Improved classification of breast cancer by analysis of genetic alterations and gene expression profiling
Horlings, H.M.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Introduction and outline of this thesis
Introduction and outline of this thesis
Introduction and outline of this thesis

Breast cancer is the most prevalent malignant disease in women worldwide, and the number of breast cancer patients is still increasing. Annually over one million women worldwide are diagnosed with breast cancer, and over 400,000 women die from this disease each year [1]. In the Netherlands 12,171 women were diagnosed with breast cancer in 2005, and 3,335 died of this disease [2].

As a result of the increasing number of choices for surgical, radiotherapy and adjuvant or neoadjuvant systemic treatment, there is a growing possibility and need for patient-tailored therapy. The prognostic and predictive factors on which the treatment choices are presently based include clinical and pathological characteristics. Prognostic factors (defined as factors that predict the course of the disease) and predictive factors (defined as factors that predict response to specific therapies) are used to guide treatment in individual patients [3]. With increasing knowledge of specific genetic alterations and gene expression profiles of tumors and the prognostic and predictive value of these genetic tumor characteristics, more refined patient therapy is starting to be possible. Currently, axillary lymph node status, tumor size, histological grade, histological type, HER2 status, and hormone receptor status are still the most important factors guiding treatment [3].

There is marked heterogeneity in breast cancer. This heterogeneous nature of breast cancer is for example reflected in the wide variety of histological breast cancer types. The World Health Organization (WHO) describes at least 18 different histological breast cancer types [4]. Invasive ductal carcinoma not otherwise specified (IDC NOS) accounts for the large majority of breast cancers (i.e. 50–80%) [4]. IDC NOS is a diagnosis by default, being defined by the WHO as a tumor that fails to exhibit sufficient morphological characteristics to be classified into one of the histological special types. Approximately 25% of invasive breast cancers are recognized as ‘special types’, and characterized by distinctive growth patterns and cytological features [4-6]. Although some of the special types are associated with prognosis, histological typing in general is a weak prognostic marker of disease outcome.

The need to recognize genetically defined subtypes of breast cancer is enhanced by the increasing availability of specific and more effective therapy regimens. A key example of these advances is the recognition of HER2 positive status in breast cancers, which predicts the response to HER2 targeted agents such as Trastuzumab (Herceptin®) [7, 8]. A significant improvement in treatment outcome could be expected from better selection of patients for specific therapies. Elucidating the molecular mechanisms of primary or acquired drug resistance can be critical to identify tumors that fail to respond to therapy, which may help design of more efficient treatment protocols [9].

The heterogeneity of breast cancer with distinctive biological characteristics and clinical behavior is also reflected at the molecular genetic level. Breast cancer, like all malignancies, arises from a multistep process of genetic alterations that affects the function of individual genes and cellular processes, and these steps are driven by alterations in oncogenes and tumor suppressor genes [10]. Activation of a proto-oncogene into an oncogene can be
due to (A) mutation (i.e. substitutions of one base pair by another), (B) DNA copy number increases from two copies present in a normal diploid genome to several copies (i.e. increased expression by gene amplification producing a “gain of function” in the affected tumor cell), (C) chromosomal translocation (i.e. producing a corresponding fusion gene) or (D) mutation of microRNAs (i.e. oncomir). Another important genetic change leading to breast cancer initiation or progression is the inactivation of tumor suppressor genes. Inactivation of a tumor suppressor gene by mutation of one tumor suppressor gene allele and complete loss of the second allele producing a “loss of function” in breast epithelial cells is the mechanism most commonly identified in breast cancer [11]. Below we have summarized important genetic alterations in oncogenes and tumor suppressor genes and epigenetic alterations in breast carcinomas; these carcinomas lead to changes in the function or quantity of important growth regulatory proteins.

**Mutations** The most frequent germ-line (inherited from parents and transmitted to offspring) mutations leading to a strongly increased breast cancer risk are mutations in BRCA1 and BRCA2 [12, 13], and the most frequent somatic (i.e. acquired in the tumor cells) mutated genes in breast cancer are TP53 and PIK3CA [14, 15].

