Breast cancer prognostication and prediction in the postgenomic era

P. E. Lønning1,2*, S. Knappskog1,3, V. Staalesen1,3, R. Chrisantha1,3 & J. R. Lillehaug3

1Section of Oncology, Institute of Medicine, University of Bergen and 2Department of Oncology, Haukeland University Hospital; 3Department of Molecular Biology, University of Bergen, Bergen, Norway

Received 3 January 2007; accepted 11 January 2007

Expanding knowledge, together with implementation of new techniques, has fuelled the area of translational medical research aiming at improving prognostication as well as prediction in cancer therapy. At the same time, new discoveries have revealed a biological complexity we were unaware of only a decade ago. Thus, we are faced with novel challenges with respect to how we may explore issues such as prognostication and predict drug resistance in vivo. While microarray analysis exploring expression of thousands of genes in concert represents a major methodological advancement, discoveries such as the finding of different mechanisms of epigenetic silencing, intronic mutations, that most gene transcripts in the human genome are subject to alternative splicing and that hypersplicing seems to be a tumour-related phenomenon, exemplifies a complex pathology that may not be explored with use of single analytical methods only. This paper discusses clinical settings for studying drug resistance in vivo together with a discussion of contemporary biology in this field. Notably, each individual parameter which has been found correlated to drug resistance in vivo so far represents either a direct drug target or a factor involved in DNA repair or apoptosis. On the basis of these findings, we suggest drug resistance may be explored on the basis of upfront biological hypotheses.

Key words: breast cancer, microarray, mutation, prediction, splicing, therapy resistance

Introduction

Most of the classical studies in translational research relates to the issue of ‘prognostication’. Thus, a simple literature search on ‘prognosis’ and ‘breast cancer’ in concert revealed >10 000 published titles from the Web of Science (6 December 2006) during the last two decades.

Contrasting this large number of publications, and rarely mentioned, is the limitations of, or, rather, the specific requirements for a prognostic factor to be of value to clinical therapy. While development of more effective but also more toxic therapy forms underlines the need of identifying patients with a good prognosis that may not need adjuvant treatment, this does not imply that all patients with a poor prognosis may benefit from therapy. Thus, a parameter predicting poor prognosis may predict drug resistance as well. This relates both to individual prognostic factors and prognostic ‘gene profiles’ generated by techniques such as microarrays.

Thanks to the extended biological knowledge and, in particular, novel methods to measure biological variables such as the use of complementary DNA (cDNA) microarrays, we have for the first time the opportunity of identifying the mechanisms of drug resistance. This expanding biological knowledge, however, has also revealed a biological diversity we were unaware of just a decade ago. Sequencing of the human genome has revealed it to contain only ~30 000 genes; on the other hand, merging evidence has revealed the complexity of this genome with individual genes coding for multiple transcripts through mechanisms such as alternative splicing [1, 2] and several promoters with different ligand-binding preferences for each gene [3]. Also, for some gene products, several thousand DNA-binding sites [4, 5] or direct and indirect transcriptional regulation of several hundred downstream genes [6, 7] have been discovered.

This review is divided into three main parts; first, we will briefly discuss the issues of prognostication and prediction in a general perspective; secondly, we will discuss contemporary findings in molecular biology with particular relevance to cancer and, thirdly, we focus on how our present knowledge can help improving therapy.

Prognostication in breast cancer

The importance of—and differences between—a prognostic and a predictive factor is discussed elsewhere [8] and may be illustrated by examining relapse-free or overall survival curves from the Oxford meta-analysis [9]. For the examples here, we have chosen curves illustrating improvement in relapse-free survival with respect to adjuvant versus no adjuvant chemotherapy among lymph node negative and node-positive
breast cancer patients (Figure 1A) obtained from an earlier meta-analysis [10, 11].

The original message from this figure was to illustrate the benefits of adjuvant chemotherapy (Figure 1B) in a low as well as a high-risk population defined by lymph node status. The figure, however, illustrates in addition the prognostic value of studying lymph node status in early breast cancer patients. Looking at 10-year outcome (Figure 1C) for patients with metastases to axillary lymph nodes (one node or more) at the time of surgery versus those without, predicted risk of having a relapse within 10 years from operation with a specificity in the range of 70% and a sensitivity of ~60% [10, 11] in the population not having adjuvant chemotherapy (calculations on the basis of figures presented in the original publications). While the group of node-positive patients may be subsequently stratified on the basis of the number of metastatic nodes [12], we are still left with the problem; how to identify a group of patients with such a low risk of relapse that it makes adjuvant therapy redundant? The figures presented here illustrate the need to develop better prognostic factors beyond lymph node negativity.

