H. Kim, B.J. Song, et al. Down regulation of bcl2 expression in invasive ductal carcinomas is both estrogen-and progesterone-receptor dependent and associated with poor prognostic factors. *Pathology oncology research: POR*, 8(1):26, 2002.

[64] S. Gruvberger, M. Ringner, Y. Chen, S. Panavally, L.H. Saal, Aake. Borg, M. Ferno, C. Peterson, and P.S. Meltzer. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Research*, 61(16):5979, 2001.

[65] K.E. De Visser, A. Eichten, and L.M. Coussens. Paradoxical roles of the immune system during cancer development. *Nature Reviews Cancer*, 6(1):24–37, 2006.

[66] V.N. Kristensen, C.J. Vaske, J. Ursini-Siegel, P. Van Loo, S.H. Nordgard, R. Sachidanandam, T. Sørlie, F. Wärnberg, V.D. Haakensen, Å. Helland, et al. Integrated molecular profiles of invasive breast tumors and ductal carcinoma in situ (dcis) reveal differential vascular and interleukin signaling. *Proceedings of the National Academy of Sciences*, 109(8):2802–2807, 2012.

[67] Robert A. Weinberg. *The Biology of Cancer*. Garland Science, 2006.

[68] P. Van Loo, S.H. Nordgard, O.C. Lingjærde, H.G. Russnes, I.H. Rye, W. Sun, V.J. Weigman, P. Marynen, A. Zetterberg, B. Naume, et al. Allele-specific copy number analysis of tumors. *Proceedings of the National Academy of Sciences*, 107(39):16910–16915, 2010.

[69] S.M. Pupa, E. Tagliabue, S. Ménard, and A. Anichini. Her-2: A biomarker at the crossroads of breast cancer immunotherapy and molecular medicine. *Journal of cellular physiology*, 205(1):10–18, 2005.

[70] K. Park, S. Han, HJ Kim, J. Kim, and E. Shin. Her2 status in pure ductal carcinoma in situ and in the intraductal and invasive components of invasive ductal carcinoma determined by fluorescence in situ hybridization and immunohistochemistry. *Histopathology*, 48(6):702–707, 2006.

Appendix A

Patient characteristics

Table A.1: **Patient and tumor characteristics**

SampleID	**PgR status (1=pos, 2=neg)**	**ER status (1=pos, 2=neg)**	**HER2 status (1=pos, 2=neg)**	**Grade (1,2,3)**	**Age**
FW08-429	1	1	2	3	42
FW08-421	2	2	1	3	82
FW08-423	n/a	n/a	n/a	2	65
FW08-440	n/a	n/a	n/a	3	56
CMB-195	n/a	n/a	n/a	1	66
BRC-28	2	2	n/a	2	n/a
FW06-001	1	1	2	2	47
BRC-183	1	1	n/a	2	n/a
BRC-239	2	2	1	3	n/a
CMB-349	n/a	n/a	n/a	3	53
FW08-413	2	2	1	3	36
FW06-007	1	1	2	2	74
BRC-288	1	1	n/a	1	n/a
FW08-419	2	1	1	3	49
FW06-24	2	2	1	2	66
FW06-26	1	1	2	2	78
FW08-432	1	1	2	3	71
BRC-19	2	1	n/a	n/a	78
BRC-190	1	1	n/a	1	n/a
CMB-17	n/a	n/a	n/a	n/a	69
FW06-41	2	1	2	2	69
FW06-008	2	2	1	3	55
FW08-400	1	1	2	3	57
FW06-23	1	1	1	2	50
FW06-68	2	2	1	1	40
CMB-354	n/a	n/a	n/a	3	44
BRC-198	1	1	2	1	n/a
BRC-98	n/a	n/a	n/a	n/a	56

Table A.2: **Patient and tumor characteristics (cont.)**

SampleID	PgR status (1=pos, 2=neg)	ER status (1=pos, 2=neg)	HER2 status (1=pos, 2=neg)	Grade (1,2,3)	Age
BRC-202	1	1	1	3	n/a
FW08-418	2	1	2	3	55
BRC-233	1	1	2	3	n/a
FW06-006	1	1	2	3	72
FW06-002	1	1	2	2	65
BRC-153	1	1	n/a	3	n/a
BRC-246	2	2	1	3	n/a
FW08-430	1	1	2	2	48
BRC-779	2	2	n/a	n/a	56
FW06-12	1	1	2	2	52
BRC-180	2	2	n/a	3	n/a
CMB-45	n/a	n/a	n/a	3	60
FW06-25	1	1	1	2	60
CMB-338	n/a	n/a	n/a	1	57
FW06-57	1	1	1	1	72
FW06-005	1	2	2	3	54
BRC-54	n/a	n/a	n/a	n/a	48
BRC-766	1	1	n/a	2	n/a
FW06-42	2	2	1	3	57
FW08-420	1	1	1	3	59
FW08-407	1	1	2	3	66
BRC-14	2	2	n/a	2	n/a
BRC-128	1	1	n/a	2	n/a
CMB-273	n/a	n/a	n/a	n/a	43
CMB-291	n/a	n/a	n/a	3	60
BRC-777	2	2	n/a	n/a	56
BRC-773	2	2	n/a	n/a	45
CMB-215	n/a	n/a	n/a	3	64
FW06-56	2	1	2	2	62
BRC-72	1	1	n/a	2	n/a

