NTNU – Trondheim
Norwegian University of
Science and Technology

Differentially Regulated pathways of Potential Importance for Treatment Response and Cardiac Toxicity after Administration of Doxorubicine to BC Patients

Tone Olsen

Biotechnology (5 year)

Supervisor: Per Bruheim, IBT
Co-supervisor: Hege Edvardsen, Radiumhospitalet

Norwegian University of Science and Technology
Department of Biotechnology

Acknowledgements

This thesis was performed as a part of the requirement for obtaining a Master's degree in Biotechnology at the Department of Biotechnology at The Norwegian University of Science and Technology (NTNU). The thesis was accomplished at Department of Genetics, Institute of Cancer Research, Norwegian Radium Hospital in Oslo.

I would like to express my gratitude to Professor Anne-Lise Børresen-Dale for giving me the opportunity to do my master thesis at the Department of Genetics. I would also like to thank Professor Vessela N. Kristensen for including me in her project group at the department and for giving me helpful feedback on my work. My special thanks goes to Dr. Hege Edvardsen for being an excellent supervisor throughout this thesis. I appreciate very much your scientific knowledge, your follow up and support and your positive spirit. Also, I would like to thank my supervisor at NTNU, Professor Per Bruheim, for taking care of all the formalities during this thesis. I am grateful to Phoung Vo who has helped me during the technical lab part, as well as Eldri Undlien Due and Anita Halvei for their help in the lab. In addition, I wish to express gratitude to Thomas Fleischer and Daniel Nebdal for their sharing of statistical- and computer knowledge. Furthermore, I would like to thank all at the Department of Genetics for all your help and support, you have made me feel very welcome at the department.

Cathrine, I am very grateful for having you as a friend and for always being there for me when I have needed it at most during this thesis. I would also like to thank all my lovely friends in Trondheim at NTNU and in Moss for being such an important part of my life.

Finally, I would like to thank my Mom, Dad and my sister Hanne for everything you have given me in life. You mean extremely much to me and I am forever grateful for your support and encouragement during my Master degree.

Oslo, May, 2012

Tone Olsen

Table of Contents

Abstract

Doxorubicin is a topoisomerase-targeting anthracycline that is one of the most effective anticancer drugs currently known. However, its clinical use is restricted by cardiotoxicity and the development of drug resistance. The main goal of this thesis has been to increase the knowledge of doxorubicin mechanism in addition to evaluate if predictive biomarkers for doxorubicin response could be identified. A total of 128 tumor samples collected from breast cancer patients before and after neoadjuvant treatment with doxorubicin were studied. mRNA expression level in tumor tissue was assessed using whole-genome mRNA microarray analysis (Agilent Human GE 4x44K microarray).

More than 5000 genes were found to be up- and down regulated following doxorubicin treatment. The molecular and cellular functions as well as canonical pathways found to be enriched in the list of genes up regulated after doxorubicin exposure were involved in among other cardiovascular system development and function, cellular movement and immune responses. p53 was found to be the transcription factor regulating the highest number of target molecules within the list of up regulated genes. RNA processing, splicing and translation were shown to be overrepresented in the list of down regulated genes. The association between doxorubicin response and changes in gene expression revealed several genes such as *CTGF*, *ITGB4* and *IGF1* to be up regulated in the samples collected from patients characterized with a partial response to doxorubicin compared to those with minimal change and/or stable disease following treatment. In addition, the gene expression profiles between samples having wild type compared to mutated *TP53* were studied, and a lower induction of expression were found for several genes such as *FGF9* and *COL11A2* in the samples having a mutated p53.

This study showed that the gene expression profile in breast cancer tumors is altered as a response to doxorubicin exposure. Identifying genes significantly altered after therapy and associate their change with response to treatment may help identify the subgroup of patients benefitting from doxorubicin treatment. Patients with little or no effect of treatment could receive alternative therapy and be spared unnecessary treatment and risk of side effects.

Sammendrag

Doxorubicin er et topoisomerase-hemmende antracyklin og er et av de mest effektive cytostatiske legemidlene brukt i kreftbehandling. Den kliniske bruken av doxorubicin er begrenset som følge av den irreversible kardiotoksiske effekten og pasienters utvikling av resistens. Hovedmålet med denne oppgaven har vært å øke forståelsen av de molekulærbiologiske mekanismene bak den cytotoksiske effekten av doxorubicin i tillegg til å prøve å identifisere prediktive biomarkører for doxorubicin respons. Totalt 128 tumorprøver samlet inn fra brystkreftpasienter før og etter neoadjuvant behandling med doxorubicin ble studert. mRNA ekspresjonsnivå i tumorvev ble målt ved hel-genom mRNA mikromatrise analyse (Agilent Human GE 4x44K microarray).

Mer enn 5000 gener ble funnet å være opp- og nedregulert etter doxorubicin behandling. De molekulære og cellulære funksjonene samt kurerte reaksjonsveiene som ble funnet overrepresentert i listen av gener oppregulert etter doxorubicin behandling var involvert i blant annet utvikling og funksjon av av hjerte-og karsystem, cellulær bevegelse og immunrespons. p53 ble funnet å være den transkripsjonsfaktoren som regulerte flest gener i listen med de oppregulerte genene. RNA prossesering, spleising og translasjon ble vist å være overrepresentert i listen med nedregulerte gener. Sammenhengen mellom doxorubicin respons og endringer i genuttrykket avdekket flere gener som *CTGF*, *ITGB4* og *IGF1* til å være oppregulert i tumorprøver fra pasienter som hadde en partiell respons til doxorubicin sammenlignet med dem som hadde en minimal endring og/eller stabil sykdom etter behandling. I tillegg ble ekspresjonsprofiler mellom prøver med villype sammenlignet med mutert *TP53* studert, og en lavere induksjon i ekspresjon ble funnet for flere gener som *FGF9* og *COL11A2* i prøvene med mutert p53.

Dette studie viser at ekspresjonsprofilen i brystkreftsvulster endres etter doxorubicin eksponering. Identifiseringen av gener signifikant endret som følge av behandling og sammenhengen mellom disse endringene mot respons på behandling kan bidra til å identifisere grupper av pasienter som kan ha utbytte av doxorubicin kjemoterapi. Pasienter med liten eller ingen effekt av behandlingen bør heller få alternativ terapi og bli spart for unødvendig behandling og risiko for bivirkninger.

1. Introduction

Cancer is one of the leading causes of death and in 2008 the cancer mortality was close to 7.6 million deaths worldwide (1). The 5-year cancer prevalence for the adult population only was 28.8 million worldwide the same year (2), while the number of people diagnosed with cancer in Norway was 27520 in 2009 (3).

1.1 Cancer as a Genetic Disease

Cancer is a genetic disease caused by either inherited mutations in genes that control genome integrity or by mutations that are acquired in somatic cells during the development of a tumor (4), often in response to aging, lifestyle or- environmental exposure. Several types of DNA changes can arise and facilitate the change of a normal cell to a tumor cell, such as mutations in single nucleotides as well as alterations in small stretches of DNA, whole genes, structural components of chromosomes or complete chromosomes (4). In addition, epigenetic changes not directly altering the DNA sequence are also involved in cancer development (5).

Two major groups of genes are repeatedly altered in human tumors (6). Genes of the first group are called proto- oncogenes and activating mutations in these genes lead to an overactive or over expressed form which can result in cancer development (6). Genes of the second group are called tumor suppressor genes and inactivating mutations in these genes lead to elimination of the gene and can promote cell transformation (6).

1.1.1 Cancer development and progression

Although there are several different types of cancer, they all have in common that genetic and epigenetic changes allow them to escape the normal regulatory machinery of the cell resulting in uncontrolled growth and proliferation (6). Cancer is a polygenic disorder, meaning that mutations in several genes are required for cancer to develop (4). Two of the models explaining tumor evolution are the clonal evolution model and the cancer stem cell model (7). In the clonal evolution model a cell acquires alterations which gives it a selective advantage (6). When this cell grows and divides its descendants may gain a substantial number of favorable genetic and epigenetic alterations which can lead to clonal expansion and formation of a tumor (6). In this model, all clones have the capacity to undergo proliferation and Darwinian

selection (7). In contrast, the cancer stem cell hypothesis is based on a cancer stem cell precursor which can both be self-renewing and also give rise to a progeny of differentiated cancer cells (7). This model suggests that only the cancer stem cells have the potential of self-renewal and can drive tumor progression (7).

Cancer development is divided into several stages (6). Cells that have obtained favorable mutations or epigenetic changes give rise to a benign tumor which is non-invasive (6). Some benign tumor develops into malignant tumors which contains cancer cells that have gain the ability to break through the basement membrane and invade the surrounding tissue (6). These cells can enter the bloodstream or lymphatic vessels and form secondary tumors in a process called metastasis (6).

1.1.2 Hallmarks of cancer

During cancer evolvement the cell acquires several capabilities to become tumorigenic and ultimately malignant which are often referred to as the "Hallmarks of cancer"(8). Hanahan and Weinberg proposed in 2000 six physiologic changes that a cell has to obtain in order to become malignant (Figure 1) (9). These hallmarks included sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis and activating invasion and metastasis (9). The authors argued that these acquired hallmarks enhance cell proliferation and decrease cell death, make the cells able to receive nutrients and oxygen by the formation of new blood vessels and eventually give the cells metastatic capabilities necessary for the establishment of new colonies at a distant site in the body.

Eleven years later Hanahan and Weinberg published a follow up article where two new emerging hallmarks were presented, deregulating cellular energetics and avoiding immune destruction (Figure 1) (8). They proposed that by modifying or reprogramming cellular metabolism the cells can use glycolytic intermediates in biosynthetic pathways to generate nucleosides, amino acids, macromolecules and organelles required for active growth and proliferation (8). In addition, the authors suggested that evading immune recognition and suppressing immune reactivity also play an important role in the development and progression of many human cancers (8).

Two enabling characteristics required for the all the eight hallmarks to evolve was also implicated by Hanahan and Weinberg (Figure 1) (8). These included genome instability and mutation and tumor-promoting inflammation (8). The authors suggested that accumulation of mutations or epigenetic alterations can generate genotypes contributing to the acquisition of the hallmark capabilities. In addition, it was proposed that inflammatory cells of the tumor microenvironment can supply several bioactive molecules to the tumor facilitating angiogenesis, invasion and metastasis (8).

Figure 1. Hallmarks of cancer. The essential characteristics that enable tumor growth and metastatic dissemination proposed by Hanahan and Weinberg are six acquired capabilities (1-6), two enabling characteristics (7-8) and two emerging hallmarks (9-10) (8) (Figure modified from (8)).

1.1.3 The tumor microenviroment

The tumor microenvironment has been implicated in cancer progression (8). The microenvironment is composed of a diversity of cell types, such as vascular endothelial cells, adipocytes, fibroblasts, macrophages and immune cells among others, which is suggested to interact with tumor cells via growth factor and cytokine networks and promote tumor cell invasion and metastasis (10). Several studies have indicated that the tumor microenvironment undergoes extensive changes during the evolution and progression of cancers. The interactions between the cells of the stromal tissue and the cancer cells are bidirectional and dynamic (11). Cancer cells recruit and activate stromal cell types such as tumor-associated fibroblasts,

macrophages, leucocytes and mast cells which in turn secrete signalling molecules promoting progression of the tumor (11).

1.2 Breast Cancer

Breast cancer is the most frequent cancer type among women, and in 2008 1.38 million women were diagnosed with breast cancer worldwide (2). In Norway the number of newly diagnosed cases each year is more than 2700 (3). As a consequence of better diagnostics and treatment the 5-year relative survival is increasing and between 2005-09 more than 88 percent of women diagnosed with breast cancer survived 5 years (all stages combined) (3). Still, the number of breast cancer deaths in Norway was the second highest among cancer deaths affecting women in 2009 (3).

1.2.1 Breast biology and development

The mammary gland is a complex organ consisting of several different cell-and tissue types (Figure 2). The two main components of the human breast are the stroma and the parenchyma. The parenchyma is the functional system of branching ducts and secretory lobules, while stroma is the supporting adipose tissue, fibroblasts, blood vessels and lymphatic vessels that surrounds the ducts and the lobules (12). The lobules of the parenchyma consist of secretory epithelial cells that synthesize milk, and from each lobule large milk ducts transport the milk produced in the lobules to the nipple (13). The lobules and ducts consist of a bi-layer of luminal epithelial cells and myoepithelial cells. The luminal epithelial cells are covering the surface of the milk ducts and are thus lining the central lumen, while the underlying myoepithelial cells are located adjacent to the basement membrane which functions as a physical barrier preventing cells from invading the surrounding tissue (Figure 2, lower section) (14). The arteries, veins and the lymphatic system of the stroma are responsible for the transportation of blood and nutrients to the breast and the exportation of waste products.

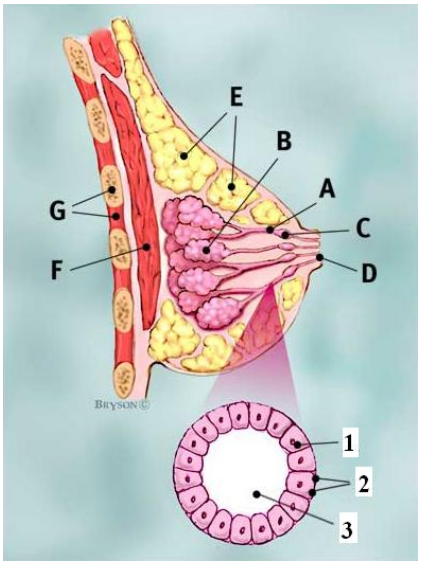

Figure 2. Anatomy of the breast. Above: The major components of the human breast are the ducts (A), the lobules (B), the dilated section of duct to hold milk (C), the nipple (D), fat (E), pectoralis major muscle (F) and chest wall/rib cage (G). Below: An cross section of a normal duct showing a monolayer of epithelial cells (1), surrounded by the basement membrane at the external side (2) and the lumen which is the center of the duct (3) at the interior (Figure modified from (15)).

Hormones play an important role in normal breast development. Between birth and puberty, the development of the mammary gland is relatively dormant. However, during puberty there is a rapid expansion of the mammary epithelium which generates an extensive ductal network driven among others by the steroid hormones estrogen and progesterone (16). Throughout life the breast is continuously undergoing changes. During the menstrual cycle there is a fluctuation of hormones and proliferation followed by apoptosis of the epithelial cells (17). And during pregnancy, a large number of milk protein- and lipid secreting luminal cells are developed in addition to further ductal branching and infiltration of the stroma (17;18).

1.2.2 Breast cancer initiation and progression

Breast tumors have different histopathological appearance depending on the origin of the tumor and if it has penetrated the basement membrane or not. Invasive breast carcinomas are the most common tumors of the female breast and consist of a group of malignant epithelial tumors. The majority of these tumors are adenocarcinomas derived from the parenchymal epithelium of the mammary gland (19). Invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC) are the two major histological types of invasive breast carcinomas, constitute approximately 80% and 10-15% of all invasive breast cancers, respectively (20). Ductal carcinoma *in situ* (DCIS) which is considered an precursor of IDC consists of cells that have not penetrated the basement membrane and accounts for about 20-25% of newly diagnosed cases (21).

Breast cancer originates when a normal epithelial cell of the breast acquires genetic and epigenetic alterations, and tumor progression is driven by further accumulations of these abnormalities (22). Telomere shortening leading to genomic instability (23) and microenviromental alterations (22) are other drivers of tumorigenisis. Breast cancer progression of ductal carcinomas is a multistep process (Figure 3) starting with atypical ductal hyperplasia (ADH) which is a premalignant lesion of abnormal epithelial cells within the duct. Next, ductal carcinoma *in situ* (DCIS) evolves when the basement membrane is degraded and the carcinoma becomes invasive. The cells can then escape the primary site and infiltrate the surrounding tissue. Metastatic carcinomas arises when the cells of the invasive carcinoma have established a new tumor at a distant site (14).

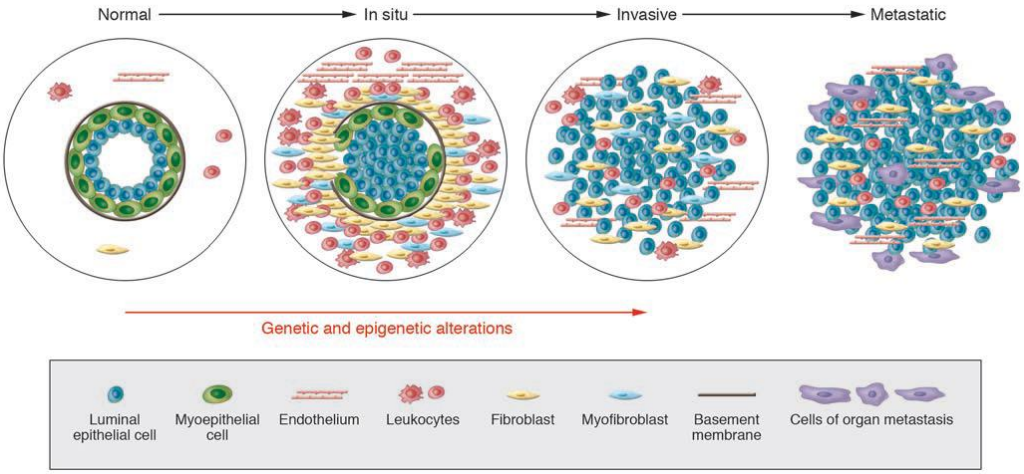

Figure 3. Schematic outline of ductal breast tumor progression. In normal breast ducts or lobules the basement membrane encapsulates the bi-layer of epithelial and myoepithelial cells. In ductal carcinoma *in situ* the epithelial cells have increased extensively in number due to increased proliferation. Degradation of the basement membrane result in invasive carcinoma, and the cells can escape the primary site. Metastatic carcinoma arises when cancer cells from the invasive carcinoma enter the vasculature of the primary stroma and seed secondary tumors in a foreign microenviroment (7).

This traditional model explains breast cancer progression in a linear, step-wise fashion. However, it has been indicated that this model is overly simplistic (24). Thus, several different models explaining breast cancer progression have been proposed and will be discussed later in the context of intratumor heterogeneity (see chapter 1.4.1).

1.2.3 Risk factors

Both hereditary and environmental risk factors are involved in the probability of developing breast cancer. Familial breast cancer constitutes about 5-10% of all breast cancer cases, while the remaining are classified as sporadic breast cancer (Figure 4). Germline mutations in high-penetrance cancer susceptibility genes, including the double-strand DNA break repair genes *BRCA1* and *BRCA2*, account for about 16% of the risk of developing familial breast cancer (Figure 4) (25). The remaining familial risk may be due to genetic or environmental factors. However, studies have indicated that inherited genetic factors predominate the excess familial risk (26). Mutations in several other breast cancer susceptibility genes, including *TP53*, *PTEN* and *ATM*, have been implicated to account for a smaller proportion of the familial risk (Figure 4) (27;28).

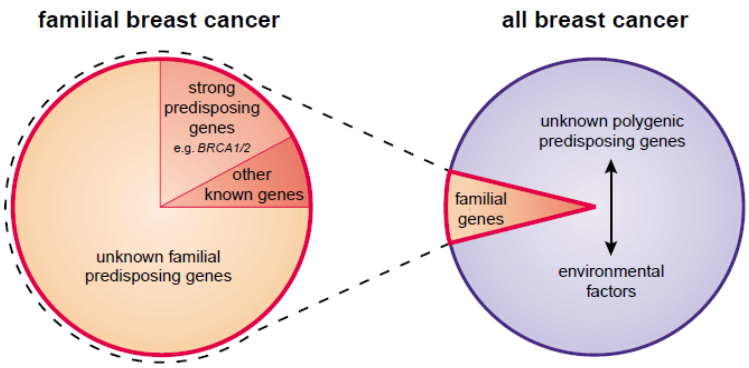

Figure 4. Familial and sporadic breast cancer. Familial breast cancer constitutes only 5-10% of total breast cancer (right panel). *BRCA1* and *BRCA2* are high-penetrance breast cancer susceptibility genes, and mutations in these genes account for about 20% of the familial risk (left panel). Most of the genetic variants that contribute to the risk of developing sporadic breast cancer are unknown (4).

The majority of all breast cancers are sporadic resulting from acquisition of numerous somatic mutations. However, most of the somatic mutations leading to sporadic breast cancer are unknown (4). Genome-wide association studies (GWAS) have identified a number of new genetic variants such as single nucleotide polymorphisms (SNPs) influencing breast cancer risk (Figure 5) (29). SNPs are inherited genomic variations in the DNA sequence and the most frequent type of variation in the human genome (30). SNPs occur when a single nucleotide at a specific locus in the genome differs between individuals in a population or between populations. The variation in a single

nucleotide has to be above 1 % in frequency in the population to be called a SNP. DNA variations at lower frequencies are known as rare variants or mutations (31).

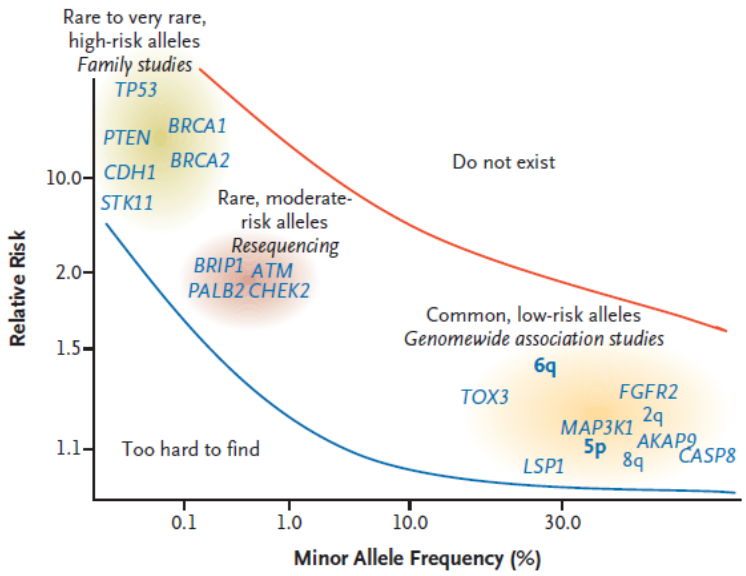

Figure 5. Genetic risk factors including breast cancer susceptibility loci and genes. All known breast cancer susceptibility genes are shown between the red and blue line. High-risk genes are highlighted in green, moderate-penetrance genes are highlighted in red while the low-risk genes are shown in orange (32).

A number of environmental and lifestyle breast cancer risk factors have been identified. Age, exposure to the hormone estrogen, diet, alcohol consumption and exposure to ionising radiation have all been implicated to increase the risk of developing breast cancer (33). Early onset of menarche, late menopause, late age at first full-term pregnancy and obesity influence the estrogen exposure and are therefore contributing factors to breast cancer risk (33). In addition, exogenous estrogen exposure such as oral contraceptives and hormone replacement therapy also increase the risk of breast cancer development although only transiently (33).

1.3 Breast Tumor Classification

Breast cancer is a highly heterogeneous disease considering the diversity in genetic and genomic variations, histopathological features and patient outcomes (14). Breast tumors can be classified by a number of approaches into distinct groups in order to assess the most favourable treatment regime or likely outcome for each patient. Both

prognostic markers that provide information regarding patient outcome and predictive markers that give information about response to a given drug or treatment are identified by tumor classification (34). Guidelines from World Health Organization classify breast cancer into several clinical categories based on histological tumor type, histological grade, tumor stage, expression of estrogen receptor alpha (ER) and the progesterone receptor (PR) and over expression of the human epidermal growth factor receptor 2 (HER2) (19). Here, the traditional clinical features for breast tumor classification will be described.

1.3.1 Histological grade

Histological grading of breast cancer is based on the differentiation of the tumor tissue (35). Three morphological features, which includes nuclear pleomorphism (changes in size, shape and color), mitotic count (the proportion of dividing cells) and to what extent the tumor has formed tubules, are evaluated during the grading (36). Together these three features are translated into a tumor grade raging from 1-3. Grade 1 corresponds to a well-differentiated tumor with high homology to normal breast tissue, showing a mild degree of cellular pleomorphism and low mitotic count. Grade 2 corresponds to a moderately differentiated tumor, and grade 3 to a poorly differentiated tumor with a high degree of nuclear atypia, frequent mitosis and no tubule formation (35). Histological grade has been shown to have a prognostic value, patients with grade 1 tumors have markedly better survival than those with grade 2 and 3 tumors (37).

1.3.2 Tumor stage

Breast tumor staging is based on the international standard tumor-node-metastasis (TNM) staging system (38). In the TNM staging system three features of the primary tumor, tumor size (**T**), regional lymph node status (**N**) and the presence of distant metastasis (**M**), are evaluated (39). Tumor size (T) is divided into four categories, where T1 includes tumors that are smaller or equal to 2 cm, T2 compromises tumors between 2 and 5 cm, T3 consist of tumors greater than 5 cm and T4 includes tumors which have extended into either the chest wall or to the skin. Regional lymph node status (N) is raging from N0-N3 and categorizes tumors according to the number and location of positive regional lymph nodes. Distant metastasis (M) reflects whether or not the primary tumor has metastasized to distant sites. The three TNM parameters are

combined in a TNM stage grouping system which classifies breast tumors into four stages (Table 1) (39). The distinct stages show prognostic differences, the relative survival is decreasing with increasing tumor stage (40).

Table 1. TNM classification of breast tumors. Stage 0 consists of carcinoma *in situ* tumors with negative node status and no distant metastasis. Stage 1 includes small (T1) and localized tumors. Stage 2 consists of tumors with different sizes (T0-T3), different nodal involvement (N0-N1) and no distant metastasis. Stage 3 includes either tumors of different size (T0-T3) with many positive lymph nodes (N1-N2) or tumors that have extended the chest wall or skin (T4) with various lymph node status (N0-N2) and no distant metastasis. Stage 4 comprises tumors of any size and lymph node status that have metastasized to distant organs (39).

TNM Stage Grouping for Breast Cancer

Stage Grouping			
0	Tis	N0	M0
I	T1*	N0	M0
IIA	T0	N1	M0
	T1*	N1	M0
	T2	N0	M0
IIB	T2	N1	M0
	T3	N0	M0
IIIA	T0	N2	M0
	T1*	N2	M0
	T2	N2	M0
	T3	N1	M0
	T3	N2	M0
IIIB	T4	N0	M0
	T4	N1	M0
	T4	N2	M0
IIIC	Any T	N3	M0
IV	Any T	Any N	M1

NOTE. Adapted with permission of the American Joint Committee on Cancer (AJCC), Chicago, IL. The original source for this material is the *AJCC Cancer Staging Manual, Sixth Edition* (2002) published by Springer-Verlag New York, www.springer-ny.com.

*T1 includes T1mic.