**Translocations** and their corresponding fusion genes are very rare. Only a few have been identified in carcinomas [16], although this might be due to a lack of technology to discover these rearrangements and more translocations may be remain to be discovered in carcinomas [17]. Only a few translocations and their fusion genes have been found in breast carcinomas and these translocations are associated with rare special types of breast carcinomas. For example, Tognon et al [18] showed that secretory carcinomas express the t(12;15)(p13;q25) ETV-6NTRK3 fusion gene. More recently, adenoid cystic carcinomas, also a special type of breast carcinomas, have been shown to harbor a t(6;9)(q22–23;p23–24) translocation corresponding to the fusion of the MYB oncogene to the transcription factor gene NFIB [19].

**Loss of heterozygosity** (LOH) LOH indicates the loss of one of the two alleles that are heterozygous at a locus. LOH on 1p, 1q, 3p, 7p, 11p, 13q, 16q, 17p, 17q, and 18q are reported to take place relatively frequently in breast cancer [20-23]. LOH may be caused by inactivation of a tumor suppressor gene, followed by loss of the wildtype allele; LOH may also lead to aberrant growth by a gene dosage effect (i.e. reduced expression of one or more genes in the deleted chromosomal region).

**DNA copy alterations** Microarray based comparative genomic hybridization (CGH) offered a fast and refined approach to screen whole genomes for a detailed analysis of DNA copy alterations [24]. In a CGH experiment, total genomic DNA is isolated from test and reference cell populations, differentially labelled, and hybridized to a representation of the genome that allows the binding of sequences at different genomic locations to be distinguished. Microarray based comparative genomic hybridization (aCGH) facilitate researchers to screen whole genomes for a detailed analysis of DNA copy alterations [24]. Frequently observed genomic alterations in breast carcinomas include gain of 1q, 8q, 16p, 17q, 20q, loss of 13q, 16q and 17p[25]. Neve et al. [26] searched by integrated use of PubMed, OMIM, Swissprot, NCICB, Geneontology databases for genomic abnormalities in breast carcinomas leading to the stable alteration of gene function. Using this approach, they identified 55 genes that were
found to have undergone copy number alterations in a proportion of breast carcinomas. For example, DNA amplification harboring known oncogenes were found in the regions of 8p11 (FGFR1) [27], 8q24 (MYC) [28], 11q13 (CCND1) [29, 30], 17q21 (ERBB2, GRB7) [28, 31, 32], and 20q13-ter (ZNF217) [33]. Deletions with known tumor suppressor genes include 13q12 (BRCA2) [34], 17p13 (TP53) [35], 17q21 (BRCA1) [36] and 16q22 (CDH1; is lost in invasive lobular carcinomas) [37, 38]. aCGH studies in breast cancer have shown that DNA copy number changes are associated with different clinical and pathological characteristics and some patterns of genomic changes have been related to different outcomes, reviewed by J.S. Reis-Filho in [25] and references herein. Since genetic alterations are the cause of cancer development, it is expected that the combination of specific genetic alterations in tumors is predictive of clinical behavior. However, until now, the use of genetic alterations as prognostic factors in breast cancer has been very limited. Most genetic alterations occur in only 10-20% of breast carcinomas; if breast carcinomas tumors are defined on the basis of combinations of genetic alterations, these subgroups become very small.

**Gene expression profiling** Given that a single genetic alteration can affect the expression at both mRNA and protein levels, gene expression profiling using micro-array analysis provides a method that can analyze the state of many regulatory pathways that may be affected by genetic alterations. This technique enabled researchers to analyze gene expression patterns in tumor samples for ten thousands of genes in one single experiment. In recent years, studies using gene expression profiling by micro-array analysis have demonstrated that breast tumors can be classified according to their expression patterns, and at least four “molecular subtypes” of breast cancer have been recognized: luminal A, luminal B, HER2 and Basal-like. Additionally several prognostic gene expression profiles have been identified: a 70-gene prognosis signature [39], a 76 gene prognosis signature [40], the “wound response signature” [41], chromosomal instability signature and [42] genomic grade index [43]. These signatures are strong independent prognostic factors. In spite of this progress in prognostic classification, the underlying mechanisms that drive these gene expression patterns remain unknown.