While probably >100 individual prognostic factors have been reported in breast cancer, in general combining several factors adds marginal benefits. Thus, by combining three to four of the most important factors, little may be gained by adding additional information. There are two main reasons for this. First, many single prognostic factors are not ‘independent variables’; looking at traditional factors such as mitotic index, flow cytometry indicators, vascular invasion and tumour grade, there is a strong correlation between these factors [13–22]. The second issue is a potential confounding ‘predictive’ effect; patients stratified for relapse risk on the basis of the prognostic factor may, in addition, express a differential sensitivity to systemic treatment (see below).

Figure 1B illustrates the predictive value, or rather lack of such, of lymph node status with respect to effect of chemotherapy. The data clearly revealed that the benefit, determined by reduction in hazard ratio of having a relapse, was similar within both groups [9]. The fact that this was the case was not something predicted upfront; actually, it was realised long after implementation of adjuvant chemotherapy on the basis of risk profile. Considering other prognostic factors, we

![Figure 1](image-url)
may be less lucky. Thus, it may not be automatically inferred that a novel prognostic factor may not involve a predictive component or, said in other words, we may not take for granted upfront that the improvement with respect to percentage reduction in relapse rate due to therapy should be similar in the ‘high risk’ and ‘low risk’ patient populations. Thus, while TP53 mutation status as well as HER-2 amplification are both known to be associated with a poor prognosis in breast cancer [23–25], TP53 mutations predict reduced sensitivity to anthracycline-containing therapy [26, 27], while HER-2 amplifications confer low sensitivity to regimens such as the combination chemotherapy with cyclophosphamide, methotrexate and fluorouracil (CMF) [28].

The potential negative consequences of selecting patients for adjuvant treatment using TP53 mutation status or any other factor associated with drug resistance as a prognostic factor to define a ‘high risk group’ is depicted in Figure 1D. Perhaps the extreme consequence of selecting patients on the basis of relapse risk only without attention to a potential predictive role of the prognostic factor may be illustrated taking an example of tamoxifen treatment with tamoxifen as an example. While long-term risk of relapse remains similar among patients with oestrogen receptor (ER)-positive and ER-negative breast cancer not undergoing adjuvant therapy, short-term prognosis, also for patients not undergoing adjuvant endocrine therapy, is worse for the ER-negative ones [29, 30]. Thus, selecting patients for adjuvant tamoxifen treatment on the basis of risk profile alone paying no attention to knowledge about the biology of the ER could mean to treat ER-negative but not the ER-positive tumours with the anti-oestrogen; the outcome would be predictable.

While there is a need for better prognostic factors, no prognostic factor should be implemented for clinical use without a proper evaluation of its predictive power with respect to the kind of therapy under consideration. Therefore, to improve not only prediction but also prognostication, we need to understand potential associations between available parameters and drug resistance. This relates to multiple-gene signatures as well as to individual biological factors.

predicting sensitivity to therapy in breast cancer

Contrasting the panel of prognostic factors available, only a limited number of individual factors associated with chemotherapy sensitivity or resistance have been identified. For reasons discussed in detail below, we find it useful to discriminate between factors predicting ‘sensitivity’ and ‘resistance’ to therapy. The finding that, for example, expression of a factor predicts therapy resistance may not implicate that all tumours devoid of the same parameter are sensitive. Further, considering sensitivity, there is evidence that some factors, such as amplification of topoisomerase II, do not predict sensitivity in absolute terms but, rather, a graded sensitivity or responsiveness to anthracycline-containing chemotherapy (see below).

Considering factors that have been found correlated to resistance or drug sensitivity so far, they all have a potential biological explanation: (i) as specific target molecules for the therapy applied, (ii) as gene products modifying therapy response through DNA metabolism or (iii) as gene products participating in the execution of apoptosis and/or growth arrest/senescence.