1.3.3 Receptor status

Evaluation of the two hormone receptors estrogen receptor (ER) and progesterone receptor (PR) in addition to the human epidermal growth factor receptor 2 (HER2) are routinely used as prognostic and predictive markers for management of breast tumors (41). ER and PR are both members of the nuclear hormone receptor family, and are stimulated by the estrogen and progesterone hormone, respectively (16). Estrogen is the main hormone that controls breast cancer proliferation by interacting through ER

(42). Expression of the two hormone receptors are assessed by immunohistochemistry (IHC) which measure the protein expression of ER and PR. ER status is critical for the identification of breast cancer patients that may benefit from anti-estrogen (endocrine) therapy such as ER antagonists (*e.g.* tamoxifen) or aromatase inhibitors (43;44). PR status is generally correlated with ER status, however the presence of this hormone receptor has less clinical significance (45). Considering the prognostic value of hormone receptor status, breast cancer patients with ER+ and/or PR+ tumors show the best overall outcome (46).

The HER2 receptor tyrosine kinase is a member of a family of transmembrane growth factor receptors that are involved in regulating normal cell proliferation and survival. The receptor is over expressed in approximately 20% of all breast tumors (47). HER2 is a marker for poor prognosis with a decreased overall survival (48;49). The level of HER2 protein expression is evaluated by IHC while HER2 gene amplification is assessed by fluorescence *in situ* hybridization (FISH) (41). HER2+ breast tumors can be treated with targeted therapy such as the monoclonal antibody trastuzumab, and treatment with such therapies has improved the overall outcome for patients in this group (50;51). Combination of the ER, PR and HER2 receptor status allow for the classification of breast tumors into different categories with distinct prognosis. Triple negative tumors (ER-/PR-/HER2-) are associated with the worst prognosis (52).

1.4 Molecular Subtypes of Breast Cancer

Breast cancer is no longer believed to be a single disease, but rather heterogenous both at the clinical and the molecular level (7). The heterogeneity of breast tumors is both intratumoral and intertumoral. Intratumor heterogeneity is the variation within a single tumor. In contrast, intertumor heterogeneity is variation between tumors from different patients (34). These variations can be captured by a diversity of methods in molecular biology and molecular genetics and used to classify the tumors into molecular subgroups. The subgroups may have different risk factors, clinical presentation, histopathological features, outcome and response to therapy (53). The aim behind molecular classification of breast cancer is to individualize the diagnosis and treatment of breast cancer and better predict the clinical outcome for patients in each subgroup (34;53).

1.4.1 Intratumor heterogeneity

As mentioned above, two proposed models explaining tumor progression are the clonal evolution and the cancer stem cell model (22). However, new insights into breast cancer progression have led to several proposed models aiming to explain the progression of breast tumors and also to explain the appearance of subpopulations within tumors (Figure 6) (24). The clonal evolution and the cancer stem cell models have both been proposed as linear models giving rise to monogenomic tumors, and also as diversity models giving rise to polygenomic tumors where multiple clones reside within the tumor (34). In addition, a separate mutator phenotype model in which tumors evolve by gradual and random accumulations of mutations as the tumor grows are also proposed and give rise to highly diverse tumors (34).

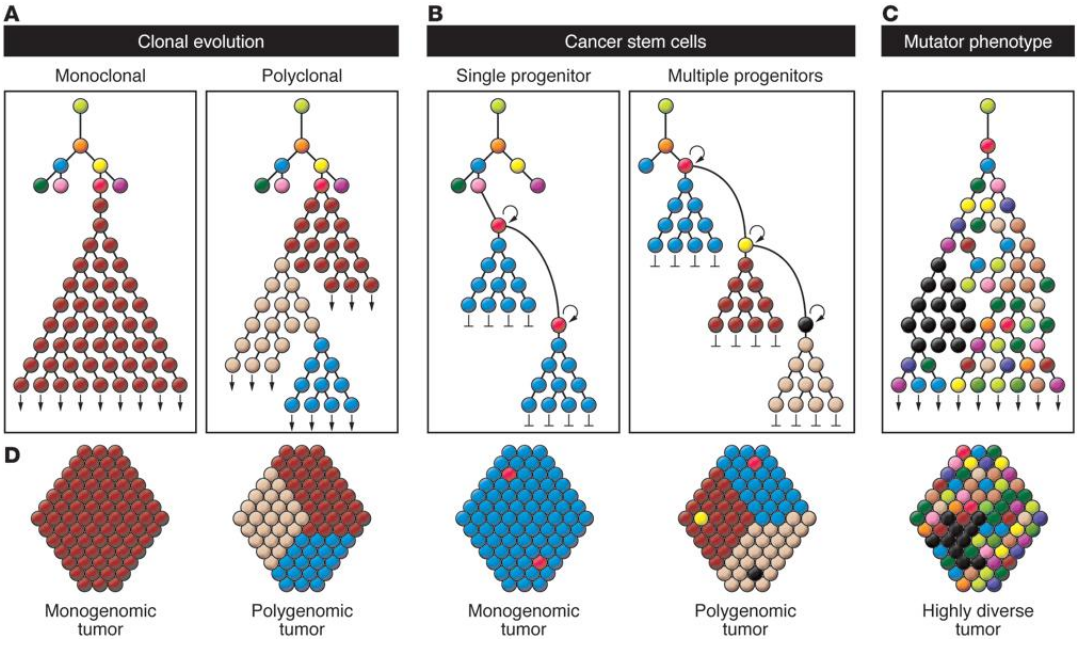

Figure 6. Three hypothetical models explaining intratumor heterogeneity. The clonal evolution (A), the cancer stem cell (B) and the mutator phenotype (C) models are different cancer progression models explaining how distinct types of intratumor heterogeneity arise. Different subpopulations of tumors (D) can be a result of the different models of tumor progression (34).

1.4.2 Intertumor heterogeneity

Gene expression studies have identified several distinct breast cancer subtypes. Perou and Sørlie classified breast cancer tumors into five subtypes based on their gene expression profile. They performed hierarchical clustering by using breast cancer specimens and a subset of genes they termed the 'intrinsic' gene set. These genes were found to be more differentially expressed between tumors from different patients, than between tumor samples from the same individual sampled twice. Based on the gene expression profiles they observed after the clustering analysis, the tumors were divided into five subtypes which included luminal A, luminal B, ERBB2+ (later designated HER2-enriched), basal-like and normal-like tumors (Figure 7) (54;55).

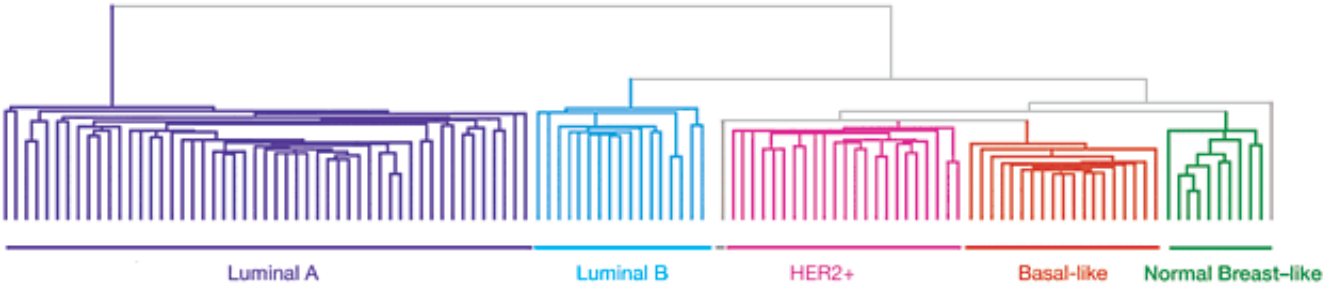

Figure 7. Breast cancer molecular subtypes classified by gene expression profile. Five distinct breast cancer subtypes were identified by Perou and Sørlie on the basis of their gene expression patterns. The molecular subtypes compromised luminal A (purple), luminal B (light blue), HER2-enriched (pink), basal-like (red) and normal-breast like (green) (Figure modified from (56)).

Luminal A and B subtypes are mostly ER positive and express many genes characteristic for breast luminal epithelial cells (55). Luminal A tumors are strongly ER positive and are also characterized as being HER2 negative, have low expression of proliferation associated genes such as Ki-67 and *TP53* tend to be wild-type (57). The Luminal B subtype is weak or moderate ER positive, HER2 positive, has higher proliferation signature and mutated *TP53* (57). In contrast to the luminal subclasses, the other three subtypes are characterized by low or absent expression of ER. The HER2-enriched subtype has high expression of several genes in the HER2 amplicon, while the basal-like tumors are often ER, PR and HER2 negative which is clinically referred to as triple negative tumors. The normal breast-like tumors express genes associated with normal breast tissue, they show strong expression of basal epithelial cells and low expression of luminal epithelial cells (54;55). Recent studies have

identified a new breast cancer intrinsic subtype called claudin-low which is characterized by mesenchymal and stem cell features (58).

Several models explaining the intertumoral heterogeneity of breast cancers have been proposed, suggesting that the intrinsic subtypes may arise from different cells of origin and have various progression pathways (7). Two of these models are the 'cell of origin'- and 'tumor subtype-specific transforming event' models (Figure 8) (7). In the 'cell of origin' model each tumor subtype is initiated in a different cell type. While in the 'tumor subtype-specific transforming event' model the cell of origin can be the same for all the different subtypes, but specific genetic and epigenetic events give rise to the distinct molecular subtypes (7).

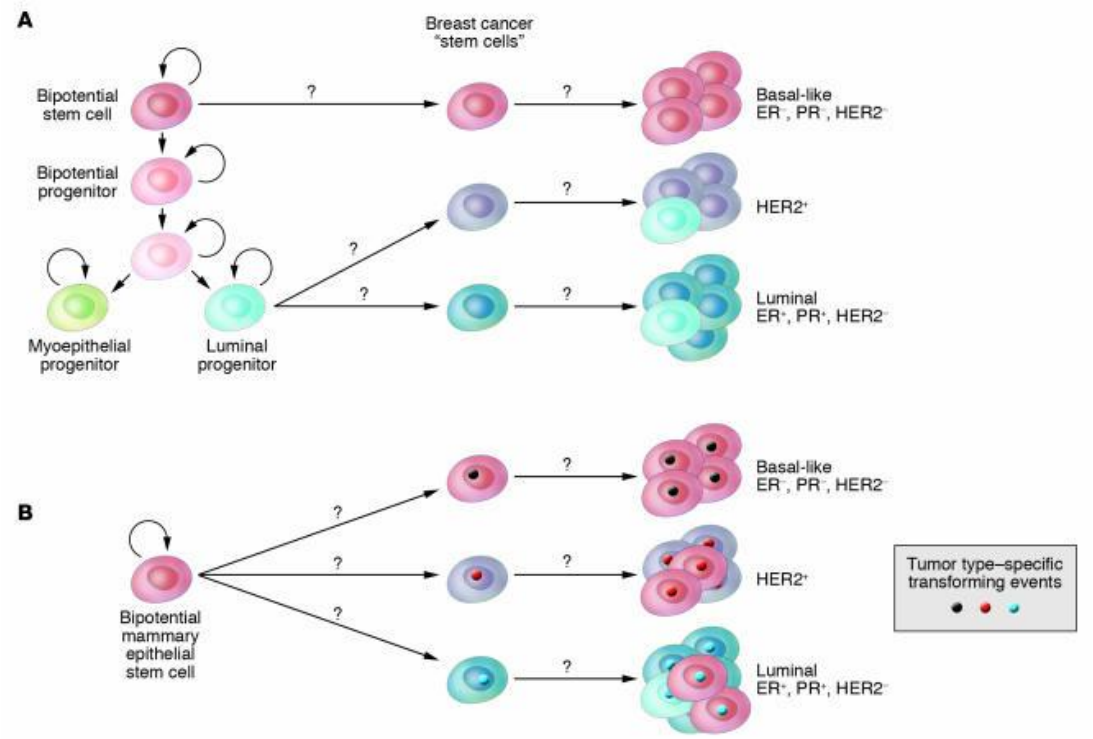

Figure 8. Two hypothesized models explaining the development of the distinct breast cancer intrinsic subtypes. Different cell types give rise to the distinct tumor subtypes in the 'cell of origin' model (A). In contrast, the same cell gives rise to the subtypes in the 'tumor subtype-specific transforming event' model. In the latter model both mutations and epigenetic changes are involved in the generation of the different subtypes (B).

Perou and Sørlie showed that the classification of breast cancer tumors based on gene expression patterns has clinical implications, and this have been validated by other studies. The subgroups display differences in overall and relapse-free survival (55). In

particular, basal-like, HER2-enriched and luminal B subtypes have the worst prognosis compared to the luminal A subtype which show the most favourable outcome (55).

The discovery of variations in gene expression between the molecular subtypes of breast cancer has lead to the development of commercial diagnostic tests to determine prognosis and predict response to treatment (59). MammaPrint DX was the first prognostic signature approved by the FDA (US Food and Drug Administration) and is a microarray-based test that measures the gene expression of 70 genes (60). The assay is designed to predict the clinical outcome of breast cancer patients (development of distant metastasis within 5 years) and in addition to give a predictive value for chemotherapy response (61). Oncotype DX is another prognostic signature that estimates the expression level of 21 genes by qRT-PCR. The assay predicts the risk of distant relapse at 10 years by a recurrence score (62) in addition to predict the chemotherapy response for patients in each risk group (59).

1.5 Breast Cancer Treatment

Treatment options for breast cancer are based on characteristics of the patient as well as characteristics of the tumor (63). The treatment of breast cancer includes surgery, radiotherapy, chemotherapy, endocrine therapy and targeted therapy (63). In Norway, treatment for breast cancer patients is standardized nationally and follows the guidelines to the Norwegian Breast Cancer Group (NBCG) (63). Patient information such as age and genetic predisposition are evaluated to tailor the treatment. Information regarding tumor size, histological stage and grade, ER, PR and HER2 receptor status and expression of the proliferation associated gene Ki-67 are used in the decision making regarding therapy regime.

1.5.1 Chemotherapy

Systemic treatment, such as chemotherapy, is traditionally administered after the tumor is surgically removed (adjuvant) to reduce the risk of recurrence and to increase the overall survival (64). In addition, it can also be administered neoadjuvant (before surgery) for instance to decrease the tumor volume pre-operatively (65). The principle behind chemotherapeutics is that they target rapidly dividing cells by interrupting cell division, thereby killing the fast growing cancer cells. Several cancer

chemotherapeutic drugs induce excessive DNA damage resulting in cell-cycle arrest and cell death. However, the distinct chemotherapeutic agents have different mechanism of action (66). Although chemotherapeutics target the rapidly diving cancer cells, many therapies fail to remove all cancer cells increasing the change of relapses. Why some cells in the tumor escape the treatment have led to several proposed theories, such as the cancer stem-cell hypothesis (67). It have been suggested that genetic variation in certain clones in the tumor may provide the opportunity of some cells to escape therapy. In addition, other non-genetic mechanisms such as epigenetic changes have been implicated to be involved in the selection by therapy (68).

1.5.2 Doxorubicin

The patients included in this thesis have received neoadjuvant treatment with doxorubicin (Figure 9). Doxorubicin is a member of the anthracycline family of antibiotic chemotherapeutic agents which are considered some of the most effective anti-cancer agents used for cancer therapy (69). Doxorubicin is the most widely administered anthracycline antibiotic and is used for the treatment of solid tumors (70;71).

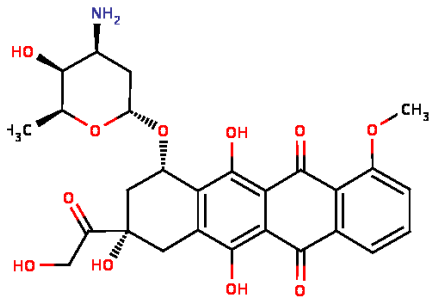

Figure 9. The chemical structure of doxorubicin. Doxorubicin is a antineoplastic antibiotic obtained from *Streptomyces peucetius* belonging to the anthracycline family of chemotherapeutics (72).

Numerous mechanisms responsible for the cytotoxic actions of doxorubicin have been proposed, and the different mechanisms have been subjected to substantial debate (71). However, the primary mechanism of action is thought to be the inhibition of topoisomerase II by the stabilization of the topoisomerase II-DNA complex (Figure 10). Topoisomerase II induces transient DNA breaks during DNA replication and other cellular processes to relax supercoiled DNA or to allow DNA strands to pass

through each other (73). When topoisomerase inhibitors are used as cancer treatment, processes such as DNA replication, transcription and repair which require changes in DNA topology are therefore interfered. The formation of double-strand DNA breaks promoted by this action leads to cell death (71;74). In addition to topoisomerase-inhibition, other mechanism including DNA intercalation, free radical formation and helicase inhibition have also been suggested to be involved in the anti-cancer actions of doxorubicin (71;74).

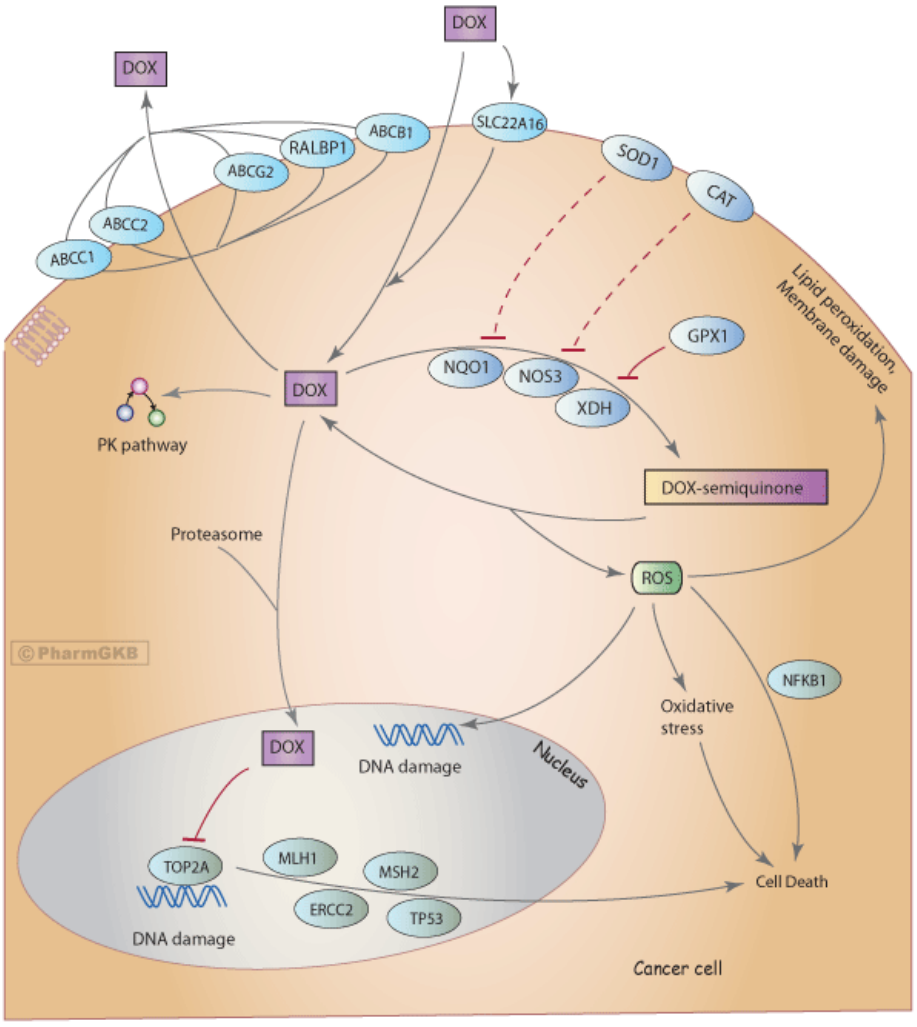

Figure 10. Doxorubicin pathway in cancer cells. Doxorubicin is transported into the cell and has several proposed mechanism of action, such as inhibition of the topoisomerase II-DNA intermediate leading to double strand breaks and generation of free radicals leading to DNA and cell membrane damage. Doxorubicin action leads ultimately to cell death which gives it an antitumor activity (75).

Cardiotoxicity and cellular resistance are the most important limitations of the clinical use of doxorubicin (69). Acute and chronic irreversible cardiotoxicity are the major adverse effect of doxorubicin therapy. Several mechanism have been implicated in the myocardial damage that may lead to congestive heart failure induced by doxorubicin (76). The mechanisms of cellular resistance to doxorubicin are thought to be different from those of the mechanisms of cardiotoxicity. Mechanisms behind resistance have been proposed to involve among other drug efflux mediated by membrane transporters such as ABCB1 and ABCG2 leading to decreased intracellular concentrations of doxorubicin (77). Mutations in the tumor suppressor protein p53, which trigger cell-cycle arrest and apoptosis in response to cellular stress signals, have also been reported to be involved in resistance to doxorubicin (78). Interindividual variation between patients in doxorubicin pharmacology including efficacy and toxicity have been hypothesized to be a result of genetic variation (70).

Increased information regarding the mechanisms of doxorubicin in addition to identification of clinical predictive biomarkers may be important to assess those patients that will benefit from doxorubicin therapy and those that will not. Since doxorubicin treatment can lead to cardiotoxicity in addition to the common side effects of chemotherapeutics such as nausea and hair loss, identification of the patients that will not benefit from the treatment regime will spear them from these unfortunate side effects. The same is true for the patients that develop chemoresistance to doxorubicin. This evaluation may lead to administration of "right drug to the right patient", which is one of the overall goals in cancer research.

2. Aim of the Study

In this study tumor specimens collected from breast cancer patients before and after neoadjuvant treatment (before surgery) with doxorubicin were analyzed. As described above doxorubicin is an anticancer anthracycline, which primary mechanism of action appears to be the inhibition of topoisomerase II in transient cleavage complex with DNA. However, the pathways underlying doxorubicin effect on tumor cells are not well understood. The overall aim in this study is to further increase our knowledge of the pathways underlying doxorubicin effect on tumor cells and to elucidate the mechanism of patients response to doxorubicin treatment, by using whole genome mRNA analysis to assess mRNA expression levels in tumor tissue before and after doxorubicin treatment. In addition, p53 has been proposed to modulate topoisomerase activity and it has been suggested that tumor cells with mutated p53 have an abnormally activated topoisomerase which may contribute to the genetic instability in tumors. Certain p53 mutations have also been connected to doxorubicin resistance. The gene expression pattern in tumor samples with mutated compared to wild type p53 was therefore examined. Some of the tumor samples included in this study have previously been classified using early cDNA microarrays containing only 8102 human genes. By utilizing mRNA arrays with a larger amount of probes and higher resolution this study also aimed to compare the previous classification with information from new microarrays.

More specifically, the aims in this study are to investigate:

1. The alterations in gene expression following treatment both at the individual gene level as well as pathway based
2. The association between mRNA expression profiles and response to therapy both at the individual gene level as well as pathway based
3. The association between mRNA expression profiles and *TP53* mutation status in the tumor samples

The main focus will be on the following topics:

- Is the gene expression profile in tumor tissue changed as a result of doxorubicin therapy? If so, which genes and pathways are implicated?
- Can changes in mRNA expression profile be connected to differences in response to doxorubicin treatment?
- Is the gene expression profile different between patients that have a mutated p53 compared to those that have wild type p53?

3. Material

The material included in this thesis was collected previously for other studies. The tumor samples were analyzed and included in a study of breast tumor classification based on mRNA expression profile using early generation cDNA expression arrays with a limited number of probes (54) in addition to a study examining the effect of *TP53* mutations on resistance to doxorubicin treatment (78).

3.1 Patient Material and Therapy Regime

The material in this thesis consists of tumor biopsies collected from breast cancer patients with locally advanced breast cancer (stage 3 and 4). The patients were between 32 and 88 years of age, with a median age of 64 years. The majority of the tumors were invasive ductal carcinomas, but some were also lobular carcinomas and tumors classified as other histological types. The tumors were graded according to the three morphological features explained in chapter 1.3.1. A total of 128 tumor specimens collected from 72 patients were included in this study. 56 of the tumors were sampled twice from the same patient, before and after doxorubicin monotherapy (14 mg/m^2) for an average of 16 weeks (range 12-23). The 'before' samples were collected prior to therapy by an open biopsy, while the 'after' samples were collected during surgery. In addition to the 56 before and after pairs collected from the same tumor, 16 'single' samples were included in this study. These samples included eight samples collected before therapy and eight samples collected after therapy from different patients. In all, the 128 samples compromised 64 before and 64 after samples. RNA from these samples was extracted using Trizol and RNeasy minikit from Qiagen.

The patients originally included in the previous studies were assessed with clinical response, according to the Union International Contre Cancer criteria, based on the alterations in tumor size after neoadjuvant doxorubicin therapy (79). Patients were categorized into the responses groups complete response (CR), partial response (PR), progressive disease (PD), stable disease and (SD) and minimal change (MC). The complete responders had disappearance of all tumor lesions while the partial responders had reduction ≥50% in the sum of all lesions. In contrast, the patients assessed with progressive disease had an increase in the diameter product of any individual tumor lesion by ≥25%, indicating a lack of sensitivity to treatment. Patients

characterized with stable disease and/or minimal change had a response between PR and PD *i.e.* reduction <50 % of all tumor lesions or increase in the diameter product of any individual tumor lesion by <25 %. The 72 patients included in this study were only assessed with partial response, stable disease and minimal change. The patients assessed with partial response were characterized as doxorubicin responders, while the patients assessed with stable disease and/or minimal change were characterized as non-responders to doxorubicin.

3.2 Clinical Data

Clinical and histopathological parameters including ER status, grade and p53 status were available for all the 128 tumor samples included in this study. The patients, which these tumor samples were collected from, were grouped according to if they obtained partial response (PR) or stable disease and/or minimal change (SD/MC) to doxorubicin treatment. The distribution of patients in the two response categories in addition to the clinical-histopatological parameters of the tumor samples are listed in Table 2. Survival data, tumor size, lymph node status and data on distant metastasis were also available but were not included in the analyses in this thesis.

Table 2. Patient characteristics. Clinical-histopatological parameters such as ER status, grade and p53 status in addition to data on response for the patients (n=128) included in this thesis. The numbers of patients in each category, in addition to the percentage are displayed.

Characteristics	No. of patients	%
ER positive	108	84.4
ER negative	20	15.6
Tumor grade 1	35	27.4
Tumor grade 2	62	48.4
Tumor grade 3	31	24.2
p53 wild type	99	77.3
p53 mutated	29	22.7
Partial response	55	43
Stable disease/minimal change	73	57

4. Methods

In this section, the different laboratory-and statistical methods used in this project are described. A table with all the reagents and equipment used are listed in Appendix A.

4.1 Microarray Technology

Microarray technology is utilized in a number of genetic experiments to analyze among other DNA sequence variation such as single nucleotide polymorphism (SNP), mRNA expression, miRNA expression, DNA methylation and protein expression. Genome-wide microarrays have made it possible to make the analysis globally. The principle behind this technology (Figure 11) is to immobilize probes that are complementary to the molecules of interest (*e.g.* DNA or RNA) on a surface. The sample molecules are labeled with substances that emit signals (*e.g.* fluorescence) followed by hybridization to the complementary probes on the array. After hybridization the array is scanned and the signals from the samples are measured. The data generated is then normalized to make the signals comparable to reveal biological differences (80). A number of different microarray platforms exist and they vary among other in genomic coverage, probe type, probe number, probe immobilization, sample hybridization and sample labeling (81).

