Gene expression profiling in breast cancer. A link between biology and clinical decision making
Hannemann, J.

Citation for published version (APA):
Chapter 1

Introduction

Breast cancer is the leading malignancy in women worldwide and the number of breast cancer cases is still increasing. About 10% of the women in the Netherlands will develop breast cancer during their lifetime, 70% of them above the age of 50. Like all malignancies, breast cancer arises as a result of the accumulation of genetic alterations, most importantly deregulation of the expression of oncogenes and tumor suppressor genes. As a consequence this will lead to highly proliferating cells that lose their differentiation and have the ability to become invasive and metastatic.

There are various genetic pathways that have been identified and it has become clear that breast cancer represents a heterogeneous disease. As for most other cancers, also breast cancer can be classified based on the cell of origin. Most of the breast cancer cases are histologically classified as invasive ductal carcinomas (IDC) not otherwise specified (nos) that develop from their non-invasive precursor lesion ductal carcinoma in situ (DCIS). The second most frequent histological breast cancer type is invasive lobular carcinoma, which develops from lobular carcinoma in situ (LCIS) and a number of other, relatively rare histological types of breast cancer are known. This heterogeneity of breast cancer is also reflected in the variable clinical courses of the disease. Some patients will develop metastases at an early stage, other tumors will never metastasize. Treatment that is effective in one patient may not show the same effectiveness in other patients with a similar tumor type.

Since the effectiveness of treatment differs between individual patients, much effort is being invested in the identification of new prognostic and predictive markers. Prognostic factors are important to estimate patient’s outcome and can be used to decide which patients will need additional adjuvant systemic treatment. On the other hand, predictive factors indicate which treatment is most effective for an individual patient. The implementation of predictive factors in clinical decision making will help to ensure that only patients that are likely to benefit from a specific treatment will receive this specific therapy. With respect to chemotherapy, no single or multigene predictor has been widely accepted yet.

Thus far, clinical and pathological factors guide important decisions in the treatment of breast cancer patients. The present clinical and pathological factors are not accurately reflecting the marked heterogeneity in the prognosis and responsiveness to various therapies. While much knowledge on the genetic mechanisms underlying breast cancer development and progression have been unraveled in the last 20 years, there still are many unanswered questions with respect to the differences in biological behavior. Approximately 10 years ago, microarrays have been established as a high throughput method to analyze the expression of thousands of mRNAs simultaneously. Many of the principles of modern microarrays have been developed in the late 1980s, when cloned cDNAs were spotted on membrane filters, hybridized, and used to quantify mRNA expression. The major improvement came in the 1990s, when the first papers were published using a two-color, internally comparative hybridization technique that allows the analysis of the relative abundance of thousands of mRNAs in one experiment. The possibility to analyze the expression of thousands of genes in one experiment, instead of performing single marker studies, has provided a powerful tool to gain new insights in cell biology and tumour behavior that will help to develop diagnostic tools that can enter clinical routine.

These microarray based gene expression studies leaves the researcher with the generation of a very high number of data points. There are several statistical methods to analyze such data sets: Unsupervised approaches make use of the gene expression information to identify similarities in the pattern of gene expression. Unsupervised hierarchical clustering is often used as a first analysis to order samples according to their overall gene expression profile, so that the underlying structure of the data can be recognized. The output of unsupervised methods is primarily a description of the relationship between genes or samples. In general, clustering is an excellent choice for initial data analysis, especially if
the experiment was designed to be exploratory. Unsupervised methods are useful to identify new subgroups of samples or functional relationships between genes that were previously unknown or when only limited additional information on the samples is available.

The second class of analyses comprises a number of supervised approaches that have a common starting point: Existing information is used to define subgroups of samples and to identify the genes that can classify these subgroups at the molecular level. Usually, the data are divided into a training and a validation set. The training set is used to identify the optimal classifier. This gene signature is then tested with the samples of the validation set to estimate the predictive accuracy of the gene expression classifier. A third possibility is the identification of gene signatures in vitro using tissue culture or cultured cell lines. These experiments are especially useful if functional relationships have to be investigated that can not be tested directly in human tumors. The results obtained in the in vitro situation can then be validated in clinical tumor samples.