Below is a summary of individual predictive factors found to be associated with response to different types of chemotherapy in vivo. Notably, with the exception of amplification of the topoisomerase II gene, for which merging evidence indicate diagnostic testing for therapy selection may be close to implementation [31–34], for none of the individual factors identified so far do we have data indicating clinical use in breast cancer in the near future. However, the findings described in details below are challenging, indicating we are approaching a conceptual understanding of biological mechanisms and pathways playing a key role in therapy resistance [35].

amplification of topoisomerase II

Anthracyclines seem to achieve their antitumour effects by affecting several intracellular processes, causing DNA intercalation as well as generation of free radicals [36–38]. A key mechanism, however, seems to be the inhibition of the DNA damage repair enzyme topoisomerase II (Topo-II) [39]. Notably, it was revealed more than a decade ago that anthracyclines administered in different doses in the adjuvant setting were associated with a ‘dose–response’ effect among tumours amplified for HER-2. This effect was not seen among HER-2-negative tumours [40]. This finding lead to suggestions that HER-2 may be associated with sensitivity to anthracycline therapy; yet, findings from in vitro studies did not support the concept [41]. However, the Topo-II is located on chromosome 17q12 in close proximity to HER-2, and the two genes are frequently, but not univocally, coamplified [32]. Thus, there is now merging evidence from studies in the neo-adjuvant as well as the adjuvant setting indicating that Topo-II amplification may sensitize tumours to treatment with anthracyclines in early breast cancer [31–34, 42–45]. For reasons (so far) unexplained, it seems that Topo-II deletions cause the same sensitization to anthracyclines that is observed for Topo-II amplifications [33].

TP53 mutations

In addition to its role in breast cancer, TP53 mutations or, rather, p53 protein immunostaining has been shown to be a prognostic factor in different malignancies, including soft tissue sarcomas, colorectal cancers, non-small-cell lung cancer, ovarian cancer and haematological malignancies among others [46–54]. Normally, p53, the protein coded by the TP53 gene, exists at very low concentrations in the cell and is not detected by immunostaining. p53 degradation occurs through MDM2 binding, followed by ubiquitination [55, 56]. As many mutated p53 variants may not bind to MDM2, current opinion is that an extended half-life with intracellular accumulation of mutated p53 explains increased levels of the protein allowing detection by immunostaining [57]. Thus, p53 immunostaining has been used as a surrogate marker for the existence of TP53 mutations.

There are, however, several pitfalls using this assumption. First, the possibility exist that in cells exposed to physiological stress there may be an accumulation of normal p53 protein.
Secondly, it is well recognised that in breast cancer mutated p53 may not always be detected by immunostaining [58–60]; about 30% of tumours harbouring TP53 mutations do not express immunostaining when exposed to the most commonly used antibodies; similar findings have been made in other cancers, such as cancer of the large bowel, endometrium and stomach [60]. In particular, we observed that many of the mutations associated with therapy resistance did not stain [59]. Some mutations, such as base insertions or deletions, may create a nonsense reading frame or an early stop codon, resulting in loss of the full-length protein product (Figure 2). Other mutations may lead to an unstable protein product as revealed for some protein products generated from mutations in the p53 upstream activators ATM and CHEK2 [61, 62]. Finally, mutated and wild-type p53 protein may exist in different three-dimensional folding, leaving the epitope exposed or unexposed [63–66].

Reviewing the literature with respect to TP53 as a predictive factor, most studies using immunostaining have not detected any correlation between p53 staining and response to chemotherapy in breast cancer [67]. In contrast, studies examining TP53 mutation status at the genomic level have reported an association between failure to anthracycline- or mitomycin C-containing chemotherapy and TP53 mutations in breast cancer [27, 59, 68]. Interestingly, a similar association has been recorded in different haematological malignancies [69–72].

While several studies have reported an association between TP53 mutations and resistance to anthracyclines and similar compounds, notably, evidence presented so far does not provide support for TP53 mutations to be used as a selection criterion with respect to therapeutic regimens. The underlying reason is the low sensitivity and specificity of such testing. A potential explanation to this finding is presented later in this paper, illustrating the need to evaluate functional pathways and not only individual genes in this setting.

**BRCA1, BRCA2 and the Fanconi proteins**

While gene products involved in DNA repair are described in the section below, due to their importance in breast cancer the BRCA1 and -2 are described separately, including the related Fanconi complex. Relevant to our discussion is the observation that somatic mutations affecting BRCA1 or -2 are rare in breast cancer.