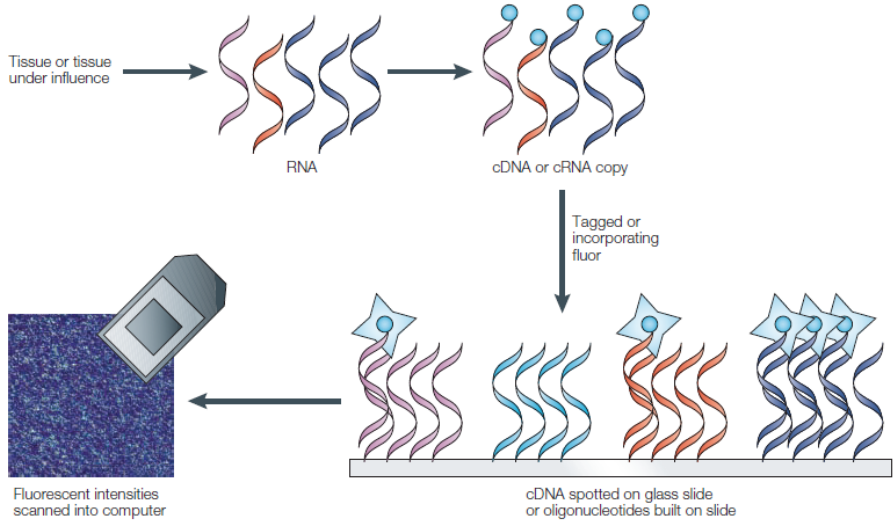

Figure 11. Schematized experimental process using microarray technology. The standard workflow for all microarray platforms measuring gene expression includes isolation of RNA or mRNA from biological specimens, generation and labeling of complementary DNA (cDNA) or complementary RNA (cRNA), hybridization to a microarray, washing of unhybridized samples and final scanning of the microarray under laser light (80).

4.2 Whole-Genome mRNA Array

The whole-genome mRNA analysis was performed according to the One-Color Microarray-Based Gene Expression Analysis Protocol from Agilent Technologies (Version 6.5, May 2010) using the Human GE 4x44K Microarray. This is a nucleic acid microarray which utilizes oligonucleotides as array elements that are attached to solid surfaces. The one-channel array provides intensity for each probe indicating a relative level of hybridization. The high throughput genome-wide array measures gene expression in experimental and control samples by the use of cyanine 3[1] (Cy3)-labeled targets (82). The procedure includes sample preparation which converts mRNA to cDNA and further to Cy3-labeled cRNA, hybridization to 60-mer oligonucleotide probes, microarray wash and final scanning and feature extraction (Figure 12) (82).

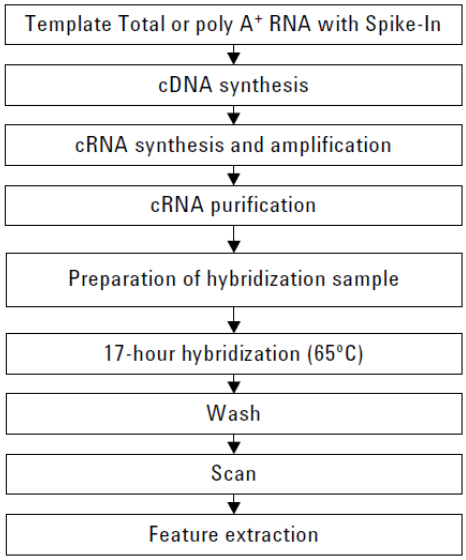

Figure 12. The standard workflow for sample preparation and array processing. Total RNA and Spike-In RNA are prepared, and the labeled cRNA generated are purified before it is hybridized to microarrays containing oligonucleotide probes. The arrays are then washed before scanning and feature extraction (82).

The Agilent One-Color Spike-In controls are required for use with the Agilent One-Color gene expression microarray workflow because gene expression microarray

[1] Cyanine 3 is a synthetic fluorescent dye which is used in microarray experiments for nucleic acid labeling.

experiments are multi-step procedures and many parameters may influence the microarray data (83). The positive controls consist of 10 *in vitro* synthesized, polyadenylated transcripts derived from the Adenovirus E1A gene which are labeled and amplified together with the experimental samples (83). The controls are optimized to anneal to complementary probes on the microarray, and after the hybridization the signal intensities for each control transcript can be used to monitor the microarray workflow from sample amplification and labeling to microarray processing performed in the experiment (83).

4.2.1 Sample preparation

In this step cDNA is synthesized from mRNA samples followed by the synthesis of fluorescent labeled cRNA (Figure 13) (82). The sample preparation process includes preparation of the one-color spike mix, preparation of the labeling reaction and the labeling reaction itself, purification of the labeled and amplified RNA and quantification of the cRNA (82).

In the labeling reaction cDNA is generated that carries a T7 promoter primer at its 5'end by using among other T7 Promoter Primer and AffinityScript Reverse Transcriptase (82). The double-stranded cDNA serves as a template for a process in which T7 RNA polymerase amplifies target material and simultaneously incorporates Cy3-labeled nucleotides (CTP) into the cRNA (82).

The amplified and labeled cRNA are purified using Qiagen's RNeasy mini spin columns and quantified using NanoDrop DN-1000 UV-VIS Spectrophotometer to ensure that the amplification is successful.

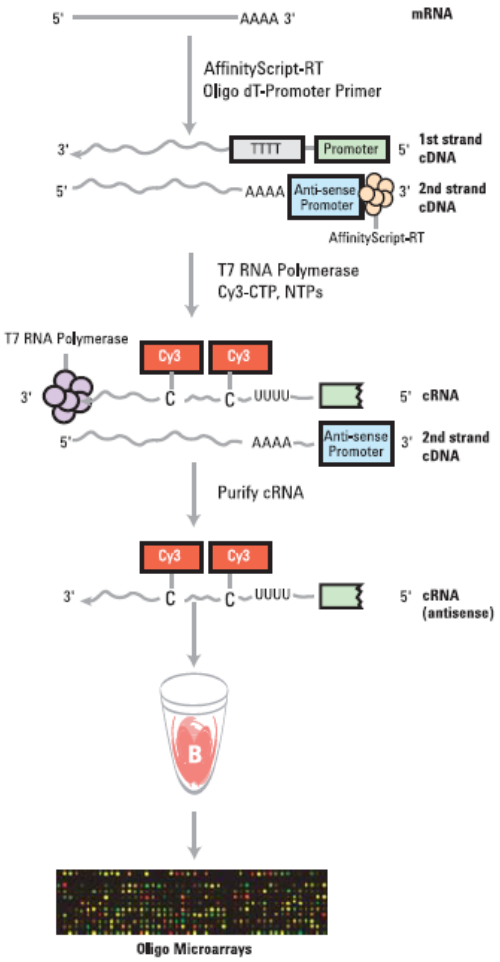

Figure 13. Generation of cRNA for one-color microarray experiment. cDNA are first generated that carries a T7 promoter at its 5'end, and the cDNA serves then as a template for a reaction in which Cy3-nucleotides are incorporated into the cRNA (82).

4.2.1.1 Procedure

Preparation of the One-Color Spike Mix:

1. Agilent One-Color Spike Mix stock solution was heated at 37 °C for 5 minutes and serial dilutions were created using Dilution Buffer appropriate for 40 ng of total RNA input amount, to make the amount of spike mix proportional to the amount of RNA input

Preparation of the labeling reaction:

Steps:

1. 2 µl of total RNA was added to 1.5 ml microcentrifuge tubes before 2 µl of diluted Spike Mix was added to each tube containing sample

2. T7 Promoter Primer Mix was prepared with 4 µl T7 Promoter Primer and 2.5 µl Nuclease-free water before 1.3 µl of the T7 Promoter Primer Mix was added to each tube containing total RNA and diluted Spike Mix
3. The primer and the template were denaturated by incubating the tubes at 65 °C for 10 minutes before the tubes were placed on ice for 5 minutes
4. 5X First Strand Buffer was prewarmed at 80 °C for 3 to 4 minutes before the cDNA Master Mix was prepared with 10 µl 5X First Strand Buffer, 5 µl DTT, 2.5 µl dNTP mix and 6 µl AffinityScript RNase Block Mix
5. 4.7 µl of cDNA Master Mix was added to each sample tube before the tubes were incubated at 40 °C for 2 hours followed by incubation at 70 °C for 15 minutes and then the tubes were placed on ice for 5 minutes
6. Transcription Master Mix was prepared with 3.75 µl Nuclease-free water, 16 µl 5X Transcription Buffer, 3 µl DTT, 5 µl NTP mix, 1.05 µl T7 RNA Polymerase Blend and 1.2 µl Cy3-CTP
7. 6 µl of Transcription Master Mix was added to each sample tube before the tubes were incubated at 40 °C for 2 hours

Purification of the labeled/amplified RNA:

Steps:

1. 84 m µl of nuclease-free water, 350 µl Buffer RLT and 250 µl ethanol (>96%) were added to the cRNA samples before 700 µl of the cRNA samples were added to RNeasy Mini Spin columns placed in 2 ml collection tubes
2. The samples were centrifuged at 4 °C for 30 seconds at 13000 rpm, the flow-through was discarded and 500 µl of buffer RPE (containing ethanol) was added to each column
3. The samples were centrifuged at 4 °C for 30 seconds at 13000 rpm, the flow-through was discarded and another 500 µl of buffer RPE (containing ethanol) was added to each column before the samples were centrifuged 4 °C for 60 seconds at 13000 rpm
4. The flow-through was discarded before the samples were centrifuged at 4 °C for 30 seconds at 13000 rpm to remove remaining traces of buffer RPE

5. The RNeasy columns were transferred to new 1.5 ml collection tubes before 30 µl nuclease-free water was added to each column to elute the cleaned cRNA samples
6. After 1 minute incubation the samples were centrifuged at 4 °C for 30 seconds at 13000 rpm before the collection tubes containing cRNA were placed on ice

Quantification of the cRNA:

Steps:

1. The cRNA samples were quantified using NanoDrop ND-1000 UV-VIS Spectrophotometer version 3.7.1 (see method chapter 4.4)
2. Cy3 dye concentration (pmol/µl), RNA absorbance ratio (260 nm/280 nm) and cRNA concentration (ng/µl) were recorded
3. The yield (µg) and specific activity (pmol Cy3 per µg cRNA) of each reaction were determined

4.2.2 Hybridization

Before hybridization RNA samples are first fragmented by a fragmentation buffer to optimize the cRNA target size (82). A blocking agent is used to minimize non-specific binding (82). A hybridization buffer is added to stop the fragmentation reaction and the samples are dispensed onto the Human GE 4x44K v2 microarray gasket slide (82). The Human GE 4x44K v2 oligo microarray slide containing 60-mer oligonucleotide probes is then placed down onto the gasket slide (82). The hybridization is prepared in Agilent microarray hybridization chamber which consists of a chamber base, a chamber cover and a clamp assembly (Figure 14) (82). The hybridization chamber is incubated in a hybridization oven (G2545A) at 65 °C and 10 rpm for 17 hours to hybridize the labeled RNA samples to the complementary probes on the microarray (82).

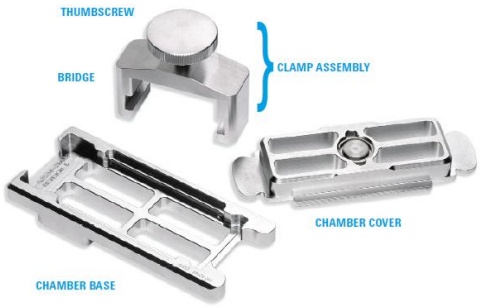

Figure 14. Agilent microarray hybridization chamber assembly. The microarray gasket slide and the oligo microarray slide are assembled in a hybridization chamber before the chamber is placed in hybridization oven to hybridize cRNA to the oligonucleotides on the oligo microarray slide (84).

4.2.2.1 Procedure

Preparation of hybridization samples:

Steps:

1. The 10X Blocking Agent was heated at 37 °C for 4 to 5 minutes
2. Fragmentation mix was made by adding 1.65 µg Cy3-labeled, amplified cRNA, 11 µl 10X Blocking Agent, Nuclease-free water (bringing the volume up to 52.8 µl) and 2.2 µl 25X Fragmentation Buffer to a 1.5 ml nuclease-free microfuge tube
3. The sample tubes were incubated at 60 °C for 30 minutes to fragment RNA before the tubes were placed on ice for 1 minute
4. 55 µl of 2x HI-RPM Hybridization Buffer was added to each sample tube to stop the fragmentation reaction before the tubes were centrifuged for 1 minute at room temperature at 13000 rpm and then placed on ice

Preparation of hybridization assembly:

Steps:

1. A clean microarray gasket slide was placed into the Agilent microarray hybridization chamber base with the barcode facing up before hybridization sample was slowly dispensed onto the gasket well
2. The oligo microarray slide, Human GE 4x44K Microarray, was slowly placed in the right orientation down onto the gasket slide before the chamber cover was placed onto the sandwiched slides and the clamp assembly was hand-tighten onto the chamber

3. The assembled chamber was vertically rotated to wet the gasket and assess the mobility of the bubbles before it was placed in a hybridization oven set to 65 °C and 10 rpm, and hybridized for 17 hours

4.2.3 Microarray wash

After the hybridization step the microarray slides are washed in Gene Expression Wash buffer 1 and 2 to remove unhybridized and non-specifically hybridized cRNA (82). Triton X-102 is added to the buffers to reduce the possibility of array wash-related background artifacts (82).

4.2.3.1 Procedure

Steps:

1. 2 ml 10 % Triton X-102 were added to wash buffer 1 and 2 in the cubitainer before Gene Expression Wash Buffer 1 was added to the disassembly dish and to the 1st wash dish
2. A slide rack was placed into the 1st wash dish before a magnetic stir bar was added and the dish was placed on a magnetic stir plate
3. Gene Expression Wash Buffer 2 was added to the 2nd wash dish, the dish was placed on a heat plate and a magnetic stir bar was added before the buffer was heated to 37 °C
4. The hybridization chamber was removed from the hybridization oven before the chamber was disassembled
5. The array-gasket sandwich was quickly transferred to the disassembly dish containing Gene Expression Wash Buffer 1 before the two slides was separated using a forceps
6. The microarray slide was placed into the slide rack in the 1st wash dish containing Gene Expression Wash Buffer 1 at room temperature for 1 minute
7. The slide rack was transferred to the 2nd wash dish containing Gene Expression Wash Buffer 2 at 37 °C for 1 minute before the slide rack was slowly removed to minimize droplets on the slides
8. A forceps was used to remove the microarray slide from the slide rack and the array slide was then placed in a light protected slide holder

4.2.4 Scanning and feature extraction

After the washing of the microarrays the slides are scanned using Agilent C Scanner (G2565CA) (82). This is a high-resolution microarray scanner that measures the fluorescence intensity of the labeled cRNA after hybridization of the samples to the microarray probes (85). The fluorescence measurement is facilitated by the use of a laser to excite the Cy3-labeled samples (85). The fluorescence from the labeled samples are detected and recorded in high-resolution TIFF images prepared for feature extraction analysis (85).

The data generated from the gene expression experiments is extracted using Agilent Feature Extraction Software (82). The software allows measurement of gene expression by extracting the information from probe features from microarray scan data (82). By the use of the Spike-In controls Agilent One Color microarray experiments can be analyzed with the Feature Extraction software which generates a Quality Control (QC) Report used for quality assessment of the microarray experiment (83). Agilent Feature Extraction Software automatically finds and places microarray grids, rejects outlier pixels, determines feature intensities and ratios, flags outlier pixels and calculates statistical confidences (86).The QC report contain statistical results and can be used to evaluate microarray performance such as the reproducibility and reliability of single microarray data (87). A QC Metric Set can be assigned the Feature Extraction (FE) Project (82), and the results appear as an evaluation table in the QC report (87). Most of the metrics included in the metric set evaluates the different laboratory steps such as the labeling, hybridization, washing and scanning (87). Thresholds are included in the metric sets and indicates if the data is in the expected range ("Good") or out of the expected range ("Evaluate") (87). The metrics can therefore determine if any problems during the experiment have arisen.

4.2.4.1 Procedure

Steps:

1. The scanner and the scanner PC was turned on and the **Agilent Scan Control** Software was opened before **Start Slot** and **End Slot** in addition to **Profile AgilentHD_GX_1Color** was selected
2. The microarray slides were inserted into slide holders and the assembled slide holders were placed in the scanner carousel, before it was verified that the

Scanner status in the main window said **Scanner Ready** and the scan was started

3. When the scanning of the microarray slides was completed, **Agilent Feature Extraction Software** was opened at the scanner PC
4. TIFF image was extracted by adding the image to be extracted to the FE Project, before setting the correct FE Project Properties and selecting the correct Extraction Set Configurations (the correct grid template and the correct protocol) before the extraction was started
5. After the extraction was completed the QC reports for each extraction set was viewed and evaluated

4.3 RNA Quality Assessment

The Agilent 6000 Nano Assay evaluates RNA quality by the use of Agilent 2100 Bioanalyzer (88). This quality characterization is important because experiments involving RNA such as genome-wide microarrays require high quality RNA. The assay contains RNA chips with an interconnected set of microchannels that is used for the separation of RNA fragments based on their size as they are driven through the channels by electrophoresis (89).

After the analysis is performed the Agilent 2100 Expert Software provides a RNA Integrity Number (RIN) for each sample (88). The RIN number is a quantitative value for RNA integrity and is ranging from 1 to 10, representing the lowest and highest RNA quality, respectively. In addition to the RIN number, an electropherogram of each sample well window is generated (88). If the total RNA run is successful two distinct peaks representing the 18S and 28S ribosomal RNA in addition to one marker peak will be displayed in the elecropherogram (Figure 15) (88).

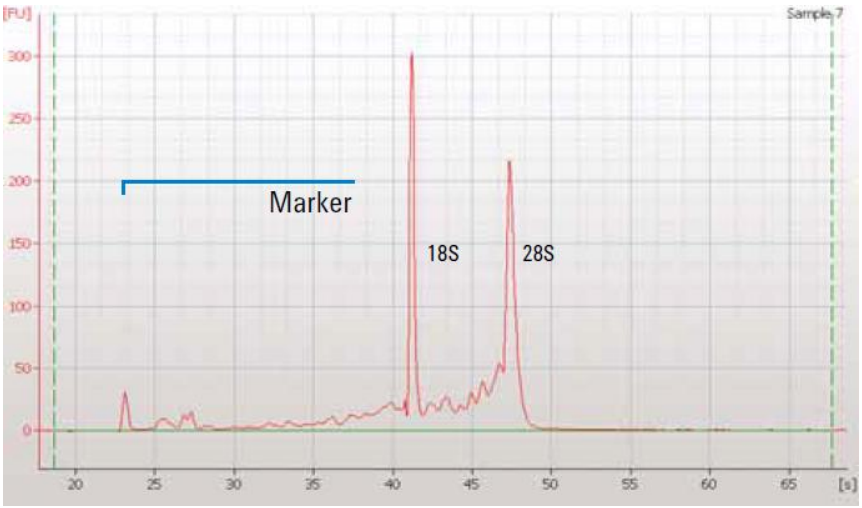

Figure 15. An electropherogram of the sample well window can be displayed after a total RNA sample run on the Agilent 2100 Bioanalyzer. One marker peak and two ribosomal peaks (18S and 28S) are the major features of a successful run (88).

4.3.3 Procedure

Steps:

1. The RNA 6000 Nano gel matrix was equilibrated to room temperature for 30 minutes before 550 µl of the matrix was dispensed to a filter spin and centrifuged at 1500 g +/- 20% for 10 minutes
2. 65 µl filtered gel was dispensed into 0.5 ml RNase-free microfuge tubes. The tubes were stored at 4ºC and must be used within one month of preparation

Preparation of the ladder:

Steps:

1. The ladder was dispensed to a 1.5 ml RNase-free microfuge tube before it was denatured at 70ºC for 2 minutes and placed directly on ice
2. Aliquots were then prepared in 0.5 ml RNase-free microfuge tubes before stored at -70ºC

Preparation of the gel-dye mix:

Steps:

1. The RNA 6000 Nano dye concentrate was equilibrated to room temperature for 30 minutes before vortexed for 10 seconds and spinned down
2. 1 µl of dye was added into a 65 µl aliquot of filtered gel and the solution was vortexed before the tube was centrifuged at 13000 g for 10 minutes, the prepared gel-dye mix had to be used within one day

Loading the gel-dye mix:

Steps:

1. A new RNA 6000 Nano chip was placed on the chip priming station before 9.0 µl of gel-dye mix equilibrated to room temperature for 30 minutes was dispensed to the well marked with a black G
2. The plugger was positioned at 1 ml before the chip priming station was closed and the plugger was pressed until it was held by the chip
3. The clip was released after 30 seconds, and after additional 5 seconds the plugger was pulled back to 1 ml position
4. The chip priming station was opened and 0.9 µl of gel-dye mix was dispensed in the two wells marked G

Loading the Agilent RNA 6000 Nano Marker:

Step:

1. 5 µl of RNA 6000 Nano marker equilibrated to room temperature for 30 minutes was dispensed in all 12 sample wells including the well marked with a ladder

Loading the ladder and samples:

Steps:

1. 1 µl of prepared ladder kept on ice was dispensed in the well marked with a ladder before 1 µl of sample was dispensed in each of the 12 sample wells after the samples had been denatured at 70ºC for 2 minutes
2. The chip was placed horizontally in the adapter of the IKA vortexer and vortexed for 1 minute at 2400 rpm
3. 350 µl RNase Zap was dispensed into an electrode cleaner and placed in the Agilent 2100 bioanalyzer for 1 minute before 350 µl RNase-free water was dispensed into another electrode cleaner and placed in the bioanalyzer for 10 seconds, then the lid was opened and the electrodes was dried for 10 seconds

Starting the Agilent 2100 Bioanalyzer:

Steps:

1. The bioanalyzer was turned on, the chip was placed in the bioanalyzer within 5 minutes after ladder and sample loading and **Agilent 2100 expert software** (Revision B.02.02 and higher) was opened before **Eukaryotic Total RNA Nano Series II** was selected at the bioanalyzer PC and the number of samples entered
2. After the analysis was completed the chip was removed before the electrode cleaner containing RNase-free water was placed in the bioanalyzer for 10 seconds

4.4 RNA Quantification with Absorbance

NanoDrop ND-1000 UV-VIS Spectrophotometer is used to measure among other the concentration of DNA and RNA and also to assess their purity (90). The spectrophotometer measures the absorbance of compounds absorbing at 230, 260 and 280 nm.

Quality assessment of template RNA and Cy-3 labeled cRNA is important for the outcome of gene expression experiments (82). The RNA should be of high quality and thus free of contaminants such as proteins, carbohydrates and organic solvents (82). In addition, the RNA should be intact and minimally degraded (82). Two ratios are used to assess high-quality RNA and include the $\mathbf{A_{260}/_{280}}$ and $\mathbf{A_{260}/A_{230}}$ ratios (82). The $\mathbf{A_{260}/_{280}}$ ratio indicates the presence of contaminating proteins and should be 1.8 to 2, indicating no protein contaminants (82). In contrast, the $\mathbf{A_{260}/A_{230}}$ ratio indicates the presence of organic compounds and cellular contaminants and should be above 2 (82).

4.4.1 Procedure

Steps:

1. The sampling arm was opened and 1.5 µl RNase-free water was applied onto the lower measurement pedestal to clean it before the arm was closed. The Nanodrop software was started and **Nucleic Acid Measurement** (for RNA and DNA) or **Microarray Measurement** (for cRNA) was selected
2. 1.5 µl distilled water was applied to initialize the instrument, and the sample type **RNA-40** was selected before 1.5 µl distilled water was used to blank the instrument
3. Then 1.5 µl of each sample was applied onto the lower measurement pedestal, and concentrations was measured (ng/µl) in addition to the $\mathbf{A_{260}/_{280}}$ and $\mathbf{A_{260}/A_{230}}$ ratios

4.5 Statistical Methods and Bioinformatics

The statistical methods and bioinformatic tools used in this thesis are described in brief below.

4.5.1 GeneSpring GX

The GeneSpring GX Software (Version 11.5.1 – Build 138755) from Agilent Technologies is used among other to pre-process and analyze mRNA expression microarray data (91). The Feature Extraction files are uploaded to GeneSpring GX and the gene expression microarray data is automatically log2 transformed to remove negative numbers and 0 (91). The log2 transformation also serves as a basic normalization of the data. Several more advanced normalization methods are available to normalize the microarray data such as quantile and percentile normalization. Normalization of the fluorescent intensities of the microarray data is performed to minimize systemic non-biological variance such as RNA quantity differences, hybridization variability between microarrays and differences between manufactured microarrays. The in between sample normalization is performed to make microarray measurements comparable and to reveal differences due to biological variation and not due to technical variation. 50% percentile normalization calculates the 50^{th} percentile for each array and aligns the arrays by their 50^{th} percentile which corresponds to the median value (91). This method assumes that the 50^{th} percentile intensity is similar for all samples. In contrast, quantile normalization assumes that the overall expression and the distribution of gene abundance is the same across arrays. When performing quantile normalization the data is therefore normalized so that the distribution of probe intensities is the same for all samples (91).

Quality control on samples can be performed in GeneSpring GX to detect outlier samples (91). Quality control on samples is assessed by Principle Component Analysis (PCA) which is a covariance analysis. Principle components are vectors that reduce the microarray data into three dimensions. The PCA components 1, 2 and 3 capture the most variance in the microarray data and are numbered according to their decreased significance (91). Detection of the major trends in the data set can be viewed in a 3D-scatter plot. The X, Y and Z axis correspond to principle component 1, 2 and 3, respectively (91). When performing PCA on samples, each sample is

plotted according to its values for the first three Principal Components. The samples having similar scores for one or more components can be considered to have similar gene expression profile. To evaluate the similarity of samples within different condition groups the samples can be colored by experimental parameters. This may also identify influent parameters involved in the experiment.

Quality control on probes is performed in GeneSpring GX by filtering the probe sets to remove low reliability probes (91). The probe sets can be filtered by expression, where a lower- and upper cutoff can be chosen. In addition, the probe sets can be filtered based on their flag values. The stringency of this filtering is adjusted by selecting specific types of flag calls and the number of samples in which the flag setting must pass (91). The default flag settings in GeneSpring GX flags the probes that are non-uniform, saturated and population outlayers as 'compromised' probes, while non-positive, non-significant and below background probes are flagged as 'not detected'.