**MicroRNAs** A new class of genetic regulators, microRNA (miRNAs) may also play a role in cancer development and progression. They represent a large family of small non-coding RNAs [44, 45]. The products of these genes consist of a single RNA strand of 19-24 nucleotides, and they guide the cleavage and/or translational repression of fully or partially sequence complementary target mRNAs and can indirectly block protein translation [45, 46]. The expression of oncogenes can also be regulated by microRNAs [47]. Mutations in such microRNAs (known as oncomirs) can lead to activation of oncogenes [48]. miRNAs are found to be dysregulated in cancer, including breast cancer (reviewed in Zoon et al. [49]), and miRNA profiling in a variety of tissues and cancers has identified cell type-specific miRNAs [50, 51], including some profiles that correlate with prognosis [52, 53]. Current evidence regarding the role of miRNAs in breast cancer is limited, although in the past few years various studies have reported differential expression in distinct classes of breast carcinomas based on miRNA microarrays or Taqman assays (reviewed by Ruepp et al. [54] and Vergehese et al.[55]. For example, miR-21, miR-10b, and miR-27a have been reported to be upregulated and miR-206, miR-17-5p, miR-27b, miR-125a and b, miR-200c, miR-126, miR-335, and let-7 family to be downregulated in breast carcinomas.
Introduction and outline of this thesis

Epigenetic alterations In addition to genetic alterations, epigenetic alterations are observed in cancer cells. They affect the expression of a locus or chromosome without changing the underlying DNA sequence [56]. DNA methylation and histone modifications are the most important epigenetic alterations that affect gene expression in cancer [57]. DNA methylation and global patterns of histone modifications have been proposed as prognostic markers in breast cancer patients [58-60]. Epigenetic changes can be reversed as opposed to mutations, and this might give new therapeutic options for patients with breast cancer. These changes are potentially reversible by a number of small molecules, DNA methyltransferases (DNMTs) and histone deacetylase (HDACs) inhibitors, and in combination with traditional chemotherapy and radiotherapy, these novel agents may have potential for treatment of breast cancer patients[61].

Next generation sequencing The arrival of massively parallel sequencing (also known as next-generation sequencing) may provide qualitative and quantitative information about any type of nucleic acid in a given sample with a still increasing throughput [62-66]. Massively parallel sequencing methods provide millions of short reads of 21 to 400 base pairs, compared to the long reads generated by conventional (Sanger) sequencing [67]. This technology can discover a full range of germline and somatic mutations and has several applications: 1) Full genome sequencing for discovery of mutations, copy number alterations and genome rearrangements; 2) analysis of transcriptomes and small noncoding (ncRNA) discovery and profiling; 3) germline DNA sequencing for gene association studies 4) DNA methylation profiling by bisulfite DNA sequencing; 5) mapping histone modifications and the locations of DNA-binding proteins; 6) mapping out the positions of nucleosomes and other determinants of DNA accessibility [62]. These multiple applications and uses of massively parallel sequencing may lead to a shift in the way breast cancers are classified. Breast tumors may be classified according to their genetic alterations, their activated pathways and most importantly, according to the treatment regimens to which they are sensitive. So far studies that used conventional sequencing of breast carcinomas showed that there are a high number of genes that are rarely mutated and only a few genes that are frequently mutated (TP53 and PIK3CA) [14, 15]. More recently three studies used next generation sequencing to characterize the whole genome of breast carcinomas. Shah et al. [68] characterized the genome of a metastatic site of an invasive lobular breast carcinoma. They found 32 non-synonymous coding mutations in the metastatic site of which 5 (ABCB11, HAUS3, SLC24A4, SNX4 and PALB2) were also present in the primary tumor removed 9 years earlier at diagnosis. Combined analysis of DNA massively sequencing with RNA sequencing revealed two new RNA-editing events that recode the amino acid sequence of SRP9 and COG3.

A second study performed by Ding et al [69] used a paired-end sequencing strategy, and generated 130.7, 124.9, 111.8 and 149.2 billion base pairs of sequence data from genomic DNA derived from blood, primary triple negative breast tumor, brain metastasis and xenograft derived from the primary tumor [69]. They found 50 novel somatic point mutations and small indels in coding sequences, non coding RNA genes and splice sites as well as 28 large deletions, 6 inversions and 7 translocations. 55% of coding point mutations were C/G to T/A transitions. Interestingly 56% C/G to T/A transitions were observed by Shah et al. [68] in a lobular breast carcinoma. 48 out of 50 point mutations were detectable in all three tumors, 20 out 48 showed relatively comparable frequencies across the 3 tumours, 26
were significantly enriched in the metastasis and/or xenograft, and 2 sites with significant enrichment in the primary tumor. The metastasis contained two de novo mutations (SNED1, FLNC) and a large deletion (CTNNA1) not present in the primary tumor. The xenograft retained all primary tumour mutations. The different amount of mutation frequencies between primary and metastasis and xenograft suggest that there is a substantial genetic heterogeneity in the cellular population at the primary site and that only a few cells have the capacity to metastasize. Of course larger numbers of primary, metastatic and normal tissue trios are needed to test the statistical significance of these observations [69].