While several studies have reported a poor prognosis in tumours arising in BRCA1 mutation carriers [73–75], the prognosis for patients harbouring BRCA2 mutations seems to be not much different from spontaneous breast cancers [76]. Interestingly, breast cancers appearing in BRCA1 carriers have

---

**Figure 2.** Different types of gene mutations and mechanisms of epigenetic silencing.
been reported to carry distinct gene profiles [77] and frequently to be associated with basal tumour-like characteristics [78], a gene profile known to be associated with a general poor prognosis [79, 80]. With respect to chemoresistance, anecdotic evidence has indicated improved sensitivity to anthracyclines in tumours expressing high mRNA levels of BRCA1 [81]; in contrast, low levels of BRCA2 messenger RNA (mRNA) has been found associated with enhanced sensitivity to taxanes [82]. The number of patients in these studies is low and these findings need further confirmation.

The Fanconi syndrome is a recessive syndrome characterised by congenital anomalies and anaemias that, in addition, includes an enhanced risk of developing haematological as well as non-haematological malignancies [83]. While the Fanconi protein complex exerts several functions, a major effect seems to be interactions with BRCA1 and -2 with respect to DNA repair. Recently, hypermethylation of FANCF, an upstream gene in the Fanconi gene complex was found associated with sensitivity to platinum drugs in ovarian cancer cell lines [84].

title other enzymes involved in DNA repair

DNA repair involves several mechanisms such as base excision repair, nucleoside excision repair, mismatch repair as well as O6-methylguanine DNA methyltransferase; the readers are referred to reviews on the subjects for details [85–87].

So far, there are little data on alterations in these enzyme systems with respect to breast cancer. However, there are several interesting findings in other cancer forms.

The mismatch repair system is responsible for removing incorrectly paired nucleotides in the genome [86]. The system contains several enzymes making heterodimers [86], and mutations affecting this system are responsible for inherited nonpolyposis cancer of the large bowel, the Lynch syndrome [88, 89]. While reported to be associated with a good prognosis in cancer patients not exposed to adjuvant chemotherapy [90], mutations in mismatch repair genes have been associated with poor sensitivity to 5-fluorouracil [91], underlining the fact that prognostic information cannot be used to make assumptions about therapy sensitivity.

Nucleotide excision is an important mechanism for the protection of cells from exogenous carcinogens. Thus, mutations affecting the nucleotide excision enzyme complex are the cause of Xeroderma pigmentosum [86], a syndrome associated with a very high risk of UV-induced skin cancers and other malignancies. The fact that this complex removes DNA-platinum adducts [92] offers an explanation why high mRNA levels of ERCC1, one of the enzymes in this complex, has been found associated with resistance to platinum compounds in both ovarian and gastric cancer [93, 94].

The O6-methylguanine DNA methyltransferase is responsible for removing large DNA adducts [95]. Interestingly, silencing of this gene through promoter hypermethylation has been reported to enhance sensitivity to alkylating agents such as BCNU and temozolomide in human glioblastomas [96–98].

While polymorphisms in the base excision pathway genes have been associated with breast cancer risk [99], we are not aware of any study reporting mutations or polymorphisms in these genes to be associated with drug sensitivity.

summary

The findings discussed here are exciting from a scientific point of view, in as much as they point to distinct biological mechanisms being the cause of drug resistance. However, with the exception of Topo-II amplification, none of these individual gene alterations identified so far has revealed a predictive power with respect to sensitivity as well as specificity warranting testing in a large-scale clinical setting. Nevertheless, they indicate biological explanations to drug resistance. Thus, these genetic changes may be considered lighthouses, or beacons, identifying functional pathways of major importance [35].

title pathologic changes leading to disturbed gene functions

The finding that changes in individual genes are associated with but not fully predictive for drug resistance mechanisms has lead to increased focus on the probability of testing not individual genes but, rather, multiple genes in concert using methods such as cDNA or oligonucleotide arrays. Before discussing such options, we will briefly consider the mechanisms leading to disturbed gene function.

somatic mutations

Somatic mutations may be considered the ‘classical way’ of gene inactivation. This may occur as single base defects (substitutions, leading to amino acid change in the coded protein), single base insertions or deletions or deletions or insertions of smaller or larger fragments (Figure 2). Obviously, any insertion or deletion of a number of base pairs not divisible by the number of three will lead to a shift of reading frame and, thus, a complete change of the amino acid sequence of the protein C-terminal of the affected codon. Such nonsense mutations often lead to truncated forms of the protein in question due to novel stop codons. Loss of wild type activity is the most common consequence.