4.5.2 Unsupervised hierarchical clustering in R

Hierarchical clustering is a method that groups both samples and genes according to the similarity in their gene expression profiles. This can be performed unsupervised which organizes the samples without any predefined parameters to identify patterns and biological mechanisms within a data set. The results of the clustering are viewed in a dendrogram. In the dendrogram the samples that have a higher similarity in their expression profiles cluster more adjacent to each other than samples that show more differences in their expression profiles. In addition, the length of the branches in the dendrogram increases as the similarity decreases. Numerous similarity metrics can be used to capture related expression values between samples by calculating the distance or the correlation, such as euclidean distance or Pearson correlation (91). The calculation of the closeness is based either on trends or magnitude. Different linkage methods are also available to link all the samples in the cluster together, *e.g.* single, average, complete and centroide. By performing average linkage the average distance between all pairs are calculated and used to cluster the samples. Hierarchical clustering can be performed among other in the open statistical system **R** (http://windowxupdate.microsoft.com). Unsupervised clustering requires normalized gene expression values for all probes for the samples you want to include in the

cluster as input data. The **R** code for hierarchical clustering using pearson distance metric and average linkage is listed below:

```
data<-read.table("allprobes_normalized.txt",
header=T,sep="\t",na.strings="NA",row.names=NULL)

data2<-data[,3:dim(data)[2]]

m<-data.matrix(data2) ; rownames(m) <- data[,1];dim(m)

c<-cor(m, method="pearson")
d<-as.dist(1-c)
hr <- hclust(d, method = "average", members=NULL)
plot(as.dendrogram(hr), edgePar=list(col=1, lwd=2), horiz=F)
```

4.5.3 4.5.8 Intrinsic subclassification

Breast tumors can be classified into distinct subclasses by a method called intrinsic subclassification. This method was developed when the five subgroups were identified based on gene expression patterns by Sørlie and colleagues (92). The genes used for the classification were selected based on their similar expression level between tumors collected before and after neoadjuvant doxorubicin therapy from the same patient. A 'intrinsic' gene list of approximately 500 genes was created that consisted of genes which were most similar expressed between tumor samples from the same patient and which in addition showed the most variation in expression among tumors from different patients. For each of the five subclasses identified, a centroid consisting of the average expression for all of the 'intrinsic' genes was created. When performing intrinsic subclassification on tumor samples today, the Pearson correlation between the expression level of the 'intrinsic' genes in each tumor sample and the five centroids are calculated. Samples are then assigned to the subtype of the centroid with the largest correlation coefficient.

4.5.4 Significance Analysis of Microarrays (SAM)

Significance Analysis of Microarrays (SAM) is a software which is used to identify significant genes in a microarray data set (93). This multiple t-test was proposed and written by coworkers at Stanford University CA. SAM requires Windows, Microsoft Excel and the latest version of **R**. SAM can be used to perform numerous statistical analyses, such as quantitative, two class (unpaired and paired), one class, multiclass, survival analysis and many more. A two class SAM analysis (either unpaired or paired) requires normalized gene expression values in addition to a variable that

divides the samples into two groups as input data. The data must be in an Excel spreadsheet in which the first row contains information about the chosen variable and the remaining row consists of gene expression data.

SAM measures the strength of the relationship between the gene expression values and the variable. Repeated permutations of the data are used to determine if any genes are differentially expressed between the two sample groups. Generally in the scientific literature the cut-off for significance is set to 5%. A p-value <0.05 indicates that the differences cannot be explained by chance. However, microarray analysis often involves many independent statistical tests performed on the same data set. False positive results are therefore expected. A parameter called delta in the SAM output window can be tuned to choose the cutoff for significance based on the false discovery rate (FDR), which corresponds to the rate of false positive results expected among the significant findings. The SAM output consists of a table with all the genes that are significant differentially expressed between the two groups chosen on the basis of the variable. Both positive and negative genes are viewed that correspond to the genes which are up- and down regulated, respectively. In addition, a fold change parameter that reflects how much the expression of each gene is different between the two groups are displayed. The gene list containing the significant genes can be used in further bioinformatic analyses. However, the gene lists often contains probes without annotations. Databases such as SOURCE (http://smd.stanford.edu/cgi-bin/source/sourceBatchSearch) contain genetic and molecular annotations and can be used for this purpose.

4.5.5 Hierarchical clustering in J-Express

J-Express (http://jexpress.bioinfo.no/site/) is a bioinformatic and statistical software used for analyses of microarray data. The software is developed by a bioinformatics group at the Department of Informatics, University of Bergen, Norway and is owned by Molmine AS. Several statistical analyses can be performed in J-Express among other hierarchical clustering. When performing hierarchical clustering in J-Express microarray data containing columns with gene identifiers, rows with sample identifiers and expression values are uploaded to the software. Next, the similarity metric and the linkage method are selected, before the hierarchical clustering is performed.

4.5.6 Ingenuity Pathway Analysis (IPA)

Ingenuity Pathway Analysis (IPA) (http://www.ingenuity.com) is a licensed software which enables pathway and gene analysis of biological data such as microarray data. The information in IPA is based on scientific literature and databases which is manually reviewed. Gene lists from *e.g.* SAM can be used as input. IPA provides information about among other networks, molecular and cellular functions, chemical and molecular interactions and disease and disorder processes for a given gene set. Many of the results in IPA are provided with a p-value, a Benjamini-Hochberg p-value which is corrected for multiple testing and a ratio which reflects the number of molecules from the input gene set that map to a specific pathway or list divided by the total numbers of molecules that are mapped by IPA to the same pathway or list.

4.5.7 Gene Set Enrichment Analysis (GSEA)

In Gene Set Enrichment Analysis (GSEA) prior-defined gene sets (*e.g.* from Gene Ontology or KEGG) are used to identify gene sets which are differentially expressed within two sample groups. GSEA can be performed in several software's, such as GeneSpring GX and J-Express. In GSEA sets of genes rather than individual genes are analyzed for differentially expression between two experimental groups. By using statistics that evaluate each gene by itself it can be difficult to identify differences in expression if the gene only shows a moderate change. Therefore, GSEA is utilized to evaluate if many genes belonging to the same gene set are changed.

4.5.8 SPSS

IMB SPSS Statistics is a software which can be utilized to perform a number of statistical tests. Cross tabulation can be performed to measure the association between two-way tables. The statistics used for measuring the relationship is among other Chi-square and Fisher's exact test. Chi-square is not recommended if more than 1/5 of the cells have less than five expected observations. In those cases the Fisher's exact test is preferred. One-Way ANOVA can also be performed in SPSS and is a parametric test which can be used for normal distributed data to compare if there is a significant difference in for instance mean expression level of a transcript between two groups. It is an extension of the two-sample t-test, and analysis of variance is used to test the hypothesis that several means are qual.

5. Results

In this study the changes in mRNA expression in relation to doxorubicin therapy (before versus after doxorubicin exposure) were investigated as well as the association between mRNA expression profile and response to doxorubicin treatment. In addition, the gene expression profiles of samples with wild type and mutated *TP53* were studied. Analyses were performed to increase the understanding of the molecular mechanisms underlying doxorubicin effect on tumors and in addition to elucidate the mechanisms behind patients' response to doxorubicin treatment.

5.1 mRNA Expression Analysis

The statistical analyses can be divided into four main parts: 1) mRNA expression in tumor tissue before versus after doxorubicin therapy, 2) mRNA expression and response to treatment, 3) mRNA expression and *TP53* mutation status and 4) molecular subclassification based on gene expression profiles. Before statistical analyses can be performed the microarray data have to be normalized and filtered. The pre-processing and quality control of the samples and probes are described below.

5.1.1 Pre-processing and quality control of microarray data

Whole-genome mRNA analysis was performed according to standard procedure (see method chapter 4.2) using the Human GE 4x44K Microarray. 2 out of 131 samples failed the technical lab quality control, leaving 129 samples for further analyses. Pre-processing of the microarray data was performed in GeneSpring GX Version 11.5.1. The microarray data were 1) log2 transformed and normalized using 50% percentile normalization, 2) filtered by expression (lower cut off 10%, upper cut off 100%) giving 26635 remaining probes and 3) filtered by flags (detected and not detected) resulting in 26270 remaining probes. The control- and spike-in probes (n=36) were deleted yielding a total of 26234 probes, that corresponded to 14046 unique genes, available for data analyses.

Principle Component Analysis (PCA) was performed to assess sample uniformity and quality. 3D-scatter plots were created on 1) normalized and unfiltered data, 2) on data filtered by expression and 3) on data filtered both by expression and by flags. By performing quality control on the original 129 experimental samples one sample (BC

40-BE) was found to be an outlier in the PCA analysis. In addition, this sample had a different distribution of signal intensities in the QC report from Feature Extraction compared to the other samples (Figure 16). When performing hierarchical clustering BC 40-BE clustered separately and based on the combination of these findings the sample was removed, leaving a total of 128 samples (56 before and after pairs and 16 'single' samples consisting of eight before and eight after samples) available for statistical-and bioinformatic analyses.

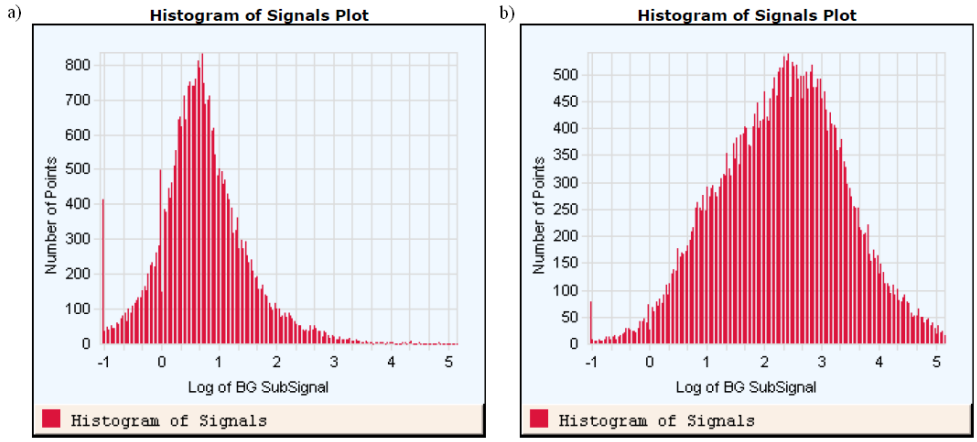

Figure 16. Probe signal distribution. The signal intensities for BC 40-BE (a) compared to another, representative sample in the experiment (b). The histogram for BC 40-BE is not as well distributed as for the other sample and show overall lower signal intensity. This sample was excluded from any further statistical analyses.

After the BC 40-BE sample was removed, new PCA plots were generated of the remaining 128 samples. One sample outlier was observed in the PCA plot created on only normalized array data (Figure 17). However, when a new plot was generated using normalized and filtered data this sample clustered together with the rest of the samples. No technical aberrations were observed for this sample and it was therefore not excluded from further statistical analyses.

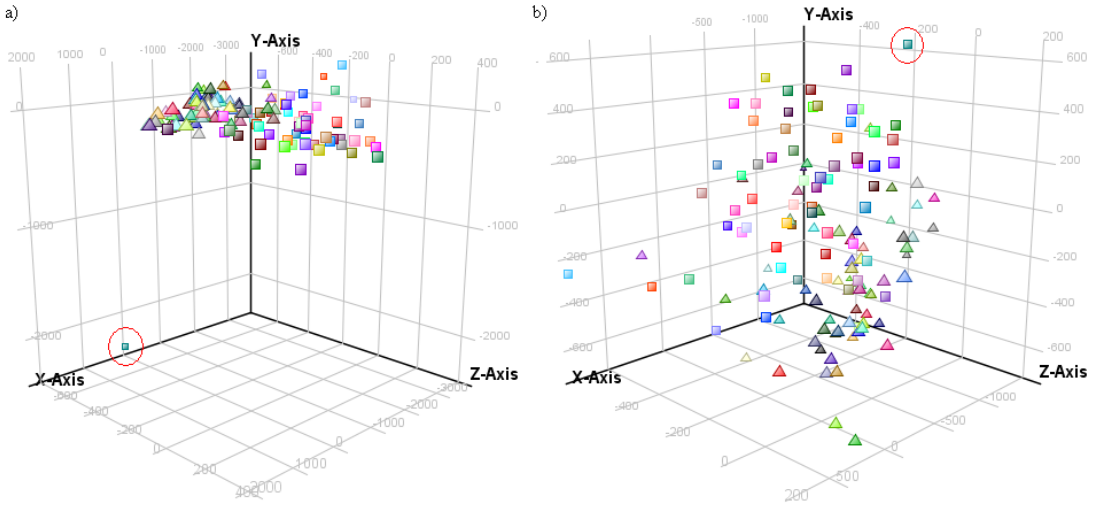

Figure 17. Principle Component Analysis (PCA). Quality control on samples was performed by PCA and the creation of 3D-scatter plots where samples are displayed in respect to the first three principle components. One sample outlier of the total 128 experimental samples is indicated with a red circle in the PCA plot created on normalized, unfiltered data (a). The PCA plots created on data filtered only by expression (data not shown) and on data filtered both by expression and by flags (b) are almost identical. The outlier sample (indicated again with a red circle) has now merged with the rest of the samples. This sample was therefore not omitted from further analyses.

5.2 mRNA Expression Profile Before and After Doxorubicin Therapy

A major part of this study was the investigation of the mRNA expression levels in breast cancer tumor tissue collected before and after doxorubicin therapy. For the samples collected before and after treatment (termed BE and AF) the analysis of the alterations in gene expression pattern may highlight the pathways induced and/or repressed by doxorubicin exposure.

5.2.1 PCA of before and after samples

A 3D PCA scatter-plot was generated of the 128 experimental samples by using the normalised and filtered probe list containing 26234 probes to evaluate the mRNA expression similarity between samples collected before (n=64) and after (n=64) doxorubicin treatment (Figure 18). The scatter plot revealed that samples collected before and after doxorubicin treatment were grouped in different areas in the plot. The before samples were mostly shifted in one area, while the after samples were mostly shifted in another indicating an overall difference between the two groups.

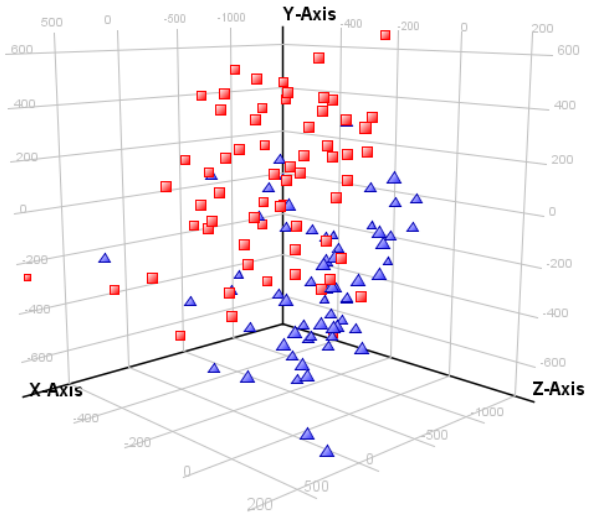

Figure 18. 3D-scatter plot of before and after samples. Before samples (n=64) are indicated in blue, while after samples (n=64) are shown in red. The PCA graph illustrates that samples within the same experimental condition *i.e.* treatment status mainly group together.

5.2.2 Unsupervised hierarchical clustering

Unsupervised hierarchical clustering by mRNA expression of the 128 tumor samples was performed in **R** version 2.14.1 (Figure 19). The normalized and filtered microarray data containing 26234 probes were used as input. Pearson correlation and average linkage were chosen as the distance metric and the linkage method, respectively. In the hierarchical cluster dendrogram the tumor samples were organized according to the similarities in their gene expression profile without any predefined parameters. The cluster analysis did not separate the before and after samples into two clear clusters. However, the subcluster that contained the majority of the samples could be divided into three main clusters (see Figure 19). Cluster 1 contained 33 after samples and 8 before samples, cluster 2 contained 31 before samples and 15 after samples while cluster 3 contained 8 after samples and 6 before samples. A significant different distribution of before and after samples between the three main clusters indicated in Figure 19 was observed using Pearson Chi-Square test (p-value=2.44E-5). The two ends on each side of the three main clusters contained both a majority of after samples.

The cluster analysis separated the majority of the before and after pairs collected from the same patient. Of the 56 pairs in the sample set, 18 of the pairs clustered together

while the rest of the pairs were separated. Nine of the 18 pairs that clustered together were collected from patients characterized as having a stable disease and/or minimal change (SD/MC) to doxorubicin therapy, and nine pairs were collected from patients assessed with a partial response (PR) to therapy. The majority of the pairs found to cluster together clustered in the ‘before’ cluster/cluster 2 in Figure 19.

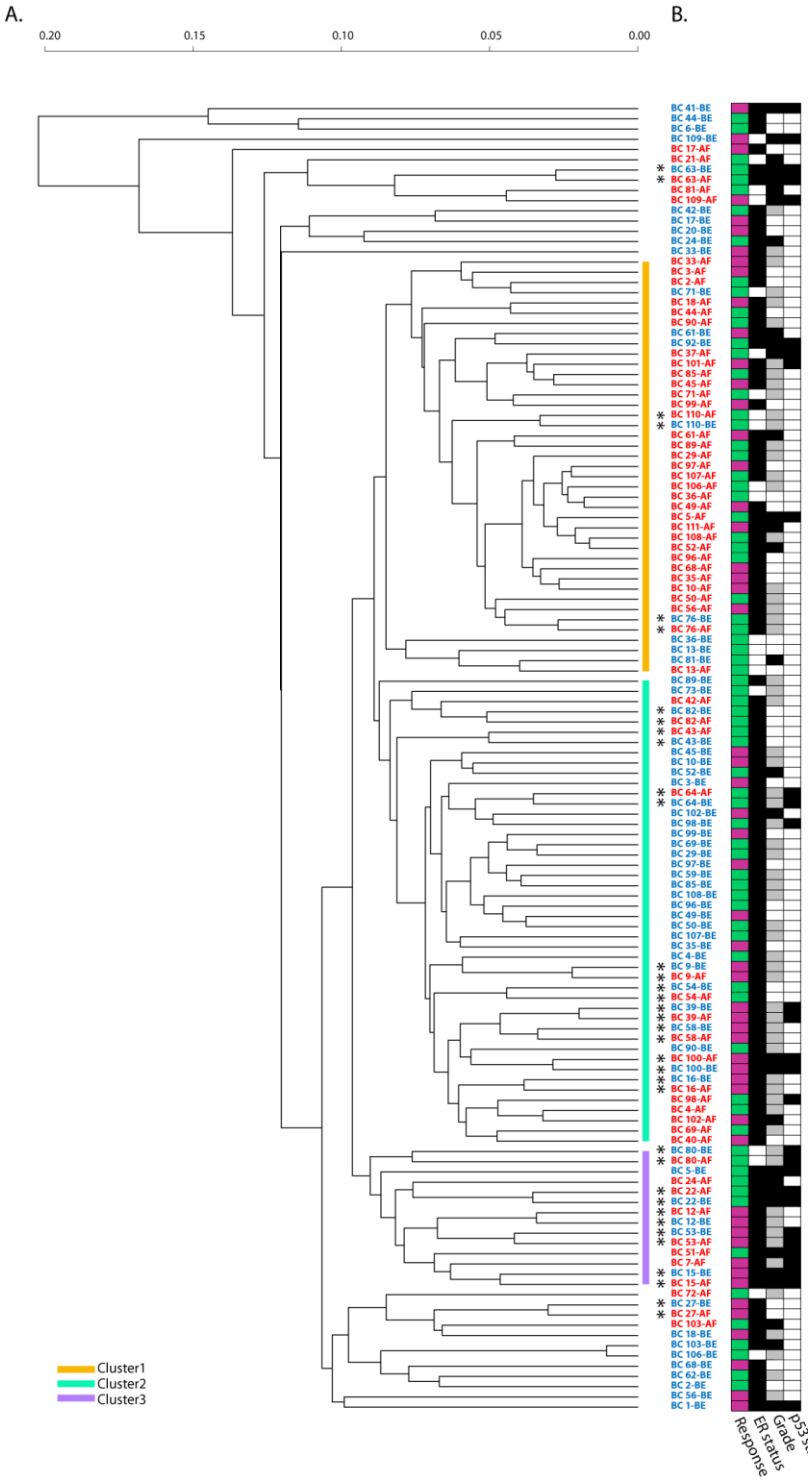

Figure 19. Unsupervised hierarchical clustering by gene expression. (A) The 128 samples were organized according to similarity in their gene expression patterns using Pearson distance metric, average linkage and 26234 probes. Samples collected before and after doxorubicin therapy are assigned BE and AF at the end of each sample name and colored blue and red, respectively. An asterisk (*) beneath the dendrogram indicates samples collected before and after doxorubicin therapy from the same patient clustering together. (B) Patient response to doxorubicin and primary tumor characteristics are indicated: response: partial response (purple), stable disease and/or minimal change (green); estrogen receptor status: ER negative (white), ER positive (black); histological grade: grade 1 (white), grade 2 (grey), grade 3 (black); *TP53* mutation status: wild type (white), mutated (black).

5.2.3 SAM analysis of tumor samples before and after therapy

SAM analysis was performed on the normalized and filtered mRNA expression data to detect genes differentially expressed between samples collected before and after doxorubicin treatment. The analysis was performed both on the total number of experimental samples (n=128, 64 before and 64 after) and on the subset of sample pairs collected before and after doxorubicin therapy from the same patient (n=112 representing 56 pairs). SAM analysis on the expression data from all samples increases theoretically the power to detect differences because more samples are included in the analysis, while the removal of the 16 'single' samples may give a "cleaner" result of the differences between the samples collected before and after doxorubicin therapy, through the embedded correction for intratumor differences.

SAM analysis performed on microarray data from all the 128 experimental samples detected 6387 probes to be significantly differentially expressed between the before and after samples (FDR≈0%) (Figure 20). In all, 2889 probes were found to be up regulated while 3498 probes were found to be down regulated in the after samples.

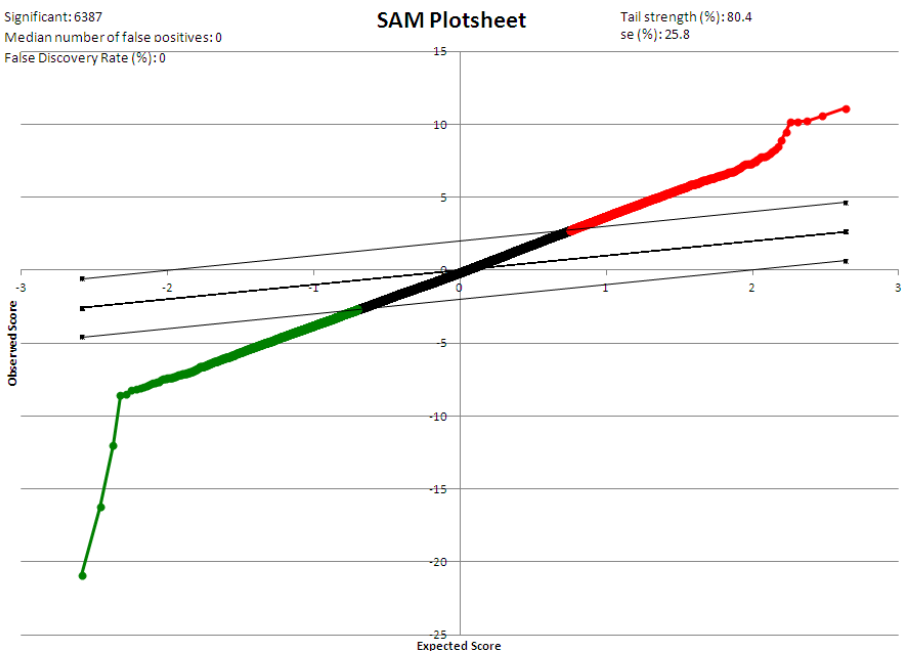

Figure 20. SAM analysis, all samples. Two class unpaired analysis revealed 6387 probes to be differentially expressed between before and after samples (FDR≈0%). The red (n=2889) and the green (n=3498) probes are up- and down regulated respectively in the after samples.

SAM analysis performed on mRNA expression data from the 56 before and after pairs revealed 6482 probes to be significant differentially expressed between the two groups (FDR≈0) (Figure 21). A total of 3040 up regulated and 3442 down regulated probes were revealed in the after samples.

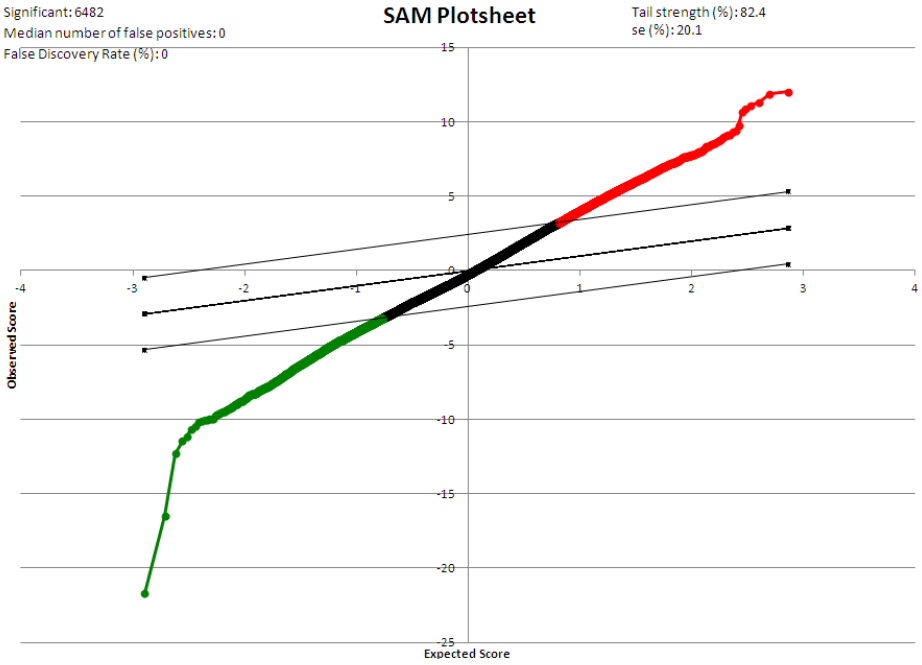

Figure 21. SAM analysis, paired samples. SAM analysis performed on the before and after pairs (n=56 pairs, 112 samples). Two class paired analysis explored 6482 probes to be significant differentially expressed between the before and after samples (FDR≈0%). The red (n=3040) and the green (n=3442) probes are up- and down regulated in the after samples respectively.

Comparing the fold change of the genes from both the unpaired and paired SAM analysis revealed a much higher fold change for the genes up regulated after therapy than for the genes down regulated after therapy. The genes showing the largest fold change in either direction, *i.e.* top 25%, from both the unpaired and the paired SAM analysis were further analyzed using Ingenuity Pathway Analysis (IPA). Selected genes that were shown to be significant differentially expressed by SAM analysis, and that were found in the list containing genes having the 25% fold change, were further investigated using IMB SPSS Statistics 18. This analysis was performed to confirm to which extent the genes were differentially expressed between the before and after samples by using One-Way ANOVA. Two of the up regulated (*FOS*, *CYR61*) and two of the down regulated (*FBXO11*, *NIT2*) genes from the unpaired analysis were

visualized by ANOVA plot (Figure 22). The mean mRNA expression level for the four genes in the before and after samples is illustrated in Figure 22. By using One-Way ANOVA all the genes were found to have a significant difference in means ($p<0.05$) between the two groups.