Studies using unsupervised hierarchical cluster analysis of gene expression data have led to the identification of several molecular subtypes of breast cancer. Based on these differences in gene expression patterns, at least four distinct groups of breast cancer can be recognized: the basal-like tumors that are mainly negative for the expression of the estrogen receptor, the progesterone receptor and HER2 and are often referred to as triple-negative tumors; the ERBB2-like tumors that are characterized by their increased expression of genes of the HER2 amplon including the HER2 gene itself; and two luminal-like subtypes (luminal A and luminal B) that are mainly ER-positive. Beyond differences in the expression of ER or HER2, these tumors show distinct gene expression patterns that may reflect different cell types of origin. It has been shown that these molecular subtypes have different clinical outcomes and respond differently to chemotherapy.

Although unsupervised approaches seem to be less biased, the possibility to identify informative molecular details of clinical subgroups is enhanced when all additional available clinical information is included in the analysis. The use of microarrays and the improvement of platforms have revealed a number of prognostic and predictive signatures. The future challenge is the evaluation of the contribution of the different levels of molecular and clinical data to the prediction of prognosis and responsiveness to therapy.

The aim of this thesis is first, to attempt to unravel mechanisms underlying chemotherapy responsiveness in breast cancer at the molecular level and second, to use gene expression profiling to understand the biology of ducal carcinoma in situ (DCIS) and the role of HER2 and E-cadherin in breast cancer development and progression.

After diagnosis, breast cancer is usually treated with a combination of surgery, radiation therapy and adjuvant systemic treatment. Adjunct systemic treatment can consist of chemotherapy and/or hormonal treatment and/or targeted therapy based on the molecular make up of the tumor. Improved knowledge of the tumor biology, especially at the molecular level, may have consequences for the treatment of individual patients, improving outcome for patients with breast cancer. It is therefore indispensable to investigate how treatment of patients can be individualized in the most effective way. To reach this aim, much additional research on breast tumor markers for classification and treatment response prediction is needed.

An overview about what is currently known about response predicting molecular markers is given in chapter 2. This chapter describes the current knowledge of well established markers such as estrogen receptor status, HER2/neu gene amplification status, and proliferation markers as well as recently described multigene predictors for response to chemotherapy.

The studies presented in this thesis embrace a number of different approaches. In chapter 3 and 4 studies of gene expression profiling to identify predictive gene expression profiles for the response to specific neoadjuvant chemotherapy regimens are described.

Chapter 5 describes a potential single predictive marker—topoisomerase IIα—and its predictive value in breast cancer patients receiving anthracycline-based chemotherapy within a large randomized trial. Topoisomerase IIα is the molecular target of anthracyclines. The gene is located near to and often co-amplified with the HER2/neu gene. It has therefore been suggested that amplification of topoisomerase IIα and not HER2/neu amplification is the actual mechanism underlying the observed sensitivity of HER2 positive tumours to anthracyclines. In a retrospective analysis we investigated whether this hypothesis can be confirmed in a large number of patients.

Ductal carcinoma in situ (DCIS) of the breast is the most common non-invasive form of breast cancer and it is assumed that most DCIS cases will progress to invasive breast cancer. This process can take years and the mechanisms of progression are not well understood. In chapter 6 we investigated the gene expression profiles of 40 in situ and 40 invasive samples to identify specific differences in gene expression between invasive and non-invasive ductal carcinomas and
differences between well and poorly differentiated DCIS samples.

Another approach to increase the knowledge on tumor behavior is to \textit{in vitro} manipulate pathways that are known to be involved in the development or progression of breast cancer. In chapter 7 we describe the effects of knocking down E-cadherin and HER2/neu expression in several breast cancer cell lines and the subsequent global changes in gene expression.