Many single base substitutions may not lead to an amino acid substitution and, thus, remain nonfunctional polymorphisms, except for the rare cases where single base substitutions modulate mRNA stability or subcellular mRNA targeting. For many single base substitutions leading to a single amino acid substitution, this leads to a protein with a biological function deviating only moderately from the original protein. Looking at germline polymorphisms (defined by convention as a single base substitution occurring in >1% of the general population), for example the p53 arg72pro leads to a marginal change in biological effects [100]. In contrast, other single base substitutions in TP53 affecting critical protein domains may lead to nonfunctional proteins. We found single base substitutions such as the codon 249 AGG-GGG (Arg-Gly) to cause lack of p53 protein staining as well as to be associated with resistance to doxorubicin in primary breast cancer [59]. Most of the germline mutations detected in Li–Fraumeni families are single base substitutions [101], so is the case for ~50% of the germline mutations of the RB gene, identified in retinoblastomas [102]. Base substitutions also represent the most frequent form of defect identified in the CDKN2A gene in view, in as much as they point to distinct biological mechanisms being the cause of drug resistance. However, with the exception of Topo-II amplification, none of these individual gene alterations identified so far has revealed a predictive power with respect to sensitivity as well as specificity warranting testing in a large-scale clinical setting. Nevertheless, they indicate biological explanations to drug resistance. Thus, these genetic changes may be considered lighthouses, or beacons, identifying functional pathways of major importance [35].

title pathologic changes leading to disturbed gene functions

The finding that changes in individual genes are associated with but not fully predictive for drug resistance mechanisms has lead to increased focus on the probability of testing not individual genes but, rather, multiple genes in concert using methods such as cDNA or oligonucleotide arrays. Before discussing such options, we will briefly consider the mechanisms leading to disturbed gene function.

somatic mutations

Somatic mutations may be considered the ‘classical way’ of gene inactivation. This may occur as single base defects (substitutions, leading to amino acid change in the coded protein), single base insertions or deletions or deletions or insertions of smaller or larger fragments (Figure 2). Obviously, any insertion or deletion of a number of base pairs not divisible by the number of three will lead to a shift of reading frame and, thus, a complete change of the amino acid sequence of the protein C-terminal of the affected codon. Such nonsense mutations often lead to truncated forms of the protein in question due to novel stop codons. Loss of wild type activity is the most common consequence.

Many single base substitutions may not lead to an amino acid substitution and, thus, remain nonfunctional polymorphisms, except for the rare cases where single base substitutions modulate mRNA stability or subcellular mRNA targeting. For many single base substitutions leading to a single amino acid substitution, this leads to a protein with a biological function deviating only moderately from the original protein. Looking at germline polymorphisms (defined by convention as a single base substitution occurring in >1% of the general population), for example the p53 arg72pro leads to a marginal change in biological effects [100]. In contrast, other single base substitutions in TP53 affecting critical protein domains may lead to nonfunctional proteins. We found single base substitutions such as the codon 249 AGG-GGG (Arg-Gly) to cause lack of p53 protein staining as well as to be associated with resistance to doxorubicin in primary breast cancer [59]. Most of the germline mutations detected in Li–Fraumeni families are single base substitutions [101], so is the case for ~50% of the germline mutations of the RB gene, identified in retinoblastomas [102]. Base substitutions also represent the most frequent form of defect identified in the CDKN2A gene in
families at risk of malignant melanomas [103]. Current work is concentrating on methods to determine the pathogenic effect of different single amino acid substitutions [104].

For other genes, such as BRCA1 and -2, the most common germline mutations appear to be nonsense mutations resulting in a truncated protein [105].

Notably, larger deletions may be difficult to detect by standard DNA sequencing, and traditional methods, such as Southern blotting, are less specific, and requires substantial amounts of material, making it unfit for testing on tumour material with a limited tissue amount supply. Recently, improved methods, such as multiplex ligation amplification (MLPA) have become available [106–109]. Yet, deletions may escape detection if they are not properly covered by the probes [110]. In such cases, combined analysis on DNA and cDNA is highly beneficial as loss of parts of the coding material with a limited tissue amount supply. Recently, alternative splicing

mutations involving noncoding regions

Recently, there has been focus on mutations affecting noncoding regions. Considering intron mutations, both mutations directly affecting the splicing sites and deep intronic mutations may lead to aberrant splicing (see below), severely affecting the function of the translated protein [111, 112]. In addition, mutations affecting promoter areas may lead to either a nonfunctional or, in some cases, ‘hyperfunctional’ promoter sequences, as recently illustrated with respect to the MDM2 gene [113].