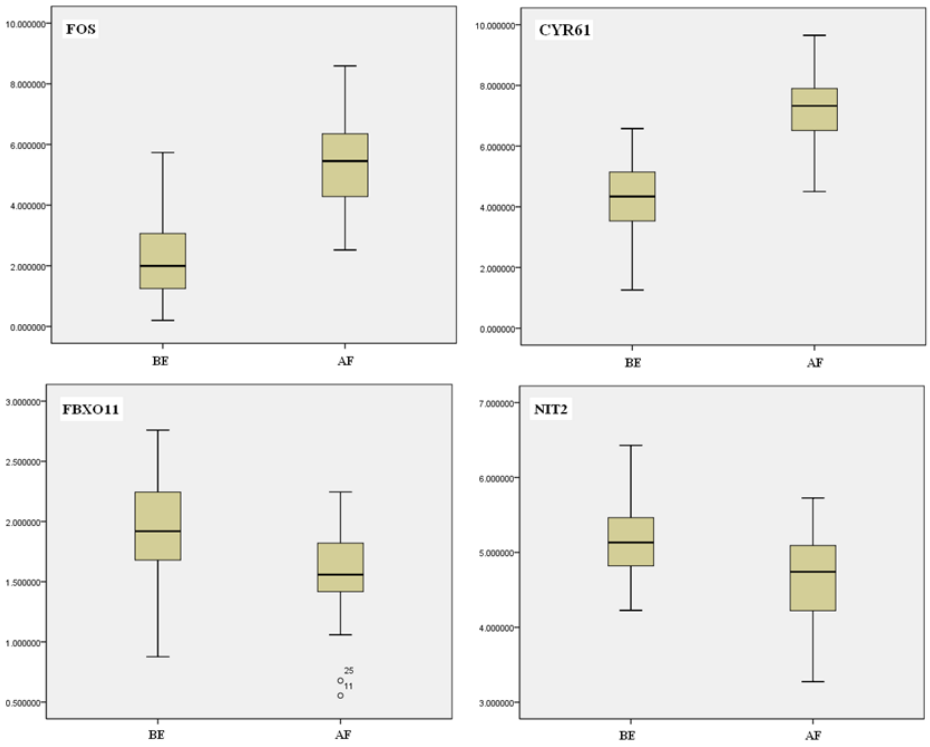

Figure 22. The distribution of mRNA expression level of selected genes in before and after samples. *FOS* and *CYR61* were found to be up regulated in the after samples, while *FBXO11* and *NIT2* were found to be down regulated in the after samples by SAM analysis. By performing One-Way ANOVA the four genes were shown to have a significant difference in means between the two groups (p-value<0.05). The distribution of expression levels between the before (BE) and the after (AF) samples is visualized in the box plot. The box plot shows the 75th (upper edge of box), 50th (line in the box) and 25th (bottom edge) percentile while the whiskers gives the 1.5 interquartile range. Outliers are shown with small black circles.

The probes found to be differentially expressed between samples collected before and after treatment from the unpaired SAM analysis (n=6387) were used to cluster the 128 samples (Figure 23). In this cluster analysis, only four pairs were found to cluster together indicating a greater separation of the before and after samples collected from the same patient compared to using all probes in the microarray data (see chapter 5.2.2).

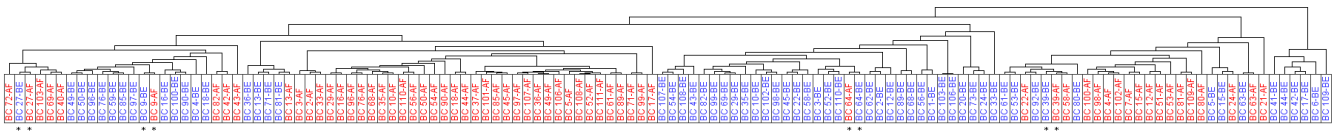

Figure 23. Hierarchical clustering using only the probes found to be differentially expressed between before and after samples from SAM. Samples collected before doxorubicin treatment are shown in blue and samples collected after treatment are shown in red. An asterisk (*) beneath the dendrogram indicates before and after pairs from the same patient that clustered together.

5.2.4 Ingenuity Pathway Analysis of genes differentially expressed between before and after samples

The probe lists containing the genes differentially expressed between before and after samples from both the unpaired and the paired SAM analysis were extracted and further analyzed using Ingenuity Pathway Analysis (IPA). Before the analyses were performed, the probe lists from SAM were pre-processed. First, the gene names of the un-annotated probes were updated where possible using SOURCE, before the remaining un-annotated probes were deleted. Probes with the top 25% fold change were extracted and the redundant genes removed before uploading the list of genes to IPA. A small subset of genes was not recognized in IPA and these genes were therefore excluded in the analyses. The number of probes and unique genes up- and down regulated, after doxorubicin therapy, in both the unpaired and paired analysis are shown in table 3.

Table 3. Overview of probes from the SAM analyses and genes included in IPA analyses. The probes up- and down regulated after doxorubicin treatment from the unpaired and paired SAM analysis are shown. Before IPA analyses were performed, the un-annotated probes were deleted and probes with the top 25% fold change extracted. The unique genes these probes represented were uploaded to IPA. The last row in the table specifies the total number of genes included in the IPA analysis.

	Unpaired		***Paired***	
	Up regulated	Down regulated	Up regulated	Down regulated
Probes from SAM	2889	3498	3040	3442
Probes after deletions	2633	2414	2770	2392
Probes with top 25% fold change	658	603	692	598
Unique genes	573	559	603	557
Genes included in IPA analysis	567	548	594	536

A total of 534 genes were found to be identical between the up regulated genes, having the top 25% fold change, from the unpaired and paired comparison of before and after samples. For the unpaired list, 39 genes were unique compared to 69 unique genes for the paired gene list. Likewise, 369 genes were found to be equal between the genes down regulated, with the top 25% fold change, after doxorubicin treatment from the unpaired and paired SAM analysis. In addition, 190 genes were found exclusively in the unpaired list while the paired list contained additional 188 unique genes. Venn diagrams illustrating the gene similarity between the up- and down regulated genes from the unpaired and paired analysis are shown in Figure 24.

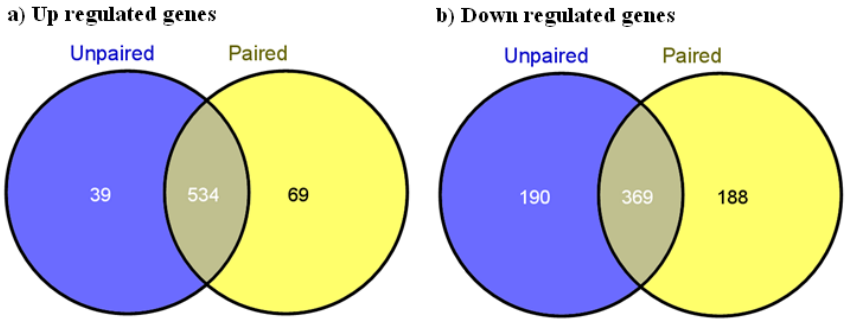

Figure 24. Venn diagrams of genes, having the top 25% fold change, from the unpaired and paired comparison of before and after samples. The Venn diagram illustrates that 534 genes are equal between the unpaired and paired analysis of genes up regulated after doxorubicin therapy (a). A total of 369 down regulated genes were found to be equal between the two gene lists containing genes down regulated after treatment (b).

For both the up- and down regulated gene lists, the genes showing the largest fold change, *i.e.* top 25%, were uploaded to IPA. IPA of genes up regulated after treatment with doxorubicin are described first, followed by IPA of the genes down regulated after therapy.

5.2.4.1 IPA on genes up regulated after doxorubicin treatment

The genes up regulated, showing the top 25% fold change, after doxorubicin therapy provided from the unpaired (n=567) and paired (n=594) before and after sample comparison were uploaded to IPA to identify functions and pathways enriched within the lists. Tables containing the molecular and cellular functions, canonical networks, transcription factors and tox lists overrepresented within the gene lists are listed below (Table 4-7).

The IPA analysis revealed many molecular and cellular functions overrepresented in the list of genes with top 25% fold change up regulated after treatment for both the unpaired and paired gene list. To identify the most significant functions a Benjamini-Hochberg threshold value was selected. A threshold value of 0.05 revealed 65 enriched functions in both the unpaired and paired gene lists. To reduce this number, a threshold value of 1.0E-13 was selected for both analyses, and revealed a total of six and eight functions significantly overrepresented in the list of genes up regulated from the unpaired and paired analysis, respectively. The IPA analyses indicated that the molecular and cellular functions overrepresented in the list of genes up regulated after treatment were almost identical between the gene lists provided from comparison of all and paired before and after samples (Table 4).

Table 4. The most significant molecular and cellular functions enriched in the list of unique genes, with top 25% fold change, up regulated after doxorubicin treatment. The functions displayed are provided from the comparison of unpaired (a) and paired (b) before and after samples. The Benjamini-Hochberg multiple testing correction p-value (B-H p-value) and the number of molecules in the uploaded data set associated with each specific function are displayed.

a) Molecular and Cellular Functions overrepresented in the list of unique genes up regulated after therapy, unpaired analysis (n=567)		
Name	*B-H P-value*	*# molecules*
Cancer	9.10E-31 – 3.40E-03	266
Reproductive System Disease	9.10E-31 – 6.85E-04	183
Cardiovascular System Development and Function	4.53E-24 – 3.43E-03	155
Organismal Development	4.52E-24 – 3.08E-03	182
Cellular Movement	5.42E-21 – 3.40E-03	160
Cell Morphology	2.49E-14 – 3.40E-03	180
b) Molecular and Cellular Functions overrepresented in the list of unique genes up regulated after therapy, paired analysis (n=594)		
Name	*B-H P-value*	*# molecules*
Cancer	2.90E-33 – 3.09E-03	282
Reproductive System Disease	2.90E-33 – 9.74E-04	192
Cardiovascular System Development and Function	3.23E-25 – 2.86E-03	164
Organismal Development	3.23E-25 – 2.92E-03	204
Cellular Movement	6.98E-23 – 2.86E-03	183
Cellular Growth and Proliferation	3.55E-15 – 2.71E-03	229
Tissue Development	4.52E-15 – 2.92E-03	242
Cell Morphology	3.65E-14 – 1.90E-03	193

The top canonical pathways overrepresented in the list of genes having the top 25% fold change up regulated after therapy, extracted from the SAM analysis of both unpaired and paired samples are shown in Table 5. A Benjamini-Hochberg threshold value was selected to detect the most significantly enriched pathways within the gene lists. A total of ten pathways were discovered to be overrepresented in the list of up regulated genes from the unpaired list when a threshold value was set to 0.05. Additionally, the same threshold value identified nine pathways enriched in the list of genes up regulated from the paired analysis. The analysis revealed that the enriched canonical pathways were strongly overlapping within both gene lists.

Table 5. The most significant canonical pathways overrepresented in the list of genes up regulated after therapy, with the top 25% fold change. The pathways are extracted from the SAM analysis of both the unpaired (a) and the paired (b) before and after samples. For each canonical pathway the Benjamini-Hochberg multiple testing correction p-value and the ratio (genes included in the imported data set/ the total number of genes in the canonical pathway) are provided.

a) Top Canonical Pathways overrepresented in the list of unique genes up regulated after therapy, unpaired analysis (n=567)		
Name	*B-H P-value*	*Ratio*
Hepatic Fibrosis/Hepatic Stellate Cell Activation	3.66E-05	20/142
Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis	2.39E-03	22/229
Human Embryonic Stem Cell Pluripotency	5.58E-03	15/144
Complement System	8.73E-03	7/34
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	8.73E-03	24/324
Atherosclerosis Signaling	1.33E-02	13/125
Wnt/β-catenin Signaling	1.7E-02	16/171
Leukocyte Extravasation Signaling	4.09E-02	16/192
Glioblastoma Multiforme Signaling	4.09E-02	14/162
PTEN Signaling	4.69E-02	11/121
b) Top Canonical Pathways overrepresented in the list of unique genes up regulated after therapy, paired analysis (n=594)		
Name	*B-H P-value*	*Ratio*
Hepatic Fibrosis/Hepatic Stellate Cell Activation	4.25E-06	22/142
Human Embryonic Stem Cell Pluripotency	2.59E-04	18/144
Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis	4.04E-03	22/229
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	4.17E-03	26/324
Caveolar-mediated Endocytosis Signaling	5.2E-03	11/83
Leukocyte Extravasation Signaling	5.2E-03	19/192
Atherosclerosis Signaling	5.66E-03	14/125
Complement System	7.42E-03	7/34
Wnt/β-catenin Signaling	8.7E-03	17/171

IPA also identifies transcription factors overrepresented as gene expression regulators within the list of genes analyzed. The top five transcription factors enriched as gene expression regulators of the list of up regulated genes, with the top 25% fold change, in the unpaired and paired analysis are listed in Table 6. The transcription factor that regulated the highest number of target molecules in both the unpaired and paired gene list was found to be p53.

Table 6. The top five transcription factors overrepresented in the list of unique genes, with the top 25% fold change, up regulated after therapy discovered by the comparison of unpaired (a) and paired (b) before and after samples. The p-value and the number of target molecules for each transcription factor are shown. The p-values are not corrected for multiple testing.

a) ***Unpaired***			***b)*** ***Paired***		
Top Transcription Factors overrepresented in the list of genes up regulated after therapy (n=567)			**Top Transcription Factors overrepresented in the list of genes up regulated after therapy (n=594)**		
Transcription Regulator	*P-value of overlap*	*# targets*	*Transcription Regulator*	*P-value of overlap*	*# targets*
TP53	9.53E-11	83	KLF2	6.13E-13	29
SP1	1.21E-10	49	SP1	1.03E-11	53
KLF2	2.00E-10	25	SP3	1.24E-11	32
SP3	3.09E-10	29	TP53	7.61E-11	87
FOXL2	3.93E-09	15	SMAD3	2.32E-10	29

The most significant tox lists enriched in the list of up regulated genes, having the top 25% fold change, identified from the comparison of all before and after sample and from only the paired samples are shown in Table 7. A Benjamini-Hochberg threshold set to 0.05 identified nine significant enriched tox lists in the list of up regulated genes from the unpaired analysis. A total of eleven significant tox lists overrepresented in the list of genes from the paired analysis were revealed with the same threshold value. The tox lists enriched in the up regulated genes from both the unpaired and paired analysis contained tox functions involved in cardiotoxicity, hepatotoxicity and nephrotoxicity.

Table 7. The most significant tox lists overrepresented within the list of genes up regulated, having the top 25% fold change, in the unpaired (a) and paired (b) before and after sample analysis. For each tox list the Benjamini-Hochberg multiple testing correction p-value and the ratio (genes included in the imported data set/ the total number of genes in the tox list) are provided.

a) Top Tox Lists overrepresented in the list of genes up regulated after therapy, unpaired analysis (n=567)		
Name	*B-H P-value*	*Ratio*
Acute Renal Failure Panel (Rat)	1.6E-08	16/62
Cardiac Hypertrophy	1.05E-06	33/318
Hepatic Fibrosis	8.36E-06	15/84
Hepatic Stellate Cell Activation	5.48E-04	8/35
Liver Proliferation	6.13E-04	18/170
Persistent Renal Ischemia-Reperfusion Injury (Mouse)	3.47E-03	6/30
Increases Renal Proliferation	8.5E-03	8/56
Cardiac Fibrosis	1.09E-02	12/124
Cardiac Necrosis/Cell Death	3.56E-02	14/179
b) Top Tox Lists overrepresented in the list of genes up regulated after therapy, paired analysis (n=594)		
Name	*B-H P-value*	*Ratio*
Acute Renal Failure Panel (Rat)	2.93E-11	19/62
Cardiac Hypertrophy	8.66E-09	38/318
Hepatic Fibrosis	3E-06	16/84
Persistent Renal Ischemia-Reperfusion Injury (Mouse)	6.66E-04	7/30
Hepatic Stellate Cell Activation	6.66E-04	8/35
Increases Renal Proliferation	2.91E-03	9/56
Liver Proliferation	2.91E-03	17/170
Cardiac Fibrosis	5.63E-03	13/124
Cardiac Necrosis/Cell Death	2.39E-02	15/179
Renal Glomerulus Panel (Human)	2.54E-02	4/17
Oxidative Stress	3.74E-02	7/57

5.2.4.2 IPA on genes down regulated after doxorubicin treatment

Genes down regulated after doxorubicin therapy, having the top 25% fold change, extracted from the SAM analysis of all before and after samples (n=548) and of paired before and after samples (n=536), were further investigated in IPA. The IPA analyses were performed to reveal information of the given gene sets, including molecular and cellular functions, top canonical pathways, top transcription factors and top tox lists overrepresented in the genes down regulated after treatment.

A Benjamini-Hocherg threshold value was selected to detect the most significant molecular and cellular functions enriched in the list of genes, with the top 25% fold change, down regulated after therapy. When the Benjamini-Hochberg threshold value

was set to 0.05, only one molecular and cellular function was identified overrepresented in the list of genes down regulated after doxorubicin therapy from the unpaired analysis. Two molecular and cellular functions were identified enriched in the list of down regulated genes from the paired analysis with the same threshold. The functions are displayed in Table 8.

Table 8. The most significant molecular and cellular functions overrepresented in the list of genes, with the top 25% fold change, down regulated after doxorubicin treatment provided from the comparison of unpaired (a) and paired (b) before and after samples. The Benjamini-Hochberg multiple testing correction p-value and the number of molecules in the data set associated with each specific function are displayed.

a) Molecular and Cellular Function overrepresented in the list of genes down regulated after therapy, unpaired analysis (n=548)		
Name	*B-H P-value*	*# molecules*
RNA Post-Transcriptional Modification	4.94E-03 – 381E-01	25
b) Molecular and Cellular Functions overrepresented in the list of genes down regulated after therapy, paired analysis (n=536)		
Name	*B-H P-value*	*# molecules*
RNA Post-Transcriptional Modification	8.00-06 – 3.60E-01	29
Cancer	3.51E-02 – 3.60E-01	46

Only one canonical pathway was revealed by IPA analysis to be enriched in the list of genes, with the top 25% fold change, down regulated after doxorubicin treatment obtained from the comparison of unpaired before and after samples. This pathway was “EIF2 Signaling” with a Benjamini-Hochberg multiple testing correction p-value (B-H p-value) of 7.3E-05 and a ratio of 19/193 (genes included in the imported data set/ the total number of genes in the canonical pathway). The same canonical pathway was found to be enriched in the list of genes down regulated after doxorubicin therapy, having the top 25% fold change, from the paired analysis. In the paired analysis “EIF2 Signaling” had a Benjamini-Hochberg p-vale of 1.84E-04 and a ratio of 18/193.

The transcription factors overrepresented as gene expression regulators within the list of down regulated genes, with the top 25% fold change, from the unpaired and paired list are shown in Table 9. Some of the transcription factors were found to overlap between the unpaired and paired analysis.

Table 9. Top five transcription factors overrepresented in the list of genes, having the top 25% fold change, down regulated after doxorubicin treatment discovered by the comparison of unpaired (a) and paired (b) before and after samples. For each transcription factor the p-value and number of target molecules are displayed. The p-values are not corrected for multiple testing.

a) Unpaired			***b) Paired***		
Top Transcription Factors overrepresented in the list of genes up regulated after therapy (n=567)			**Top Transcription Factors overrepresented in the list of genes up regulated after therapy (n=594)**		
Transcription Regulator	*P-value of overlap*	*# targets*	*Transcription Regulator*	*P-value of overlap*	*# targets*
TP53	9.53E-11	83	KLF2	6.13E-13	29
SP1	1.21E-10	49	SP1	1.03E-11	53
KLF2	2.00E-10	25	SP3	1.24E-11	32
SP3	3.09E-10	29	TP53	7.61E-11	87
FOXL2	3.93E-09	15	SMAD3	2.32E-10	29

No tox lists were identified as being significantly overrepresented within the list of genes down regulated after doxorubicin treatment obtained from the before and after comparison of either the unpaired or paired samples.

5.3 mRNA Expression Profile and Response to Doxorubicin

The patients, in which the tumor samples included in this thesis were collected from, were evaluated with three different responses to doxorubicin. The patients who obtained a stable disease and/or minimal change were categorized in the same group as non-responders (SD/MC). The patients which obtained a partial response were categorized as responders (PR).

5.3.1 The association between clinical factors and response

To investigate the relationship between treatment response and different clinical parameters such as ER status, grade and *TP53* mutation status cross tabulation was performed in IMB SPSS Statistics 18. The analysis revealed a significant association between ER status of the samples collected after doxorubicin therapy and response (Fisher's Exact 2-sided p-value=0.017). No significant association was found between grade or *TP53* mutation status and response.

5.3.2 PCA of two different response groups

PCA analysis was performed to investigate if it were possible to separate the two different response groups based on their overall gene expression profiles (Figure 25). The normalized and filtered probe list containing 26234 probes was used in the analysis. The PCA scatter plot illustrated that samples collected from doxorubicin responders (n=55) and non-responders (n=73) clustered together. This may indicate that the expression profiles of the samples in the two response groups are similar.

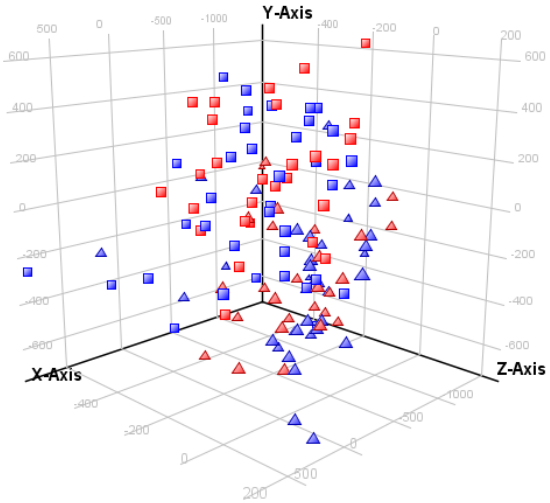

Figure 25. 3D-scatter plot of the samples collected both before and after treatment from patients with different response to doxorubicin. Samples collected from patients assessed with stable disease and/or minimal change (SD/MC) (n=73) are shown in blue, while samples taken from patients assessed with partial response (PR) (n=55) are indicated in red. The PCA plot does not show an apparent separation of the two groups.

5.3.3 SAM analysis of tumor samples from responders and non-responders

To identify genes differentially expressed between doxorubicin responders and non-responders SAM analysis was performed. The normalized and filtered mRNA expression data were used as input. SAM was performed on the before and after samples separately, in addition to the delta values for each of the before and after pairs (the differences in expression between the before and the after sample for a given patient, calculated by extracting the expression value of the after sample from the expression value of the before sample for each pair). The analyses of before and after samples were performed to evaluate if the expression profiles were predictive for response. The analysis of delta values was performed to assess if the two response groups had different capability to induce or repress the expression of genes. The

results from the three SAM analyses of responders and non-responders revealed that no genes were significantly differentially expressed between the patients who obtained a stable disease and/or minimal change and the patients who achieved a partial response to doxorubicin treatment. Another approach was then performed by using only the probes which were shown to be differentially expressed between the before and after samples revealed by SAM analysis (described in chapter 5.2.3). SAM analyses were performed on before samples, after samples and delta expression values utilizing the 6482 probes found to be differentially expressed between the paired before and after sample analysis. However, no significant genes were revealed to be differentially expressed between the two response groups.

5.3.4 Gene Set Enrichment Analysis (GSEA)

Since no individual genes were found to be differentially expressed by the comparison of doxorubicin responders and non-responders, Gene Set Enrichment Analysis (GSEA) was performed. This method analyses sets of genes rather than looking at each gene by itself. GSEA was performed in GeneSpring GX Version 12 by using the normalized and filtered microarray data. The GSEA did not reveal any significant gene sets to be differentially expressed between the doxorubicin responders and non-responders.

5.3.5 Statistical testing of p53 target genes

The IPA analysis of genes differentially expressed between samples collected before and after doxorubicin treatment revealed p53 as the transcription factor regulating the highest number of target genes with the top 25% fold change up regulated after treatment (see chapter 5.2.4.1). Given that p53 status has previously been associated to doxorubicin action and resistance, SAM analysis was performed using the p53 target genes (n=87) from the paired before and after sample analysis to evaluate if these genes could be connected to doxorubicin response. Analyses of before (n=64) and after (n=64) samples in addition to delta expression values from the before and after pairs (n=56) uncovered several significantly differentially expressed genes in samples collected from patients assessed with partial response to doxorubicin compared to those with stable disease and/or minimal change. A total of six (FDR≈0%) and 30 (FDR=3-4%) significant up regulated genes were found in the samples collected from the responders using before and after samples, respectively.

By using delta values seven genes (FDR≈0%) were revealed to have a higher delta in the doxorubicin responders representing a bigger change in expression following treatment. The genes from the SAM analyses of before samples and delta values had a FDR≈0%, while the genes from the analysis of after samples had a higher FDR. Hence only the genes revealed to be up regulated in samples collected from patients assessed as doxorubicin responders from the SAM analysis of before samples, and the genes that were shown to have a higher delta in samples from patients with partial response by using delta values are shown in Table 10.

Table 10. p53 target genes differentially expressed between samples collected from patients characterized as doxorubicin responders (PR) and non-responders (SD/MC). Six genes (FDR≈0%) were found to be significant up regulated in before samples from patients with a partial response compared to those with stable disease and/or minimal change (a). When applying the delta expression values seven genes (FDR≈0%) were revealed to have a higher delta in samples collected from patients assessed as doxorubicin responders (b). The average expression values and delta values in addition to the absolute difference in expression and delta values of the samples within the two response groups are shown. The list is sorted according to gene name.

a) Before samples			
Gene name	*Average expression, PR*	*Average expression, SD/MC*	*Absolute difference in expression*
CDC42EP3	1.59	1.23	0.36
CTGF	2.75	2.11	0.64
DKK3	4.73	4.3	0.43
ITGB4	3.04	2.63	0.41
JUNB	5.54	5.26	0.28
PHILDA1	-0.79	-1.16	0.37
b) Delta expression values			
Gene name	*Average delta, PR*	*Average delta, SD/MC*	*Absolute difference in delta*
AKAP12	0.58	-0.52	1.1
COL14A1	2.18	0.94	1.24
DUSP1	2.57	1.74	0.83
EGR1	3.3	2.53	0.77
HIC1	1.46	0.85	0.61
IGF1	2.32	1	1.32
MMPP23B	2.18	0.94	1.24

A graphical illustration of the relationship between the p53 target molecules were generated in IPA and are shown in Figure 26. All direct relationships between the uploaded genes observed in humans and cancer specifically were selected in IPA. IPA analysis also revealed three canonical pathways enriched, "PTEN Signaling", "IGF-1 Signaling" and "HER-2 Signaling in Breast Cancer", within the list of p53 target molecules when a Benjamini-Hocherg threshold value was set to 0.001.