More recently Stephens et al. [70] used a paired-end sequencing strategy to identify somatic rearrangements in 24 breast cancer genomes, including 15 primary tumors and 9 immortal cell lines. They found more rearrangements than previously appreciated. Most rearrangements are found to be intrachromosomal. Tandem duplications are most likely the most common subclass of mutations and are often found in estrogen and progesterone negative breast tumors, whereas estrogen positive tumors have few rearrangements. Many new in-frame fusion genes were identified, although none were found to be recurrent and are therefore likely to be passenger events and not driver mutations. Much larger series will be required to investigate the likelihood of recurrent cancer causing rearrangements in breast tumors.

**Aim and outline of this thesis**

Combining assessment of genetic alterations with established gene expression signatures has the potential to improve biological understanding of breast cancer development and progression and may also result in improved prediction of prognosis and response to breast cancer therapy [71-73]. The aim of this thesis is to investigate the use of genetic alterations (e.g. mutations, DNA copy number alterations, gene amplification and gene deletion, and miRNA genes) and gene expression profiling of breast tumors to improve the categorization of breast cancer, and to better understand the biological mechanisms underlying the histological and molecular subtypes of breast cancer. 

**Chapter 1** is a general introduction of genetic breast cancer research and the different techniques to analyse genetic alterations in breast cancer. **Chapter 2** is a review of microarray analysis focusing on its applications in clinical oncology. Different bioinformatics approaches for the analysis of cancer gene expression profiles (e.g. unsupervised, data driven or supervised and hypothesis driven analysis) are discussed in this chapter.  

**Chapter 3** describes a two-class supervised algorithm for the analysis of DNA copy number data and the identification of regions of chromosomal alteration, SIRAC algorithm (Supervised Identification of Regions of Aberration in aCGH data sets), to identify DNA amplifications or deletions across tumors that may reveal key genes involved in cancer and improve our understanding of the underlying biological processes associated with breast carcinomas.  

In **Chapter 4** we report DNA copy number data of 68 human breast carcinomas for which gene expression and clinical data were available. We identified regions of chromosomal alterations that are associated with genes in the altered regions for which the expression level is significantly correlated with the copy number and the results were validated in publicly available data sets. **Chapter 5** is a letter to the editor of Nature Genetics on the
publication of Holst et al. [74]. In the original report of Holst et al: “Estrogen receptor alpha (ESR1) gene amplification is frequent in breast cancer” the authors investigated the clinical relevance of ESR1 (estrogen receptor alpha) gene amplification in human breast cancer and reported that the ESR1 gene is amplified in 21% of breast carcinomas. We report that in view of published array-based comparative genomic hybridization (aCGH) studies as well as our own unpublished aCGH data, the frequency of ESR1 gene amplification is less than 5%.

In Chapter 6 we aimed to refine the breast cancer type classification systems by analyzing a series of 11 histological special types [invasive lobular carcinoma (ILC), tubular, mucinous A, mucinous B, neuroendocrine, apocrine, IDC with osteoclastic giant cells, micropapillary, adenoid cystic, metaplastic, and medullary carcinoma] using immunohistochemistry and genome-wide gene expression profiling. Chapter 7 presents a study in which we used a large-scale RNA interference screen to discover genes involved in trastuzumab resistance in breast cancer. Chapter 8 describes a study intended to identify a microRNA gene expression profile as prognostic factor in triple negative breast cancer, as mRNA gene expression profiling was not successful in identifying prognostic subgroups. Secondly this study aimed to determine whether there are differences in microRNA expression in molecular subtypes of breast cancer. In Chapter 9 we tried to evaluate the association of PIK3CA and TP53 mutation status with various gene expression signatures; and to combine PIK3CA and TP53 mutation status with gene expression signatures in association with clinicopathological factors for patients with early stage breast cancer. Chapter 10 is a discussion of translating genomic alterations of breast cancer into clinical application (as a perspective accompanying the article by Russnes et al: “Genomic architecture characterizes tumor progression paths and fate in breast cancer patients”[75]. Chapter 11 is a summary of this thesis in English and Dutch.
Reference List


Introduction and outline of this thesis