Appendix B

Required Reagents and Equipment

Table B.1: **Required reagents, vendor and part number**

Reagent	**Vendor**	**Part number**
RNAse Away	VWR	732-2352
DNase/RNase-free water	Invitrogen	10977-035
100% Ethanol	Sentrallager	RH20173
RNA isolation		
TRIzol	Invitrogen	15596018
Chloroform		
Isopropanol		
RNeasy Mini Kit	Qiagen	74104
Bioanalyzer		
Agilent RNA 6000 Nano Kit	Agilent	5067-1511
RNAse ZAP	Ambion	AM9782
Gene Expression Analysis		
Low Input Quick Amp Labeling Kit, One-Color	Agilent	5190-2305
Rna Spike-In Kit, One-Color	Agilent	5188-5282
Gene Expression Hybridization Kit	Agilent	5188-5242
Gene Expression Wash Buffer Kit	Agilent	5188-5327
RNeasy Mini Kit	Qiagen	74104

Table B.2: **Required equipment**

Equipment	**Vendor**	**Part number**
Agilent Microarray Scanner	Agilent	G2565BA
Hybridization chamber	Agilent	G2534A
Hybridization chamber gasket slides	Agilent	G2534-60014
Hybridization oven	Agilent	G2545A
Hybridization oven rotator	Agilent	G2530-60029
Nuclease-free 1.5 mL microfuge tubes	Ambion	12400
Magnetic stir bar (×2)	Corning	401435
Magnetic stir plate (×2)	Corning	6795-410
Microcentrifuge	Eppendorf	5417R
NanoDrop ND-1000 UV-VIS spectrophotometer	NanoDrop	ND-1000
Slide-staining dish, with slide rack (×3)	Thermo Shandon	121
Bioanalyzer 2100	Agilent	
Circulating water baths or heat blocks		
Clean forceps		
Ice bucket		
Pipetman micropipettors, (P-10, P-20, P-200, P-1000)		
Sterile, nuclease-free aerosol barrier pipette tips		
Vortex mixer		
Timer		
Nitrogen purge box for slide storage		

Appendix C

Experimental data

Table C.1: **Data generated from the experiments**

FW06	Diagnose (DCIS=1, IDC=2)	c [ng/ul] RNA	260/280	260/230	RNA integrity number	
1	1	570	1,82	2,35	3,0	
2	1	86	1,78	2,26	4,0	
5	1	1518	1,81	2,31	7,9	
6	1	550	1,85	2,22	5,6	
7	1	1001	1,83	2,03	5,7	
8	1	303	1,74	2,23	6,0	
12	2	226	1,78	1,98	7,3	
23	1	979	1,85	2,29	8,1	
24	1	328	1,65	2,10	4,2	
25	2	25	1,34	0,95	3,2	
26	2	624	1,81	2,30	7,7	
41	1	264	1,84	2,18	7,1	
42	1	111	2,01	1,78	6,4	
56	2	144	1,72	1,63	1,6	
57	2	251	1,79	1,79	4,8	
68	2	266	1,81	1,81	6,2	
FW08	**Diagnose (DCIS=1, IDC=2)**	**c [ng/ul] RNA**	**260/280**	**260/230**	**RNA integrity number**	
400	1	48	1,41	1,19	n/a	
407	1	23	1,41	4,36	n/a	
413	1	79	1,49	1,37	6,0	
418	1	446	1,95	1,97	3,1	
419	1	116	1,64	1,52	7,6	
420	1	321	1,45	0,35	n/a	
421	1	311	1,76	1,14	6	
423	1	447	1,88	1,72	4,4	
427	1	370	1,37	0,32	n/a	Failed
429	1	158	1,73	1,68	8,2	
430	1	408	1,77	1,43	5,9	
432	1	183	1,65	1,80	5,1	
440	1	546	1,85	1,73	5,9	

Table C.2: **Data generated from the experiments (cont.)**

BRCNr	Diagnose (DCIS=1, IDC=2)	c [ng/ul] RNA	260/280	260/230	RNA integrity number	
19	1	137	1,65	1,14	3,6	
54	1	201	1,66	1,20	5,9	
98	1	240	1,68	1,24	6,5	
773	1	254	1,50	1,00	2,0	
777	1	260	1,38	0,62	n/a	
779	1	277	1,62	1,10	6,3	
190	2	285	1,48	0,78	5,0	
149	2	280	1,43	0,61	n/a	Failed
202	2	132	1,65	1,09	7,1	
233	2	239	1,81	1,53	8,2	
28	2	282	1,78	1,23	2,7	
128	2	305	1,83	1,13	n/a	
183	2	388	1,76	1,40	5,4	
246	2	554	1,83	1,68	5,5	
288	2	578	1,58	1,06	6,3	
766	2	512	1,46	1,04	n/a	
198	2	46	1,51	1,00	6,7	
14	2	452	1,84	2,20	6,5	
72	2	348	1,25	0,71	n/a	
153	2	387	1,77	2,09	7,7	
180	2	600	1,57	1,44	4	
239	2	584	1,72	2,16	8,6	
259	2	620	1,34	1,40	n/a	Failed
738	2	312	1,41	1,30	n/a	Failed
CMB	**Diagnose (DCIS=1, IDC=2)**	**c [ng/ul] RNA**	**260/280**	**260/230**	**RNA integrity number**	
215	1	77	1,66	1,07	8,7	
273	1	165	2,03	1,96	7	
291	1	74	1,88	1,55	6,9	
17	1	571	1,79	2,12	8,4	
349	1	41	1,58	0,93	5,1	
354	1	123	1,71	1,67	7,0	
45	1	25	2,00	2,18	6,9	
195	1	234	1,86	0,36	8,6	
338	1	30	1,70	1,07	6,2	