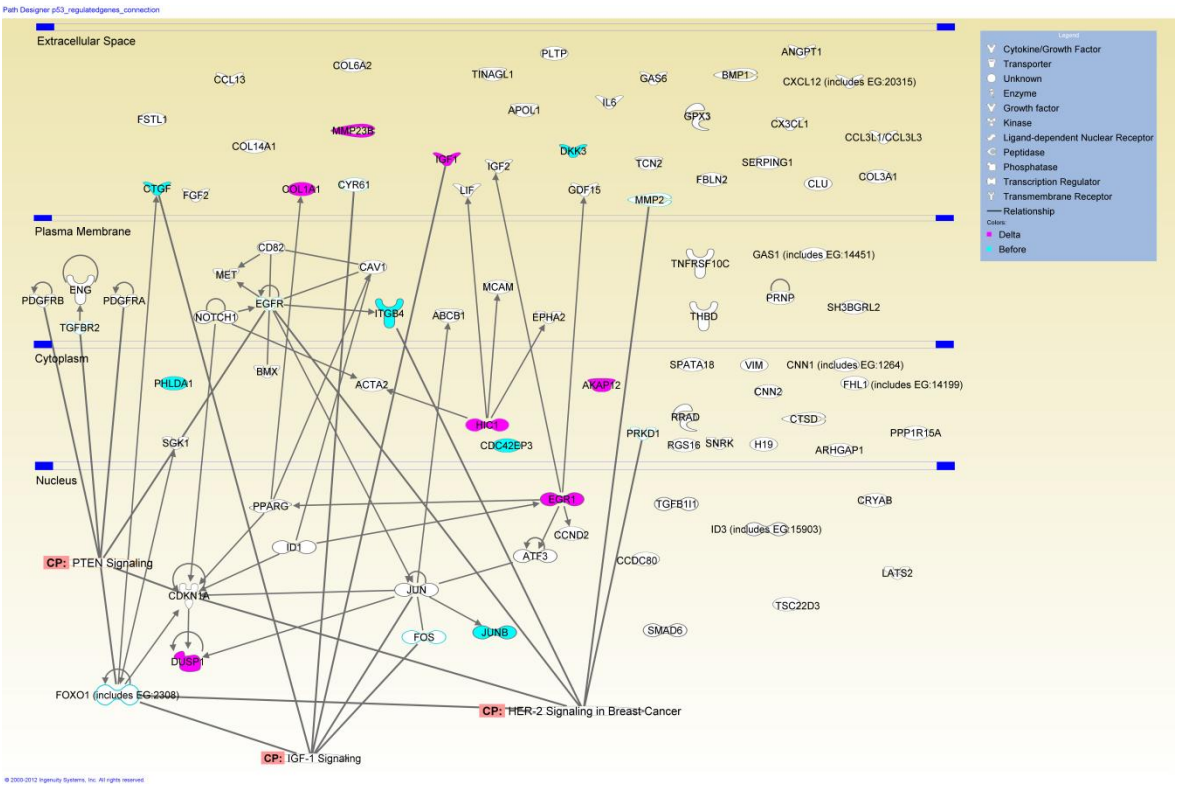

Figure 26. The relationship between p53 target molecules. The p53 target genes that were found to be up regulated in doxorubicin responders from the analysis of before samples are illustrated in blue, while the p53 target genes found to have a higher delta in doxorubicin responders are shown in pink. The location of the molecules in addition to the connection between them is illustrated in the figure. The canonical pathways (CP) included are enriched signaling pathways within the uploaded genes.

5.4 mRNA Expression Profile and *TP53* Mutation Status

Certain *TP53* mutations have been shown to predict for resistance to doxorubicin in breast cancer patients. Additionally, p53 has also been shown to modulate topoisomerase activity. The comparison between the mRNA expression profiles of samples containing wild type and mutated p53 are described below.

5.4.1 PCA of samples with different *TP53* mutation status

PCA was performed on both before and after samples with wild type (n=99) and mutated (n=29) p53 to evaluate the differences in gene expression between the two groups (Figure 27), by using the normalized and filtered probe list containing 26234 probes. The scatter plot indicated that samples with different *TP53* mutation status do not separate completely. However, it appears that there is a certain distribution of the samples within the two groups. Samples with mutated p53 group mainly in one side of the PCA plot, while samples with wild type p53 group for the most part in another side of the plot.

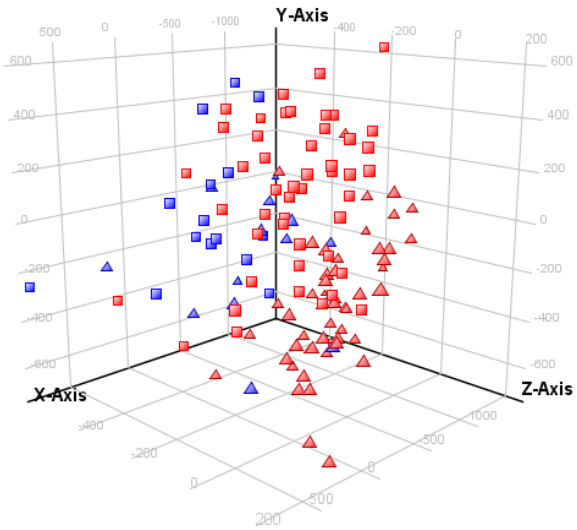

Figure 27. 3D scatter-plot of samples with different *TP53* mutation status. Samples with wild type p53 (n=99) are shown in red, while samples containing mutated p53 (n=29) are shown in blue. The PCA plot reveals that the samples do not separate perfectly. However, the graph may give the impression that samples with mutated p53 group mainly to the left, while samples with wild type p53 group mainly to the right. It is important to notice that the PCA plot includes both before and after samples. When focusing on the before and after samples separately the same trend was seen (data not shown).

5.4.2 SAM analysis of tumor samples with wild type and mutated p53

To investigate if genes were differentially expressed between the samples having a wild type p53 compared to those that had mutated p53 SAM analysis was performed. Delta expression values were used in the analysis, which correspond to the differences in expression between the samples collected before and after doxorubicin therapy from the same patient. The 112 samples representing 56 pairs were therefore used in this analysis. The comparison of the samples which had wild type 53 to those that had mutated p53 revealed 48 probes with a smaller delta in samples with *TP53* mutation, *i.e.* smaller differences in expression between the before and after samples for those genes (Figure 28). No genes were found to have a significantly higher delta in samples with mutated p53.

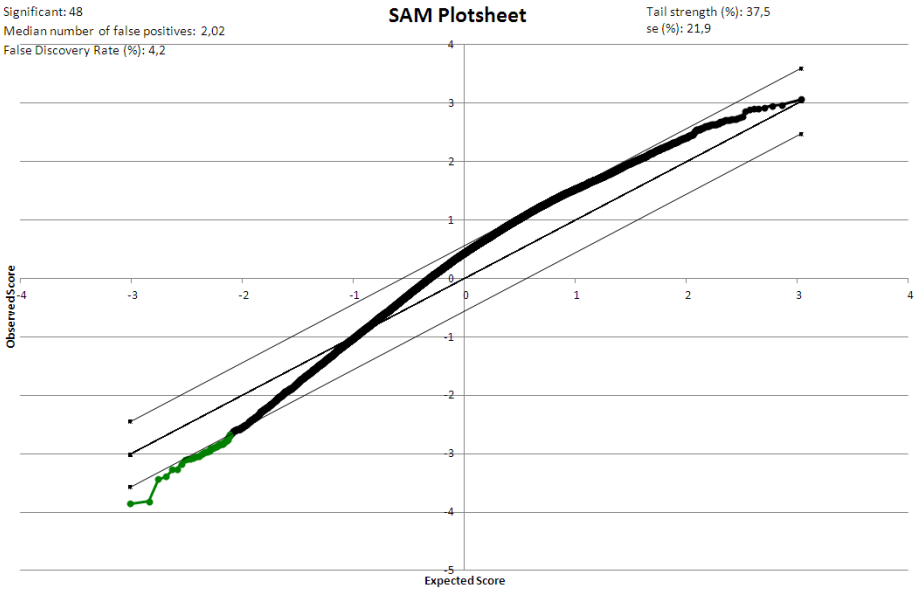

Figure 28. SAM analysis, paired samples. SAM analysis was performed on tumor sample pairs collected before and after doxorubicin treatment from the same patient (n=56), which either had wild type (n=45 pairs) or mutated (n=11 pairs) p53. The delta values were used as input (the differences in expression between the before and after sample for a given patient). Two class paired analysis revealed 48 probes having a significant difference in delta value between the wild type and the mutated p53 samples. These probes (FDR≈0-4.2%) had a lower delta value in the samples with a mutation in *TP53* representing a smaller change in expression following doxorubicin treatment.

For each of the 48 probes found to have a smaller delta in samples with *TP53* mutation, the average delta was calculated for the group of samples with different *TP53* status. The genes and the average delta in addition to the absolute difference in delta values are shown in Table 11.

Table 11. The 48 probes found to have a lower delta in tumor samples collected from patients with *TP53* mutation. For each probe, gene name, average delta value for both the *TP53* wild type and mutated samples and the absolute difference in delta values are displayed. A positive delta value corresponds to an induction in expression, while a negative delta value corresponds to repression. The list is sorted according to gene name.

Probe ID	**Gene name**	**Average delta *TP53* wt (n=45)**	**Average delta *TP53* mutated (n=11)**	**Absolute difference in delta**
A_24_P398790	*C14orf102*	0.58	-0.31	0.90
A_23_P70359	*C6orf59*	0.66	-0.52	1.18
A_24_P215653	*CLEC14A*	2.15	0.72	1.43
A_23_P42322	*COL11A2*	0.41	-0.95	1.37
A_23_P105803	*FGF9*	0.95	-0.92	1.87
A_24_P59471	*FLJ36840*	-0.26	-1.31	1.05
A_23_P55897	*KANK2*	0.37	-0.48	0.85
A_23_P127406	*KDM4D*	1.31	0.35	0.97
A_32_P100830	*KIF19*	1.79	-0.49	2.28
A_24_P184937	*KLHL36*	0.74	0.00	0.74
A_23_P32175	*LHX6*	1.45	-0.61	2.06
A_32_P199263	*LOC389831*	-0.45	-1.36	0.91
A_23_P397293	*LY6K*	1.01	-1.28	2.28
A_32_P138348	*LY6K*	0.97	-1.18	2.15
A_24_P392280	*MCTS1*	-0.19	-1.03	0.85
A_24_P142885	*PANK2*	0.12	-0.75	0.87
A_23_P71928	*SH2D3C*	1.75	0.68	1.07
A_24_P101314	*SHF*	0.52	-0.46	0.98
A_24_P356916	*SLC13A3*	0.57	-0.63	1.19
A_24_P102151	*SNRK*	0.16	-0.97	1.14
A_23_P254816	*TCF15*	1.24	-0.24	1.49
A_23_P40611	*TCN2*	1.58	0.44	1.14
A_23_P154566	*TOX2*	1.38	0.01	1.37
A_23_P144746	*ZNF454*	0.86	-0.63	1.49
A_23_P144704		-0.47	-1.44	0.97
A_23_P385084		0.75	-0.71	1.46
A_23_P57482		0.58	-0.61	1.18
A_24_P128361		1.13	-0.76	1.90
A_24_P204976		-0.01	-1.01	0.99
A_24_P221092		-0.13	-1.12	0.99
A_24_P247774		1.02	-0.88	1.90
A_24_P248255		0.65	-0.64	1.30
A_24_P367804		0.05	-1.09	1.14
A_24_P640212		-0.43	-1.43	0.99
A_24_P735073		-0.61	-1.58	0.97
A_24_P753849		-0.33	-1.18	0.85
A_24_P940615		0.20	-0.85	1.05
A_32_P111235		-0.08	-0.89	0.81
A_32_P113404		0.11	-1.02	1.13
A_32_P116088		0.19	-1.17	1.36
A_32_P139311		-0.27	-1.34	1.07
A_32_P196287		-0.14	-1.22	1.08
A_32_P198791		0.45	-0.91	1.36
A_32_P5432		0.09	-1.19	1.28
A_32_P6442		0.39	-0.41	0.80
A_32_P73039		-0.24	-1.58	1.34
A_32_P73580		-0.17	-1.04	0.87
A_32_P930953		-0.14	-1.24	1.10

5.4.3 Ingenuity Pathway Analysis of genes differentially expressed between samples with wild type and mutated p53

The genes identified by SAM analysis to have a different delta value between samples with wild type and mutated p53 were further investigated in IPA. First, the probe list with the 48 probes was processed. Un-annotated probes were updated using SOURCE where possible, while the probes that remained un-annotated were deleted. This resulted in 28 unique genes which were uploaded to IPA. The IPA analysis revealed information about the gene set, including molecular and cellular functions and transcription factors responsible for the gene expression changes (Table 12). Selecting a Benjamini-Hochberg threshold value of 0.05 identified 40 molecular and cellular functions enriched in the genes which had a lower delta value in samples with mutated p53. However, the majority of the functions did only have one molecule associated with the specific function. Therefore, another threshold was set to only including the functions which had 4 or more molecules involved, and they are listed in Table 12. No canonical pathways or tox lists were found to be significantly overrepresented in the genes having a smaller delta in the mutated p53 samples after correction.

Table 12. The molecular and cellular functions (a) and the five top transcription factors (b) overrepresented in the list of genes with a smaller delta in samples collected from patients with mutation in *TP53*. For each function, the Benjamini-Hochberg multiple testing correction p-value and the genes associated with the function are provided. The p-value, not corrected for multiple testing, and the number of target molecules are displayed for each transcription factor.

a) Molecular and Cellular Functions overrepresented in the list of genes with a smaller delta in p53 mutated samples (n=28)		
Name	*B-H P-value*	*Genes (≥4)*
Organ Morphology	1.58E-02 – 1.03E-01	*FGF9, KDM4D, SH2D3C, COL11A2*
Tissue Development	1.58E-02 – 1.00E-01	*COL11A2, FGF9, TCF15, LHX6, SH2D3C*
Cellular Development	1.58E-02 – 6.24E-02	*COL11A2, FGF9, TCF15, LHX6*
Tissue Morphology	1.58E-02 – 9.85E-01	*FGF9, TCF15, SH2D3C, MCTS1*
Small Molecule Biochemistry	2.54E-02 – 1.00E-01	*SLC13A3, TCN2, PANK2, FGF9*
b) Top Transcription Factors overrepresented in the list of genes with a smaller delta in p53 mutated samples (n=28)		
Transcription Regulator	*P-value of overlap*	*Target molecules*
PAX3	6.05E-04	*SNRK, FGF9, TCN2*
TOB1	6.66E-04	*SNRK, TCN2*
FUBP3	9.73E-04	*FGF9*
ZNF219	2.92E-03	*COL11A2*
HEXIM1	6.79E-03	*FGF9*

5.5 Intrinsic Subclassification

Intrinsic subclassification was performed both on the samples collected before and after doxorubicin therapy. Some of the samples included in this study were previously classified with subtype using cDNA microarrays. The classification of before samples was performed to evaluate if the gene expression experiment using new mRNA arrays would give the same or a different subtype of the samples that were previously classified. The classification of the after samples was performed to assess if the subtype of the after sample changed compared to the before sample from the same tumor as a result of doxorubicin therapy.

The ‘intrinsic’ gene list described by Sørlie and colleagues containing CloneIDs were updated with Unigene IDs using the Stanford SOURCE Search website. A total of 491 genes of the original 552 intrinsic genes had valid Unigene IDs. These were

matched with the normalized and filtered mRNA expression data for the 128 samples collected before and after doxorubicin treatment. In all, there were 590 genes that matched the intrinsic gene set because several genes had more than one probe that represented the same gene. The genes having more than one probe were excluded from the analysis, yielding a total of 239 genes with unique Unigene IDs that matched the intrinsic gene list. The before and after samples were classified with tumor subtype by calculating the Pearson correlation between each sample and the five centroids. The subtype of each sample was determined by using the centroid with the largest correlation coefficient.

5.5.1 Subclassification of before samples using cDNA- and Agilent 44K microarrays

A total of 30 before samples had previously been classified with cDNA microarrays. Comparing the new subtype classification of the before samples to the original subtype of these samples revealed that 11 of the samples were scored with a different subtype than the original (Figure 29). Six of the samples which were original classified in the HER2-enriched subgroup were scored as either luminal A or luminal B with the new classification. The rest of the samples with original luminal A or luminal B subtype were classified with the opposite luminal subtype.

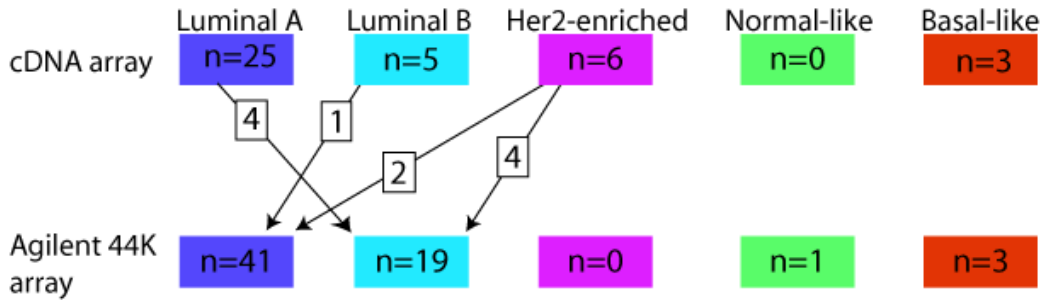

Figure 29. Intrinsic subclassification of samples collected before doxorubicin therapy. A total of 30 samples collected before treatment included in this study were previously classified using cDNA arrays. The new subclassification performed in this study was carried out using Agilent 44K/mRNA arrays (n=128). Of the 30 samples previously classified, 11 samples were scored with a different subtype using the new arrays. The black boxes indicate the change in subclassification between the two arrays for the overlapping samples.

5.5.2 Comparison of subclassification of before and after samples

Comparing the subclass of the before and after pairs classified with the new mRNA arrays, revealed that nine after samples were scored with a different subtype than the before sample from the same tumor (Table 13). Of the nine pairs, one sample was collected from a patient characterized as doxorubicin responder (PR) while the rest of the samples were collected from non-responders (SD/MC). In addition, when looking at the unsupervised cluster of all the samples included in this study, only the BC 80 before and after sample clustered together (see Figure 19).

Table 13. The nine before and after pairs in which the after samples were scored with another subtype following doxorubicin treatment compared to the before sample from the same patient. The correlation coefficients of the five centroids of the before (a) and after (b) sample of each pair are provided. The centroid with the largest correlation coefficient for each sample is indicated in red.

a) Before Sample	Lum A	Lum B	HER2-enriched	Basal-like	Normal-like
BC 10	0.086	0.130	-0.210	-0.172	-0.381
BC 103	0.168	0.262	-0.270	-0.156	-0.369
BC 106	0.184	0.289	-0.273	-0.172	-0.414
BC 85	0.175	0.186	-0.235	-0.173	-0.374
BC 50	0.221	0.008	-0.303	-0.254	-0.145
BC 71	0.003	-0.004	-0.102	-0.045	-0.067
BC 80	-0.131	0.074	-0.029	0.060	-0.285
BC 5	-0.091	0.412	-0.184	0.083	-0.460
BC 81	0.050	-0.139	-0.257	-0.037	0.124
b) After Sample	Lum A	Lum B	HER2-enriched	Basal-like	Normal-like
BC 10	0.159	-0.071	-0.252	-0.195	-0.069
BC 103	0.252	0.156	-0.328	-0.212	-0.341
BC 106	0.079	-0.071	-0.220	-0.080	0.028
BC 85	0.174	0.157	-0.185	-0.162	-0.291
BC 50	0.036	0.053	-0.165	-0.129	-0.232
BC 71	-0.026	0.213	-0.036	-0.045	-0.319
BC 80	-0.139	0.221	-0.104	0.271	-0.222
BC 5	0.006	-0.085	-0.179	-0.048	0.010
BC 81	-0.306	0.137	0.034	0.336	-0.183

6 Discussion

This section is divided into two parts. Biological considerations regarding the results are discussed first, followed by a discussion concerning the methodological considerations.

6.1 Gene Expression Analysis and Biological Considerations

A discussion of the results obtained in this study is described below in the same order as they are presented under the result section.

6.1.1 Gene expression pattern in tumor samples before and after doxorubicin treatment

Comparing the gene expression profile in tumor samples collected pre- and post neoadjuvant doxorubicin treatment may give information regarding doxorubicin action both at the individual gene level as well as pathway based. This may further explore the underlying molecular mechanisms of doxorubicin cytotoxicity.

The separation of the samples collected before and after doxorubicin treatment in the PCA analysis indicated an overall difference in gene expression between samples collected pre- and post therapy. We wanted to include both all samples and only the paired samples collected from the same tumor in different analyses. Including only the paired samples would make it possible to correct for intratumor differences to some degree, while including all samples would increase the power of the statistical analyses.

6.1.1.1 Unsupervised clustering

Unsupervised clustering of all the 128 experimental samples was performed to investigate the overall sample relations based on gene expression patterns. The cluster analysis by all probes did not separate the samples collected before and after doxorubicin therapy into two clear clusters. In other words, doxorubicin treatment does not overall lead to a distinct change in expression profile which can completely separate the samples collected from patients before and after doxorubicin treatment. When examining the subcluster that contained the majority of the samples, the cluster could be divided into three main clusters. A significant different distribution of the before and after samples in these clusters was revealed. The two external clusters on each end of the dendrogram were compromised of a majority of after samples.

Of the total 56 matched pairs consisting of tumor samples collected pre- and post therapy from the same patient, as many as 38 (67.9%) of the pairs clustered separately. This may indicate that doxorubicin therapy in some cases have a stronger impact in the gene expression profile than the molecular program of the primary tumor itself. But it may also point to intratumor transcriptional heterogeneity and be a result of sampling of different parts of the tumor. Interestingly, the majority of the pairs that clustered together were found in the 'before cluster' *i.e.* the cluster that contained the majority of the before samples. This may indicate a lower induction of genes after doxorubicin treatment for those pairs.

Previous studies comparing gene expression patterns in tumor samples drawn from the same patient before and after therapy have shown more similarity between samples collected from the same patient than between samples from different patients collected at the same time-point (54). Perou *et.al.* examined the gene expression pattern in 20 tumor pairs collected before and after doxorubicin treatment from the same patient using cDNA microarrays. In all, 15 pairs (75%) were revealed to cluster together. A total of 15 of these tumor pairs were also included in this study. The clustering analysis revealed that 10 of these 15 pairs were separated using new mRNA arrays, while 5 pairs still clustered together. The differences in frequency of pairs clustering together may be that the previous study utilized fewer genes *i.e.*1753 to organize the samples based on their gene expression profile. The clustering in this thesis was performed using 26234 probes corresponding to 14046 unique genes. The higher amount of probes may detect more differences in expression, therefore separating more of the tumor samples drawn from the same patient. Our results from the clustering analysis indicates that for a subset of the samples the alterations in gene expression profile following treatment is high enough to separate the before and after samples collected from the same tumor.

6.1.1.2 Genes differentially expressed between before and after samples

When SAM analysis was performed to identify differentially expressed genes between before and after samples, a high number of probes were revealed to be up- and down regulated after treatment for both the unpaired and paired analysis. We used the probes that were differentially expressed from the unpaired comparison of before and after samples to cluster the 128 experimental samples. When using this subset of

probes, only four before and after pairs still clustered together. This indicates a greater separation of the samples compared to when all probes from the microarray analysis were used.

We selected the genes from the SAM analysis showing the top 25% fold change to be further studied in pathway analyses. The genes, with the top 25% fold change, showed high concordance between the unpaired and paired gene lists. This may imply that intratumor heterogeneity is not affecting the results to a high degree when focusing on the genes with the highest fold change. The analysis of the list of unique genes which were shown to be up regulated after doxorubicin treatment are discussed first, followed by a discussion of the list of genes down regulated after treatment.

6.1.1.3 Analysis of genes up regulated after therapy

Analysis of selected up regulated genes in IPA revealed several identical functions, canonical pathways, transcription regulators and tox lists between the unpaired and paired analysis. Since the concordance between the genes having the top 25% fold change in the unpaired and paired gene list was high, this is a result we expected. The similarity implicated that either including or excluding the samples collected only before or after therapy from a given patient do not seem to affect the results when focusing on the genes with the highest fold change. Including more of the genes identified in the SAM analysis may have increased the difference between the unpaired and paired analysis, but this have to be explored in future analysis and are not included in this thesis.

The three functions “cardiovascular system development and function”, “organismal development” and “cellular movement” found to be enriched in the list of up regulated genes from both analysis involved among other development of blood vessels, vasculogenesis, angiogenesis, migration of endothelial cells and migration of tumor cells. This may indicate that doxorubicin therapy influences these processes. If doxorubicin treatment causes cell cycle arrest due to inhibition of topoisomerase II and the creation of DNA breaks, the tumor may use other mechanism such as angiogenesis to continue to grow and proliferate. A study using a transgenic breast cancer model implicated that the anti-tumor activity of doxorubicin was enhanced by inhibiting the angiogentic signaling receptor VEGF2 (94). Doxorubicin was not

reported to influence tumor angiogenesis in a study focusing on chemotherapy and angiogenesis (95), but this should be further examined. Another study from metastatic breast cancer in mouse models reported that doxorubicin stimulated both motility and invasion of breast cancer cells by activating TGFβ, transforming growth factor β, signaling (96). The researchers reported that this has the potential to generate drug resistant cancer cells. Likewise, a study using a transgenic model for breast cancer metastasis showed that doxorubicin increased the level of TGFβ as well as circulating tumor cells (97). Among the genes with top 25% fold change up regulated after therapy in our study, *MMP2, matrix metallopeptidase 2,* was involved in several of the functions and pathways enriched within the gene lists. Matrix metallopeptidases are responsible for the degradation of the extracellular matrix and have been reported to be involved in breast cancer invasion, metastasis and tumor angiogenesis (98;99). Together, it can be speculated that these findings indicate that doxorubicin not only induce a cytotoxic effect on cancer cells, but may in addition also induce cell migration, invasion and angiogenesis. This is only a hypothesis so far, and further investigation will be needed to explore these possible functions of doxorubicin.

Several of the canonical pathways overrepresented in the list of genes with the top 25% fold change up regulated after therapy included immune responses such as pathways involved in Rheumatoid Arthritis, "complement system" and "leukocyte extravasation signaling", indicating an activation of immune responses following treatment. Chemotherapeutic drugs have been shown to both induce immune responses which can promote tumor growth and responses which can lead to undesirable side effects (100). The biological effect of the complement system includes among other clearance of immune complexes and apoptotic cells. The cytotoxic effect of doxorubicin leading to cell death may generate cell debris that needs to be cleared. Likewise, the leukocyte extravasation which is the migration of leukocytes from blood to tissue during inflammation may be a result of the inflammatory response which arises in the tumor after doxorubicin treatment (101). The pathways "Wnt/β-catenin signaling" and "human embryonic stem cell pluripotency" were also found to be enriched in the list of genes up regulated after treatment in both the unpaired and paired analysis. Wnt/β-catenin signaling compromises extracellular growth factors involved in various aspects of development

such as cell differentiation, cell polarity and cell proliferation. Interestingly, Wnt is involved in controlling stem cell self-renewal and differentiation (102). The Wnt/β-catenin pathway has been reported to be deregulated in cancer or cancer stem cells (102). It has also been implicated to play a role in doxorubicin chemoresistance in osteosarcoma and nevroblastoma tumor cells (103;104). A study of early lung metastasis provided a link between the Wnt signaling pathway and epithelial-mesenchymal transition in basal-like breast cancer, involved in the metastatic cascade to increase cells motility (105). When inhibiting the Wnt signaling, the cancer cells decreased the capacity to self renew and drive tumorigenesis. The canonical pathway "PTEN signaling" was only found to be enriched in the gene list containing up regulated genes from the unpaired analysis. PTEN is a tumor suppressor phosphatase which regulates signalling pathways involved in cell growth, migration and apoptosis (106). Since PTEN signaling is overrepresented in the list of genes up regulated after doxorubicin treatment, this may imply that genes in the PTEN signalling pathway are involved in the doxorubicin induced apoptosis. A study using PTEN-null cells reported that apoptosis following doxorubicin treatment was decreased in the cells with truncated PTEN protein compared to cells expressing a functional PTEN (107).

Interestingly, when looking more into the up regulated genes with top 25% fold change involved in the different functions and canonical pathways, we found several genes reported to have a transcriptional response to doxorubicin treatment in tumors such as *ABCB1*, *ABCG2* and *FGF2*. *ABCB1* and *ABCG2* are ATP-dependent membrane transporters which promote drug efflux leading to decreased drug accumulation. Both genes have been reported to mediate drug resistance (108). Likewise, *FGF2, fibroblast growth factor 2,* have been implicated in the development of doxorubicin resistance (109).

When comparing the transcription factors revealed by IPA to be overrepresented as gene expression regulators within the unpaired and paired gene lists, *TP53* was found to regulate the highest number of genes within both lists. p53 plays a major role in DNA damage response and apoptosis (110). Mutations in p53 have previously been associated with resistance to doxorubicin therapy in breast cancer patients (111). A study of p53 knockout cells showed that these cells are more resistant to doxorubicin than cells having a wild type p53 (112). Likewise, a study of doxorubicin action in

mice models revealed that mice lacking p53 had a growth delay when treated with doxorubicin, indicating that doxorubicin efficacy is modulated by p53 (113).

Other transcription factors such as SP1, SP3 and KLF2 were also found to regulate a high number of target genes found in both the unpaired and paired lists containing up regulated genes. SP1 and SP3 are members of the Specificity Protein/Kruppel-like factor transcriptions regulators that have roles in various cellular processes. SP1 and SP3 have been associated with tumorigenesis, and the expression of both genes have been found to be increased in cancer cells (114). KLF2 is a member of Kruppel-like factor proteins that have tumor suppressor functions and are reported to inhibit proliferation, migration and angiogenesis while inducing apoptosis and adhesion (115). KLF2 expression has been shown to be reduced in many malignancies such as ovarian cancer (116). Additionally, it has been reported to be up regulated in a doxorubicin resistant osteosarcoma cell line and therefore indicated to be involved in a drug resistant phenotype (117).

During the IPA analysis we also investigated the tox lists which were found to be enriched in the genes with top 25% fold change up regulated after doxorubicin therapy for both the unpaired and paired analysis. The tox lists were compromised of functions involved in cardiotoxicity, hepatotoxicity and nephrotoxicity. The main adverse effect of doxorubicin is known to be irreversible cardiotoxicity often leading to degenerative cardiomyopathy/heart failure (118). Many mechanisms of the cardiotoxicity induced by doxorubicin have been proposed and studied. Several studies have implicated that oxidative stress involving production of reactive oxygen species (ROS) during doxorubicin metabolism causes the doxorubicin-induced cardiotoxicity (119). Interestingly, “oxidative stress” was one of the tox list found to be enriched in the paired gene list from the comparison of pre- and post doxorubicin samples. The other tox functions involving liver and kidney toxicity may be a result of the pharmacokinetics of doxorubicin. The biotransformation of doxorubicin occurs primarily in the liver, and doxorubicin clearance is mediated by bile excretion and renal clearance (120). An important consideration regarding the analysis of the tox list enriched after doxorubicin treatment is that we have only measured the gene expression in tumor tissue and not in heart, liver or kidney tissue. However, since doxorubicin is administered systemic we assume that the treatment effects on gene

expression are similar in other tissues such as the heart but this would have to be confirmed *e.g.* in animal models.

6.1.1.4 Analysis of genes down regulated after therapy

The genes with the top 25% fold change found to be down regulated following therapy for the unpaired and paired analysis were uploaded to IPA. The cellular and molecular function "RNA post-transcriptional modification" was revealed by IPA to be enriched in the list of genes down regulated after treatment for both the unpaired and paired analysis. This pathway involves among other RNA- and mRNA processing and RNA- and mRNA splicing. In concordance with our results, a study where doxorubicin-induced gene expression changes in human breast cancers were examined reported that the down regulated genes were involved in among other RNA splicing, RNA processing, mRNA processing and mRNA metabolism (121). When looking at the canonical pathways found to be enriched in the list of genes down regulated after doxorubicin therapy only "EIF2 signaling" was found to be overrepresented within the unpaired and paired gene lists. EIFs are eukaryotic translation initiations factors involved in mRNA translation (122). These findings suggest that tumors alter processes involved in RNA/mRNA processing and metabolism in addition to protein synthesis in general after doxorubicin exposure, maybe to focus on other mechanisms more important for cell survival.

The HNF4A transcription factor regulated the highest number of targets genes within the unpaired as well as the paired gene lists. This transcription factor is involved in hepatocyte differentiation, and its loss has been associated with hepatocellular carcinoma development (123). The transcription factor MYCN regulated also a high number of the genes down regulated after doxorubicin therapy. MYCN controls numerous cellular processes such as cell growth and proliferation, cell-cycle progression, transcription, differentiation, apoptosis and cell motility (124). It is frequently amplified in neuroblastomas and MYCN over expression have been found in breast carcinomas (125;126). However, neither of these transcription factors have previously been connected to doxorubicin treatment.

All together, our results implicates that the gene expression in tumor is altered as a result of doxorubicin treatment. Several genes were found to be up- and down

regulated following treatment. However, we cannot exclude the possibility that some of the differences in mRNA expression between the before and after samples in the paired analysis is a result of different clones collected before and after therapy from the same tumor. It is known that breast cancer is a highly heterogeneous disease, and the collection of distinct clones may have an effect on the gene expression analysis.

6.1.2 Gene expression pattern and response to therapy

The gene expression pattern in samples collected from patients characterized as responders (PR) and non-responders (SD/MC) to doxorubicin were examined with the aim to identify a predictive profile which could be used to tailor the treatment for each patient. Before the gene expression profile of the samples in the two response categories was examined, we investigated the association between clinical parameters such as ER status, grade and *TP53* mutations status and doxorubicin response. ER status of the samples collected after therapy was found to be associated with response. This is in agreement with a study of breast cancer response to neoadjuvant anthracycline-based chemotherapy that reported that ER negativity was correlated with better response (127). However, this connection was functionally unexplained by the authors.

PCA analysis performed on samples collected from the responder and non-responder patients did not discriminate the samples in the two response groups, indicating an overall gene expression similarity. When SAM analysis was performed on the before samples, the after samples and on the delta expression values no significant genes were found to be differentially expressed between the two response groups. In addition, when only the probes revealed to be differentially expressed by SAM analysis of before and after samples were used no genes were found to be differentially expressed between doxorubicin responders and non-responders. We then tried a different approach by using gene set enrichment analysis to detect differences in gene expression of whole gene sets, but no differences were observed between the two response groups for any gene set.

Since p53 has been connected to doxorubicin resistance, we selected p53 target genes found in the IPA analysis of paired before and after samples and evaluated the connection between these genes and response to doxorubicin. Performing SAM

analysis of these genes only, revealed several genes up regulated in before and after samples collected from patients characterized as having a partial response to doxorubicin. By using delta expression values several genes were found to have a higher delta in samples from doxorubicin responders, indicating an induction in expression of those genes. The up regulated genes from the analysis of after samples had a higher false discovery rate, we therefore chose to focus on the genes revealed to be connected to response from the analysis of before samples and delta values. Some of the genes are discussed below.

CTGF, connective tissue growth factor, encodes an extracellular matrix-associated signaling molecule and has been shown to promote endothelial cell growth, migration, adhesion and survival (128). There is also strong evidence for that that this protein is involved in angiogenesis, a study of breast cancer cells showed that connective tissue growth factors including CTGF were involved in breast cancer angiogenesis (129). A study of neoadjuvant doxorubicin and cyclophosphamide treatment of breast cancer patients revealed that CTGF was more expressed in samples collected after than before treatment, and it was implicated that this gene could be connected to drug resistance (130). *ITGB4* encodes integrin, beta 4, a member of the integrin family transmembrane receptors which mediates cell-matrix or cell-cell adhesion in addition to transducing signals that regulates gene expression and cell growth. ITG4B has been proposed to be involved in carcinoma progression by regulating the migration, invasion and survival of carcinoma cells (131). This gene has also been found to promote tumor angiogenesis in a mouse model (132). *IGF1* was found to have a higher delta in the samples collected from patients assessed with partial response to doxorubicin. It is known to have an important function in cancer biology and has been reported to mediate resistance to chemotherapy (133). A study of atypical teratoid rhabdoid tumor cells revealed that inhibition of the IGF1 receptor sensitized the tumor cells to doxorubicin and cisplatin treatment (134). Another study showed that IGF1 attenuated the response of breast cancer cells to doxorubicin by inducing proliferation and inhibiting apoptosis (135). The results from this study may also imply that IGF1 is involved in response to doxorubicin treatment.

The genes found to be up regulated in doxorubicin responders should be further studied in the continuation of this thesis by performing *e.g.* experiments in cell lines.

mRNA expression data from untreated and doxorubicin treated xenograft models have been generated in the lab previously, and the results from this study should also be compared to the data from the animal models.

It could be speculated that no prominent predictive gene expression profile was observed when performing SAM on the total microarray data because the tumors included in this study had a less effective response to doxorubicin therapy. No samples were characterized with a complete response. The small difference in partial response compared to stable disease and/or minimal change may also explain why predictive expression profiles were difficult to explore. Although our results indicate that some genes can be connected to responses to doxorubicin, this does not exclude that other predictive profiles exists. Several genes differentially expressed between patients obtaining partial response and stable disease and/or minimal change may become apparent when a larger number of patients can be evaluated. Additionally, response classifiers may be revealed when patients with more distinct responses to doxorubicin are included in the analysis.

Doxorubicin pharmacology has been shown to have large inter-individual variation, and this may also contribute to the response (70). Both doxorubicin efficacy and toxicity have been implicated to be influenced by genetic variants effecting the expression of genes responsible for transport, metabolism and drug action. In the continuation of this study it would be interesting to evaluate these aspects. Whole genome genotyping of the germline DNA as well as tumor DNA collected from the patients included in this study have already been performed. The analysis from the genotyping experiment may elucidate the influence of genetic variation on both mRNA expression level and on response to doxorubicin.

6.1.3 Gene expression pattern and *TP53* mutation status

Since p53 have been shown to modulate the activity of topoisomerase II which is the proposed target of doxorubicin and mutations in *TP53* have been associated with doxorubicin resistance, we examined the differences in expression profiles between samples with wild type and mutated p53. The PCA analysis did not discriminate the samples completely, but a skewed distribution of the samples was still observed. From the SAM analysis 48 probes representing 28 unique, annotated genes with a lower

delta were revealed in the samples with p53 mutations. The lower delta indicates a lower induction of these genes. The IPA analysis uncovered several general molecular and cellular functions such as organ- and tissue morphology and cellular- and tissue development to be enriched within the list of 28 unique genes. Doxorubicin has previously been found to induce morphological changes in H9c2 myoblasts (136), but information connected to the finding form IPA are limited.

The transcription factor PAX3 was found to regulate three of the genes which had a smaller delta in the samples with p53 mutation. PAX3 is a transcription factor involved in normal embryonic development, and has been implicated to be involved in tumorigenisis (137;138). Interestingly, PAX3 has been reported to suppress p53 accumulation and p53-dependent apoptosis (139). The TOB1 transcription factor regulated two of the genes found to have a lower delta in the samples with mutation in p53. This transcription factor has been proposed to function as a tumor suppressor by inhibiting cell cycle progression in breast cancer specimens as well as breast cancer cell lines, therefore suppressing tumorigenisis (140). However, the link between PAX3 and TOB1 to p53 mutations and the effect of doxorubicin will require further analyses. It should be clarified that the results from the gene expression analysis should be interpreted with cautions because only a few samples had a mutated p53 and few genes were uploaded to IPA, decreasing the power of the analysis. A better insight into the role of p53 mutations and effect of doxorubicin may become apparent if more patients with a mutation in *TP53* were included in the analysis.

6.1.4 Subclassification

The subclassification of the tumor samples included in this study was performed by using 239 unique genes from the ‘intrinsic’ gene list provided by Sørlie and colleagues (92). The intrinsic gene list divided the breast cancer samples into five subgroups: luminal A, luminal B, HER2-enriched, basal-like and normal-like. In all, 30 of the before samples included in this study had previously been classified using cDNA microarrays. When comparing the subclass of the same samples, 11 samples changed subtype classification. Of the six samples assessed as HER2-enriched subtype in the previously characterization, all obtained another subtype with the new classification. It would be difficult to conclude if this is a result of the new mRNA arrays which have higher resolutions than the old cDNA arrays, if the isolation of new

mRNA give another representation of the phenotype of a heterogeneous tumor or if this is due to the fact that only a proportion of the intrinsic genes were used in the re-classification.

Clonal selection during cancer evolution and also during cancer therapy have recently been implicated to exist (68). One of the main challenges in cancer treatment is that some cells escape therapy. In order to evaluate if doxorubicin has a positive selection of certain clones, we compared the subclass of the before and after samples from the same patient. This revealed that nine of the after samples were scored with a different subclass than the before sample. Four pairs (BC 10, BC 103, BC 85 and BC 81) had a high correlation coefficient indicating a strong change in the subclass. Of these pairs, three changed subtype from luminal B to luminal A and one changed from normal-like to basal-like following doxorubicin treatment. The change from luminal B to luminal A subtype may indicate an reduction of proliferation after treatment, which may be connected to a more favorable outcome (141). In contrast, the change from normal-like to basal-like is possible indicating a more aggressive phenotype following therapy (142). Of the nine pairs that changed subtype after doxorubicin therapy, all except from one pair was collected from non-responders (SD/MC). These observations may also, as discussed above, be a result of using a subset of the original intrinsic gene list. Additionally, the correlation coefficient of either the before or the after sample of five pairs was below 0.1, which is very low and the results should therefore be interpreted with caution for these.

To increase the confidence in the subclassification it would be important to increase the number of probes used for the correlation analysis. To address the problem of replicate probes, we could have used the average of the expression of the replicate probes. This would only be possible if the different probes showed the same trends with regards to expression. Another option is to blast the sequences of the original CloneIDs for the intrinsic genes to find the exact location of the probes and include the probes from our data set that are closest to this position. Both alternatives would have increased the number of intrinsic genes included in the classification, giving the classification a stronger power. The problem with both alternatives is that the included probes will still be an approximation for the original clones. Because of time

limitations neither alternatives were prioritized in this study but should be performed in the continuation of the project.

6.2 Gene Expression Analysis and Methodological Considerations

The methodological considerations regarding the gene expression analysis are described below.

6.2.1 Patient samples and RNA quality control

The tumor samples used in this thesis were collected previously for other studies. The RNA from the samples was isolated using Trizol and RNeasy minikit from Qiagen and stored at -80 ºC. Information regarding the RNA quantity existed from previous measurements. For those samples that were listed with very high concentrations, the quantity was measured again by NanoDrop ND-1000 UV-VIS Spectrophotometer (Thermo Scientific, Wilmington, Delaware, USA) because we speculated that these values were not reliable. RNA to be used in the gene expression microarray analysis was subtracted from the stock solutions and a dilution of 40 ng RNA was made for each sample. For the RNA samples included in this study, only a minority were reported with a RIN value. The quality of the samples was assessed by Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA) to ensure high quality RNA. However, the amount of RNA in the stock solution was low for many samples and RNA assessment was therefore only performed on a selection of the samples. Because RNA from only a subset of the samples had been quality assessed, we cannot exclude the possibility that some effects may be related to sample quality. However, since the generated cRNA in the gene expression procedure are quantified using NanoDrop ND-1000 UV-VIS Spectrophotometer before the samples are hybridized to the arrays, this step will reveal samples of low quality. Of the total 131 samples included in this study, two samples had very low Cy3 specific activity after the amplification and labelling phase. These samples were therefore excluded from the hybridization step. Additionally, some of the samples had a lower cRNA concentration or Cy3 specific activity than recommended in the protocol from the gene expression experiment after the amplification step. However, these samples were accepted with good quality in additional quality controls performed as a part of the array scoring and post-wet lab analysis.

6.2.2 Technical variability and data quality control

Gene expression microarray experiment is a sensitive method, and technical variability during the lab procedure can influence the outcome. It is therefore important to minimize the sources of the technical variability. The variability can be a result of *e.g.* differences in hybridization or dissimilarities between microarrays (batch effects). During the microarray experiment performed in this study, the same batch of chemicals was used for the all the samples on each array. In addition, the experiment was performed in a time interval of two months by only two persons. The before and after samples and the samples in the different response categories were distributed on all the microarrays to minimize the probability of batch effects. The Agilent Technologies 4x44K Whole Human Genome microarrays used in this study contains a number of control probes which are utilized to monitor the experiment workflow. For each microarray a quality report summarizing the experiment was generated and evaluated. A few samples did not pass the quality control the first time for several technical reasons (such as low hybridization of the positive controls), and the experiment was repeated for these samples. All samples included in this thesis passed the quality control. Taken together, we concluded that the microarray data was of good quality.

After the mRNA expression analysis data were extracted from the scanned images using the Feature Extraction software. The data were then uploaded to GeneSpring GX, in this process the data were automatically log2 transformed. The data were further normalized using 50% percentile shift. 50% percentile normalization was chosen because it aligns the sub-arrays by their median expression value, and do not scale all the expression intensities such as *e.g.* quantile normalization does. We therefore speculated that smaller differences in expression levels will be remained to a higher extent with percentile normalization compared to quantile normalization. Quantile normalization is a more strict normalization method, and the differences in expression revealed by further analysis may therefore be even more reliable. On the other hand, genes with smaller differences in expression may still have important biological functions and with this in mind we decided to normalize the data by 50% percentile normalization.

Principle Component Analysis (PCA) was performed to assess sample uniformity and quality in GeneSpring GX. Of the original 129 included samples, one was revealed to be an outlier. Going back to the quality control report for this sample, it was detected that the histogram of the signal intensities for this sample was different than for the other samples. In addition, when an unsupervised cluster of all the experimental samples was generated, the sample clustered entirely by itself. Based on these findings, the sample was discarded from further statistical analyses. The quality control on probes was also performed in GeneSpring GX. The probes were first filtered by expression, where a lower cut-off was set to 10% and a higher cut-off was set to 100%. Probes which were expressed less than 10% were therefore discarded because we concluded that probes with such low intensities could be noise or background. The probes were then filtered by flags, where probes flagged as 'compromised' (non-uniform, saturated and population outliers) were omitted from further analysis. We chose to include the flag call named 'not detected' (non-positive, non-significant and below background) because by removing the probes less than 10% expressed we expected the non-positive, non-significant and below background probes to also be removed from the probe list. It could be speculated that this may have resulted in that some probes with low expression were included in the further statistical analyses. In spite of this, the thresholds that we chose later in the analyses (*e.g.* only selected the genes with highest fold change etc.) will likely exclude these probes.

6.2.3 Statistical analysis and bioinformatics

mRNA expression experiments using microarrays generates a large amount of data and requires statistical- and bioinformatic analytic tools that can tolerate this extensive amount. Another challenge is the multiple testing problem which arises when many independent statistical tests are performed on the same data set.

Unsupervised clustering was performed in **R** because of the number of probes extended 26000 and clustering of this high amount of probes was difficult to perform in *e.g.* GeneSpring GX or J-Express. The unsupervised clustering was performed with all probes and without dividing the dataset into subgroups. Given that gene expression data following the log2 transformation are normally distributed and that we wanted the closeness of the samples in the cluster to be calculated based on trends and not on

magnitude, we chose Pearson correlation as the distance metric. The average linkage method was chosen because we evaluated that the most appropriate way to calculate the distance in the cluster was to use the average distance between all pairs (not the *e.g.* smallest or the biggest distance as for single and complete linkage method, respectively).

To investigate if genes were differentially expressed between tumor samples divided into groups based on experimental condition (*e.g.* before and after samples, doxorubicin responders- and non-responders) Significance Analysis of Microarrays (SAM) was performed. SAM was selected because of its capacity to handle large data sets. In addition, SAM provides a delta parameter which can be used to choose the threshold for significance based on the false discovery rate (FDR). This makes the results from SAM more reliable because they are corrected for multiple testing. To confirm the results from the SAM analysis we selected a few probes which were shown to be differentially expressed between two groups for further investigation in IMB SPSS Statistics 18. One-Way ANOVA revealed a significant difference in mean expression between the before and after samples for the selected genes. This led us to the conclusion that the results from SAM were trustworthy.

A large number of probes were shown to be differentially expressed when comparing samples collected before and after doxorubicin treatment by SAM analysis (>6000). We decided therefore to only include the probes which had the largest fold change, *i.e.* top 25%, for pathway based investigation. Additionally, we evaluated that the probes having the highest fold change (*i.e.* the probes showing the largest difference in expression between the before and after samples) were most interesting to examine first. However, for further analysis on this material it would be interesting to include all the differentially expressed probes to evaluate what effect this would have on the results.

Ingenuity Pathway Analysis (IPA) was chosen for the pathway analyses. All genes are weighted equal in the analysis, and a Fisher's Exact p-value are displayed as a default p-value. In order to correct for the multiple testing issue, we selected a Benjamini-Hochberg p-value which is corrected for multiple testing for the molecular and cellular functions, canonical pathways and tox list provided from the IPA analyses.

We selected the statistical and bioinformatic methods based on the ones we evaluated to be most correct for this experiment. It should however be noted that the different pre-processing methods and statistical analyses available may have an influence on the final outcome of microarray analysis in general. Interesting results from large scale microarray experiments should be validated by *e.g.* quantitative RT-PCR. Additionally, gene expression experiments measures only the abundance of the different mRNA molecules in the cells and will not necessarily correlate to the protein level. To increase the power of the results obtained in this study, investigation of other biological levels such as protein abundance should be performed in *e.g.* cell lines.

7 Conclusions and Future Perspectives

Here the mRNA expression changes following doxorubicin exposure have been investigated in a panel of samples collected from breast cancer patients before and after treatment with doxorubicin. The goal has been to shed light on the molecular biology underlying the tumor response to doxorubicin by identifying genes and pathways affected by the therapy. Doxorubicin was found to be involved in numerous cellular processes when examining the genes up- and down regulated after treatment. Some pathways were thought to serve for its anticancer effect, such as PTEN signalling and the involvement of p53 in DNA damage responses. In contrast, other mechanisms such as cell movement and angiogenesis may have the opposite effect on tumor cells causing the cells to continue to grow and invade. Mechanisms enriched within the down regulated gene list such as RNA processing and splicing in addition to protein translation may imply that doxorubicin induced damage causes the cells to focus on more important processes for survival, such as DNA repair following treatment. *TP53* was shown to regulate the highest number of molecules within the list of up regulated genes, and several genes were found to have a smaller change in expression in samples with *TP53* mutations. These findings support the involvement of *TP53* in doxorubicin response.

The genes found to be significantly differentially expressed between samples collected from doxorubicin responders compared to non-responders may serve as biomarkers for doxorubicin response. These genes may be important to identify the patients that will benefit from doxorubicin therapy and those that will not.

The alterations in gene expression profile between samples collected before and after therapy revealed in this study may contribute to obtaining a better insight into the underlying molecular mechanisms of doxorubicin antitumor activity. Additionally, the genes found to be differentially expressed between doxorubicin responders and non-responders may be a step forward in the direction to individualize doxorubicin treatment.

Further evaluations of the genes and pathways found to be altered following doxorubicin treatment should be performed to support the finding in this study. In the continuation of this thesis the entire gene list containing genes found to be

significantly differentially expressed should be studied, and not only the genes showing the highest fold change. Microarray data from cell lines and mouse models untreated and treated with doxorubicin exists, and this should be compared to our results. To evaluate the importance for the genes found to be involved in doxorubicin response, experiments in cell lines using miRNA/siRNA could be performed to evaluate the viability of the cells after these genes are inhibited.

References

(1) World Health Organization (2012), Cancer mortality and morbidity [online]. Available from: http://www.who.int/gho/ncd/mortality_morbidity/cancer/en/index.html, [Downloaded 12.03.2012].

(2) International Agency for Research on Cancer (IARC) (2010), GLOBOCAN 2008 [online]. Available from: http://globocan.iarc.fr/ [Downloaded 15.03.2012].

(3) Kreftregisteret (2011), Kreftstatistikk for 2009 [online]. Available from: http://kreftregisteret.no/no/registrene/kreftstatistikk/ [Downloaded 15.03.2012].

(4) Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. *Nat Genet* 2003 March;**33** Suppl:238-44.:238-44.

(5) Hamilton JP. Epigenetics: principles and practice. *Dig Dis* 2011;**29**(2):130-5.

(6) Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. *Molecular Biology of THE CELL*. 5th ed. Garland Science; 2008.

(7) Polyak K. Breast cancer: origins and evolution. *J Clin Invest* 2007 November;**117**(11):3155-63.

(8) Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011 March 4;**144**(5):646-74.

(9) Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* 2000 January 7;**100**(1):57-70.

(10) Korkaya H, Liu S, Wicha MS. Breast cancer stem cells, cytokine networks, and the tumor microenvironment. *J Clin Invest* 2011 October;**121**(10):3804-9.

(11) Spano D, Zollo M. Tumor microenvironment: a main actor in the metastasis process. *Clin Exp Metastasis* 2012 April;**29**(4):381-95.

(12) Medina D. The mammary gland: a unique organ for the study of development and tumorigenesis. *J Mammary Gland Biol Neoplasia* 1996 January;**1**(1):5-19.

(13) Ramsay DT, Kent JC, Hartmann RA, Hartmann PE. Anatomy of the lactating human breast redefined with ultrasound imaging. *J Anat* 2005 June;**206**(6):525-34.

(14) Vargo-Gogola T, Rosen JM. Modelling breast cancer: one size does not fit all. *Nat Rev Cancer* 2007 September;**7**(9):659-72.

(15) BREASTCANCER.ORG (2012), Image-Breast Anatomy [online]. Available from: http://www.breastcancer.org/symptoms/understand_bc/what_is_bc.jsp [Downloaded 25.03.2012].

(16) McNally S, Martin F. Molecular regulators of pubertal mammary gland development. *Ann Med* 2011 May;**43**(3):212-34.

(17) Howard BA, Gusterson BA. Human breast development. J *Mammary Gland Biol Neoplasia* 2000 April;**5**(2):119-37.

(18) Stingl J. Estrogen and progesterone in normal mammary gland development and in cancer. *Horm Cancer* 2011 April;**2**(2):85-90.

(19) World Health Organization. *Classification of Tumours, Pathology and Genetics of Tumours of the Breast and Female Genital Organs*. Third ed. International Agency for Research on Cancer (IARC); 2003.

(20) Harris JR, Lippmann ME, Morrow M, Osborne CK. *Diseases of the Breast*. Second ed. Philadelphia: Lippincott Willams & Wilkins; 2000.

(21) Ernster VL, Ballard-Barbash R, Barlow WE, Zheng Y, Weaver DL, Cutter G et al. Detection of ductal carcinoma in situ in women undergoing screening mammography. *J Natl Cancer Inst* 2002 October 16;**94**(20):1546-54.

(22) Polyak K. Is breast tumor progression really linear? *Clin Cancer Res* 2008 January 15;**14**(2):339-41.

(23) Chin K, de Solorzano CO, Knowles D, Jones A, Chou W, Rodriguez EG et al. In situ analyses of genome instability in breast cancer. *Nat Genet* 2004 September;**36**(9):984-8.

(24) Schnitt SJ. Molecular biology of breast tumor progression: a view from the other side. I*nt J Surg Pathol* 2010 June;**18**(3 Suppl):170S-3S.

(25) Peto J, Collins N, Barfoot R, Seal S, Warren W, Rahman N et al. Prevalence of BRCA1 and BRCA2 gene mutations in patients with early-onset breast cancer. *J Natl Cancer Inst* 1999 June 2;**91**(11):943-9.

(26) Peto J, Mack TM. High constant incidence in twins and other relatives of women with breast cancer. *Nat Genet* 2000 December;**26**(4):411-4.

(27) Stratton MR, Rahman N. The emerging landscape of breast cancer susceptibility. *Nat Genet* 2008 January;**40**(1):17-22.

(28) Walsh T, King MC. Ten genes for inherited breast cancer. *Cancer Cell* 2007 February;**11**(2):103-5.

(29) Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nat Genet* 2009 May;**41**(5):579-84.

(30) Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 1998 May 15;**280**(5366):1077-82.

(31) Brookes AJ. The essence of SNPs. *Gene* 1999 July 8;**234**(2):177-86.

(32) Foulkes WD. Inherited susceptibility to common cancers. *N Engl J Med* 2008 November 13;**359**(20):2143-53.

(33) McPherson K, Steel CM, Dixon JM. ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. *BMJ* 2000 September 9;**321**(7261):624-8.

(34) Russnes HG, Navin N, Hicks J, Borresen-Dale AL. Insight into the heterogeneity of breast cancer through next-generation sequencing. *J Clin Invest* 2011 October 3;**121**(10):3810-8.

(35) Rakha EA, Reis-Filho JS, Baehner F, Dabbs DJ, Decker T, Eusebi V et al. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res* 2010;**12**(4):207.

(36) Ignatiadis M, Sotiriou C. Understanding the molecular basis of histologic grade. *Pathobiology* 2008;**75**(2):104-11.

(37) Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* 1991 November;**19**(5):403-10.

(38) Singletary SE, Connolly JL. Breast cancer staging: working with the sixth edition of the AJCC Cancer Staging Manual. *CA Cancer J Clin* 2006 January;**56**(1):37-47.

(39) Singletary SE, Allred C, Ashley P, Bassett LW, Berry D, Bland KI et al. Revision of the American Joint Committee on Cancer staging system for breast cancer. *J Clin Oncol* 2002 September 1;**20**(17):3628-36.

(40) Bland KI, Menck HR, Scott-Conner CE, Morrow M, Winchester DJ, Winchester DP. The National Cancer Data Base 10-year survey of breast carcinoma treatment at hospitals in the United States. *Cancer* 1998 September 15;**83**(6):1262-73.

(41) Bertos NR, Park M. Breast cancer - one term, many entities? *J Clin Invest* 2011 October;**121**(10):3789-96.

(42) Welboren WJ, Sweep FC, Span PN, Stunnenberg HG. Genomic actions of estrogen receptor alpha: what are the targets and how are they regulated? *Endocr Relat Cancer* 2009 December;**16**(4):1073-89.

(43) Jordan VC, Brodie AM. Development and evolution of therapies targeted to the estrogen receptor for the treatment and prevention of breast cancer. *Steroids* 2007 January;**72(**1):7-25.

(44) Patel RR, Sharma CG, Jordan VC. Optimizing the antihormonal treatment and prevention of breast cancer. *Breast Cancer* 2007;**14**(2):113-22.

(45) Bartlett JM, Brookes CL, Robson T, van de Velde CJ, Billingham LJ, Campbell FM et al. Estrogen receptor and progesterone receptor as predictive biomarkers of response to endocrine therapy: a prospectively powered pathology study in the Tamoxifen and Exemestane Adjuvant Multinational trial. *J Clin Oncol* 2011 April;20;**29**(12):1531-8.

(46) Dunnwald LK, Rossing MA, Li CI. Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. *Breast Cancer Res* 2007;**9**(1):R6.

(47) Andrulis IL, Bull SB, Blackstein ME, Sutherland D, Mak C, Sidlofsky S et al. neu/erbB-2 amplification identifies a poor-prognosis group of women with node-negative breast cancer. Toronto Breast Cancer Study Group. *J Clin Oncol* 1998 April;**16**(4):1340-9.

(48) Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 1987 January 9;**235**(4785):177-82.

(49) Sjogren S, Inganas M, Lindgren A, Holmberg L, Bergh J. Prognostic and predictive value of c-erbB-2 overexpression in primary breast cancer, alone and in combination with other prognostic markers. *J Clin Oncol* 1998 February;**16**(2):462-9.

(50) Junttila TT, Akita RW, Parsons K, Fields C, Lewis Phillips GD, Friedman LS et al. Ligand-independent HER2/HER3/PI3K complex is disrupted by trastuzumab and is effectively inhibited by the PI3K inhibitor GDC-0941. *Cancer Cell* 2009 May 5;**15**(5):429-40.

(51) Murphy CG, Fornier M. HER2-positive breast cancer: beyond trastuzumab. *Oncology* (Williston Park) 2010 April 30;**24**(5):410-5.

(52) Nishimura R, Arima N. Is triple negative a prognostic factor in breast cancer? *Breast Cancer* 2008;**15**(4):303-8.

(53) Reis-Filho JS, Weigelt B, Fumagalli D, Sotiriou C. Molecular profiling: moving away from tumor philately. *Sci Transl Med* 2010 September 1;**2**(47):47ps43.

(54) Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA et al. Molecular portraits of human breast tumours. *Nature* 2000 August 17;**406**(6797):747-52.

(55) Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci* U S A 2001 September 11;**98**(19):10869-74.

(56) Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA* 2006 June 7;**295**(21):2492-502.

(57) Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. *Molecular Oncology* 2010 November 24;**5**:5-23.

(58) Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol* 2007;**8**(5):R76.

(59) Reis-Filho JS, Pusztai L. Gene expression profiling in breast cancer: classification, prognostication, and prediction. *Lancet* 2011 November;19;**378**(9805):1812-23.

(60) de SF, Bender R, Glas A, Rutgers E. Gene expression profiling: decoding breast cancer. *Surg Oncol* 2009 December;**18**(4):366-78.

(61) van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002 January 31;**415**(6871):530-6.

(62) Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004 December 30;**351**(27):2817-26.

(63) Norsk Bryst Cancer Gruppe (NBCG) (2011), Blåboka [online]. Available from: http://nbcg.no/nbcg.blaaboka.html [Downloaded 02.04.2012].

(64) Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 2005 May 14;**365**(9472):1687-717.

(65) D'Souza N, Darmanin G, Fedorowicz Z. Immediate versus delayed reconstruction following surgery for breast cancer. *Cochrane Database Syst Rev* 2011 July **6**;(7):CD008674.

(66) Helleday T, Petermann E, Lundin C, Hodgson B, Sharma RA. DNA repair pathways as targets for cancer therapy. *Nat Rev Cancer* 2008 March;**8**(3):193-204.

(67) Jain P, Alahari SK. Breast cancer stem cells: a new challenge for breast cancer treatment. *Front Biosci* 2011 January 1;**16**:1824-32.:1824-32.

(68) Greaves M, Maley CC. Clonal evolution in cancer. *Nature* 2012 January 18;**481**(7381):306-13.

(69) Weiss RB. The anthracyclines: will we ever find a better doxorubicin? *Semin Oncol* 1992 December;**19**(6):670-86.

(70) Jamieson D, Boddy AV. Pharmacogenetics of genes across the doxorubicin pathway. *Expert Opin Drug Metab Toxicol* 2011 October;**7**(10):1201-10.

(71) Gewirtz DA. A critical evaluation of the mechanisms of action proposed for the antitumor effects of the anthracycline antibiotics adriamycin and daunorubicin. *Biochem Pharmacol* 1999 April 1;**57**(7):727-41.

(72) DrugBank (2012), Doxorubicin [online]. Available from: http://www.drugbank.ca/drugs/DB00997 [Downloaded 02.04.2012].

(73) Wang JC. Cellular roles of DNA topoisomerases: a molecular perspective. *Nat Rev Mol Cell Biol* 2002 June;**3**(6):430-40.

(74) Attia SM. Comparative aneugenicity of doxorubicin and its derivative idarubicin using fluorescence in situ hybridization techniques. *Mutat Res* 2011 October 1;**715**(1-2):79-87.

(75) PharmGKB (2012), Doxorubicin Pathway [online]. Available from: http://www.pharmgkb.org/pathway/PA165292163 [Downloaded 03.04.2012].

(76) Minotti G, Menna P, Salvatorelli E, Cairo G, Gianni L. Anthracyclines: molecular advances and pharmacologic developments in antitumor activity and cardiotoxicity. *Pharmacol Rev* 2004 June;**56**(2):185-229.

(77) Kaye S, Merry S. Tumour cell resistance to anthracyclines--a review. *Cancer Chemother Pharmacol* 1985;**14**(2):96-103.

(78) Geisler S, Lonning PE, Aas T, Johnsen H, Fluge O, Haugen DF et al. Influence of TP53 gene alterations and c-erbB-2 expression on the response to treatment with doxorubicin in locally advanced breast cancer. *Cancer Res* 2001 March 15;**61**(6):2505-12.

(79) Hayward JL, Carbone PP, Heusen JC, Kumaoka S, Segaloff A, Rubens RD. Assessment of response to therapy in advanced breast cancer. *Br J Cancer* 1977 March;**35**(3):292-8.

(80) Butte A. The use and analysis of microarray data. *Nat Rev Drug Discov* 2002 December;**1**(12):951-60.

(81) Stears RL, Martinsky T, Schena M. Trends in microarray analysis. *Nat Med* 2003 January;**9**(1):140-5.

(82) Agilent Technologies Inc. One-Color Microarray-Based Gene Expression Analaysis, Low Input Quick Amp Labeling. Version 6.5. 2010.

(83) Agilent Technologies Inc. Agilent One Color RNA Spike-In Kit. First edition. 2005.

(84) Agilent Technologies Inc. Agilent microarray hybridization chamber user guide. Version 2.0. 2006.

(85) Agilent Technologies Inc. Agilent G2565CA Microarray Scanner System (Scan Control Software 8.5) User Guide. 2010.

(86) Agilent Technologies (2012), Feature Extraction Software [online]. Available from: http://www.genomics.agilent.com/CollectionSubpage.aspx?PageType=Product&SubPageType=ProductDetail&PageID=1379 [Downloaded 03.04.2012].

(87) Agilent Technologies Inc. Agilent Feature Extraction Software Referance Guide. Seventh Edition. 2009.

(88) Agilent Technologies Inc. Agilent RNA 6000 Nano Kit Guide. 2006.

(89) Agilent Technologies Inc. Agilent RNA 6000 Nano Kit Quick Start Guide. 2007.

(90) NanoDrop Technologies Inc. ND-1000 Spectrophotometer V3.5 User's Manual. Rev. 7. 2007.

(91) Agilent Technologies (2012), GeneSpringGX Manual [online]. Available from: http://www.chem.agilent.com/cag/bsp/products/gsgx/manuals/genespring-manual.pdf [Downloaded 15.04.2012].

(92) Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci* U S A 2003 July 8;**100**(14):8418-23.

(93) Chu G, Li J, Narasimhan B, Tibshirani R, Tusher V (2012), SAM "Significant Analysis of Microarrays" Users guide and technical document Release 4.0 [online]. Available from: http://www-stat.stanford.edu/~tibs/SAM/sam.pdf [Downloaded 16.04.2012].

(94) Lynn KD, Udugamasooriya DG, Roland CL, Castrillon DH, Kodadek TJ, Brekken RA. GU81, a VEGFR2 antagonist peptoid, enhances the anti-tumor activity of doxorubicin in the murine MMTV-PyMT transgenic model of breast cancer. *BMC Cancer* 2010 July 30;**10**:397.:397.

(95) Miller KD, Sweeney CJ, Sledge GW, Jr. Redefining the target: chemotherapeutics as antiangiogenics. *J Clin Oncol* 2001 February 15;**19**(4):1195-206.

(96) Bandyopadhyay A, Wang L, Agyin J, Tang Y, Lin S, Yeh IT et al. Doxorubicin in combination with a small TGFbeta inhibitor: a potential novel therapy for metastatic breast cancer in mouse models. *PLoS One* 2010 April 28;**5**(4):e10365.

(97) Biswas S, Guix M, Rinehart C, Dugger TC, Chytil A, Moses HL et al. Inhibition of TGF-beta with neutralizing antibodies prevents radiation-induced acceleration of metastatic cancer progression. *J Clin Invest* 2007 May;**117**(5):1305-13.

(98) Djonov V, Cresto N, Aebersold DM, Burri PH, Altermatt HJ, Hristic M et al. Tumor cell specific expression of MMP-2 correlates with tumor vascularisation in breast cancer. *Int J Oncol* 2002 July;*21*(1):25-30.

(99) Duffy MJ, Maguire TM, Hill A, McDermott E, O'Higgins N. Metalloproteinases: role in breast carcinogenesis, invasion and metastasis. *Breast Cancer Res* 2000;**2**(4):252-7.

(100) Zitvogel L, Apetoh L, Ghiringhelli F, Andre F, Tesniere A, Kroemer G. The anticancer immune response: indispensable for therapeutic success? *J Clin Invest* 2008 June;**118**(6):1991-2001.

(101) Sauter KA, Wood LJ, Wong J, Iordanov M, Magun BE. Doxorubicin and daunorubicin induce processing and release of interleukin-1beta through activation of the NLRP3 inflammasome. *Cancer Biol* Ther 2011 June 15;**11**(12):1008-16.

(102) Takebe N, Harris PJ, Warren RQ, Ivy SP. Targeting cancer stem cells by inhibiting Wnt, Notch, and Hedgehog pathways. *Nat Rev Clin Oncol* 2011 February;**8**(2):97-106.

(103) Dieudonne FX, Marion A, Marie PJ, Modrowski D. Targeted inhibition of TCF activity promotes syndecan-2 expression and sensitization to doxorubicin in osteosarcoma cells and bone tumors in mice. *J Bone Miner Res* 2012 May 1;**10**.

(104) Flahaut M, Meier R, Coulon A, Nardou KA, Niggli FK, Martinet D et al. The Wnt receptor FZD1 mediates chemoresistance in neuroblastoma through activation of the Wnt/beta-catenin pathway. *Oncogene* 2009 June 11;**28**(23):2245-56.

(105) DiMeo TA, Anderson K, Phadke P, Fan C, Perou CM, Naber S et al. A novel lung metastasis signature links Wnt signaling with cancer cell self-renewal and epithelial-mesenchymal transition in basal-like breast cancer. *Cancer Res* 2009 July 1;**69**(13):5364-73.

(106) Chalhoub N, Baker SJ. PTEN and the PI3-kinase pathway in cancer. *Annu Rev Pathol* 2009;**4**:127-50.:127-50.

(107) Wan X, Li J, Xie X, Lu W. PTEN augments doxorubicin-induced apoptosis in PTEN-null Ishikawa cells. *Int J Gynecol Cancer* 2007 July;**17**(4):808-12.

(108) Gottesman MM, Fojo T, Bates SE. Multidrug resistance in cancer: role of ATP-dependent transporters. *Nat Rev Cancer* 2002 January;**2**(1):48-58.

(109) Song S, Wientjes MG, Gan Y, Au JL. Fibroblast growth factors: an epigenetic mechanism of broad spectrum resistance to anticancer drugs. *Proc Natl Acad Sci* U S A 2000 July 18;**97**(15):8658-63.

(110) Brady CA, Attardi LD. p53 at a glance. *J Cell* Sci 2010 August 1;**123**(Pt 15):2527-32.

(111) Aas T, Borresen AL, Geisler S, Smith-Sorensen B, Johnsen H, Varhaug JE et al. Specific P53 mutations are associated with de novo resistance to doxorubicin in breast cancer patients. *Nat Med* 1996 July;**2**(7):811-4.

(112) Dunkern TR, Wedemeyer I, Baumgartner M, Fritz G, Kaina B. Resistance of p53 knockout cells to doxorubicin is related to reduced formation of DNA strand breaks rather than impaired apoptotic signaling. *DNA Repair* (Amst) 2003 January 2;**2**(1):49-60.

(113) Dart DA, Picksley SM, Cooper PA, Double JA, Bibby MC. The role of p53 in the chemotherapeutic responses to cisplatin, doxorubicin and 5-fluorouracil treatment. *Int J Oncol* 2004 January;**24**(1):115-25.

(114) Li L, Davie JR. The role of Sp1 and Sp3 in normal and cancer cell biology. *Ann Anat* 2010 September;20;**192**(5):275-83.

(115) Black AR, Black JD, Azizkhan-Clifford J. Sp1 and kruppel-like factor family of transcription factors in cell growth regulation and cancer. *J Cell Physiol* 2001 August;**188**(2):143-60.

(116) Wang F, Zhu Y, Huang Y, McAvoy S, Johnson WB, Cheung TH et al. Transcriptional repression of WEE1 by Kruppel-like factor 2 is involved in DNA damage-induced apoptosis. *Oncogene* 2005 June 2;**24**(24):3875-85.

(117) Rajkumar T, Yamuna M. Multiple pathways are involved in drug resistance to doxorubicin in an osteosarcoma cell line. *Anticancer Drugs* 2008 March;**19**(3):257-65.

(118) Singal PK, Iliskovic N. Doxorubicin-induced cardiomyopathy. *N Engl J Med* 1998 September 24;**339**(13):900-5.

(119) Simunek T, Sterba M, Popelova O, Adamcova M, Hrdina R, Gersl V. Anthracycline-induced cardiotoxicity: overview of studies examining the roles of oxidative stress and free cellular iron. *Pharmacol Rep* 2009 January;**61**(1):154-71.

(120) Lal S, Mahajan A, Chen WN, Chowbay B. Pharmacogenetics of target genes across doxorubicin disposition pathway: a review. *Curr Drug Metab* 2010 January;**11**(1):115-28.

(121) Lee SC, Xu X, Lim YW, Iau P, Sukri N, Lim SE et al. Chemotherapy-induced tumor gene expression changes in human breast cancers. *Pharmacogenet Genomics* 2009 March;**19**(3):181-92.

(122) Stolboushkina EA, Garber MB. Eukaryotic type translation initiation factor 2: structure-functional aspects. *Biochemistry* (Mosc) 2011 March;**76(**3):283-94.

(123) Bonzo JA, Ferry CH, Matsubara T, Kim JH, Gonzalez FJ. Suppression of hepatocyte proliferation by hepatocyte nuclear factor 4alpha in adult mice. *J Biol Chem* 2012 March 2;**287**(10):7345-56.

(124) Albihn A, Johnsen JI, Henriksson MA. MYC in oncogenesis and as a target for cancer therapies. *Adv Cancer Res* 2010;107:**163-**224.:163-224.

(125) Westermark UK, Wilhelm M, Frenzel A, Henriksson MA. The MYCN oncogene and differentiation in neuroblastoma. *Semin Cancer Biol* 2011 October;**21**(4):256-66.

(126) Vita M, Henriksson M. The Myc oncoprotein as a therapeutic target for human cancer. *Semin Cancer Biol* 2006 August;**16**(4):318-30.

(127) Faneyte IF, Schrama JG, Peterse JL, Remijnse PL, Rodenhuis S, van de Vijver MJ. Breast cancer response to neoadjuvant chemotherapy: predictive markers and relation with outcome. *Br J Cancer* 2003 February 10;**88**(3):406-12.

(128) Brigstock DR. Regulation of angiogenesis and endothelial cell function by connective tissue growth factor (CTGF) and cysteine-rich 61 (CYR61). *Angiogenesis* 2002;**5**(3):153-65.

(129) Chien W, O'Kelly J, Lu D, Leiter A, Sohn J, Yin D et al. Expression of connective tissue growth factor (CTGF/CCN2) in breast cancer cells is associated with increased migration and angiogenesis. *Int J Oncol* 2011 June;**38**(6):1741-7.

(130) Koike Folgueira MA, Brentani H, Carraro DM, De Camargo Barros FM, Hirata Katayama ML, Santana de Abreu AP et al. Gene expression profile of residual breast cancer after doxorubicin and cyclophosphamide neoadjuvant chemotherapy. *Oncol Rep* 2009 October;**22**(4):805-13.

(131) Lipscomb EA, Mercurio AM. Mobilization and activation of a signaling competent alpha6beta4integrin underlies its contribution to carcinoma progression. *Cancer Metastasis Rev* 2005 September;**24**(3):413-23.

(132) Nikolopoulos SN, Blaikie P, Yoshioka T, Guo W, Giancotti FG. Integrin beta4 signaling promotes tumor angiogenesis. *Cancer Cell* 2004 November;**6**(5):471-83.

(133) Arnaldez FI, Helman LJ. Targeting the insulin growth factor receptor 1. *Hematol Oncol Clin North Am* 2012 June;**26**(3):527-42.

(134) D'cunja J, Shalaby T, Rivera P, von BA, Patti R, Heppner FL et al. Antisense treatment of IGF-IR induces apoptosis and enhances chemosensitivity in central nervous system atypical teratoid/rhabdoid tumours cells. *Eur J Cancer* 2007 July;**43**(10):1581-9.

(135) Gooch JL, Van Den Berg CL, Yee D. Insulin-like growth factor (IGF)-I rescues breast cancer cells from chemotherapy-induced cell death--proliferative and anti-apoptotic effects. *Breast Cancer Res Treat* 1999 July;**56**(1):1-10.

(136) Sardao VA, Oliveira PJ, Holy J, Oliveira CR, Wallace KB. Morphological alterations induced by doxorubicin on H9c2 myoblasts: nuclear, mitochondrial, and cytoskeletal targets. *Cell Biol Toxicol* 2009 June;**25**(3):227-43.

(137) Medic S, Ziman M. PAX3 across the spectrum: from melanoblast to melanoma. *Crit Rev Biochem* Mol Biol 2009 June;**44**(2-3):85-97.

(138) Robson EJ, He SJ, Eccles MR. A PANorama of PAX genes in cancer and development. *Nat Rev Cancer* 2006 January;**6**(1):52-62.

(139) Underwood TJ, Amin J, Lillycrop KA, Blaydes JP. Dissection of the functional interaction between p53 and the embryonic proto-oncoprotein PAX3. *FEBS Lett* 2007 December 22;**581**(30):5831-5.

(140) O'Malley S, Su H, Zhang T, Ng C, Ge H, Tang CK. TOB suppresses breast cancer tumorigenesis. *Int J Cancer* 2009 October 15;**125**(8):1805-13.

(141) Cheang MC, Chia SK, Voduc D, Gao D, Leung S, Snider J et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *J Natl Cancer Inst* 2009 May;20;**101**(10):736-50.

(142) Rakha EA, Reis-Filho JS, Ellis IO. Basal-like breast cancer: a critical review. *J Clin Oncol* 2008 May;20;**26**(15):2568-81.

Appendix A: Reagents and equipment

Table 14. Reagents and equipment used in this thesis.

Name	Supplier	Catalog nr./Part nr.
Spike-Mix	Agilent Technologies	5188-5282
Dilution Buffer	Agilent Technologies	5188-5282
T7 Promotor Primer	Agilent Technologies	5190-2305
5X First Stand Buffer	Agilent Technologies	5190-2305
0.1 M DTT	Agilent Technologies	5190-2305
10 mM dNTP Mix	Agilent Technologies	5190-2305
AffinityScript RNase Block Mix	Agilent Technologies	5190-2305
5X Transcription Buffer	Agilent Technologies	5190-2305
NTP Mix	Agilent Technologies	5190-2305
T7 RNA Polymerase Blend	Agilent Technologies	5190-2305
Nuclease-free Water	Agilent Technologies	5190-2305
Cyanine 3-CTP	Agilent Technologies	5190-2305
RNeasy Mini Spin Columns	Qiagen	74106
Collection Tubes	Qiagen	74106
RNase-free Buffers (RLT, RPE)	Qiagen	74106
2X Hi-RPM Hybridization Buffer	Agilent Technologies	5188-5242
25X Fragmentation Buffer	Agilent Technologies	5188-5242
10X Gene Blocking Agent	Agilent Technologies	5188-5242
Microarray hybridization chamber assemblies	Agilent Technologies	G2534-60001
Microarray Gasket slide	Agilent Technologies	G2534-60011
Human GE 4x44 Microarray	Agilent Technologies	G4112F
Gene Expression Wash Buffer 1	Agilent Technologies	5188-5327
Gene Expression Wash Buffer 2	Agilent Technologies	5188-5327
Triton X-102 (10%)	Agilent Technologies	5188-5327
RNaseZAP	Applied Biosystems	AM9780
RNase-free water	Invitrogen	10977-035
Agilent RNA 6000 Nano Chip	Agilent Technologies	5067-1511
Agilent RNA 6000 Ladder	Agilent Technologies	5067-1511
RNA Nano Dye Concentrate	Agilent Technologies	5067-1511
Agilent RNA 6000 Nano Marker	Agilent Technologies	5067-1511
Agilent RNA 6000 Nano Gel Matrix	Agilent Technologies	5067-1511
Filters for Gel Matrix	Agilent Technologies	5185-5990
Chip Priming Station	Agilent Technologies	5065-4401
Agilent Syringe Kit	Agilent Technologies	G2938-68706
Ethanol (>96%)	Antibac AS	6000068