In rare cases, mutations may lead to creation of a new functional promoter. The latter is illustrated by the observation that an intronic single base substitution generated a novel promoter element allowing gene activation leading to a variant of thalassaemia [114].

alternative splicing

The last decade has focused in particular on the issue of alternative splicing, i.e. the generation of several processed RNA transcripts from a single primary RNA transcript (Figure 3 and 4). For most genes, this is manifested by synthesis of a main full-length transcript with the addition of several truncated variants. Alternative splicing is a physiological phenomenon; the readers are referred to recent comprehensive reviews on the subject [115]. While there is evidence that perhaps as much as 75% of transcribed genes in the human genome may be subject to alternative splicing [1], the phenomenon seems more extensive in cancer compared with normal tissue [116]. Considering MDM2 and Chk2, known to be of critical importance to p53 degradation and activation in response to DNA damage, respectively [56, 117], >40 and 60 alternative transcripts have been identified [118, 119]: for Chk2, the number of transcripts identified were higher in breast cancer compared with benign tissue. The finding of hypersplicing in cancer is likely due to mechanisms other than mutations in the splicing sites, as the splice variants often are detected together with the main transcript and no mutation may be identified. Interestingly, for MDM2, it seems that activation of the different promoters may lead to a different ratio between main and truncated RNA transcripts [120]. For the TP53 analogues TP63 and TP73, a family of truncated transcripts are generated [121], and functional active alternative transcripts for TP53 have been identified as well [121, 122]. For many alternative transcripts it is not known whether they are translated into a protein product or, if so, whether the protein may be functional, while for some genes, such as the TP53 analogue TP73, alternative transcripts coding for proteins with an effect antagonistic to the wild-type protein has been identified [123].

gene amplification

Chromosomal disturbance is a hallmark of cancer [124, 125]. Thus, for many genes such as Topo-II (see above), MDM2 [126] as well as cyclins D and E [127, 128] amplifications, and not mutations, are the pathological event recorded in neoplasias. Notably, amplification seems to implicate mRNA up-regulation in about two of three cases in breast cancer [129], underlining that in one of three cases amplification of a large gene segment may not be associated with mRNA up-regulation of each gene in that domain. The importance of gene amplification may be illustrated using Topo-II as an example (see above) as well as MDM2 and its sister gene MDMX. The main function of the MDM2 and MDMX gene products is to mediate p53 degradation, and amplification of MDM2, and more recently MDMX [130], has been revealed as an alternative way of p53 inactivation.

epigene\(t\)ic mechanisms

Gene imprinting is a critical physiological process, and the readers are referred to contemporary reviews on the subject [131, 132]. Genetic imprinting plays a crucial role in embryonic life but during the whole lifetime of the individual as well [133]. Basically, two main mechanisms; promoter hypermethylations [134] and histone deacetylations [135], have been discovered (Figure 2).

Importantly, deregulated promoter hypermethylation is a frequent form of gene inactivation in cancer, and hypermethylation has been detected in critical genes such as CDKN2A, the gene coding for the cdk inhibitor p16, in melanomas [136], and the DNA repair enzyme methylguanine DNA methyltransferase in glioblastomas. Interestingly, O\(^{-}\)methylguanine DNA methyltransferase promoter hypermethylation was found to correlate to response to temozolomide in this setting [97, 98].

While promoter hypermethylation is relatively easy to detect, we currently lack convenient methods for detecting protein deacetylations. However, substantial experimental evidence links histone deacetylation to gene transcriptional control [135, 137], and deacetylase inhibitors are currently in development for cancer therapy [138]. Thus, appropriate diagnosis not only of gene silencing but also the specific mechanism responsible may be a key issue for future targeted therapies.

exploring gene dysfunction using microarrays

The invention of microarrays, allowing investigations of multiple genes in concert, represents an important step
forward in our attempts to explore biological mechanisms of chemoresistance in cancer. Thus, microarray studies have been used to explore the biology of most tumour types. So far, two major goals with respect to breast cancer have been achieved:

First, using hierarchical clustering, Perou et al. [139] and Sørlie et al. [79] were able to show that breast cancers may be classified on the basis of global gene ‘signatures’. This classification has been found robust, confirmed in several studies [80, 140, 141]. The main finding probably was the identification of the ‘basal’, or ‘triple-negative’ class, revealing tumours lacking expression of the ER as well as HER-2 not only to be devoid of these factors but, actually, to represent a distinct class of tumours, identified by a specific gene profile.

Second was the finding that hierarchical [79] as well as supervised [142–144] clustering was able to reveal strong prognostic signatures. While the specificity of the Amsterdam 70 gene profile has been challenged [145], this gene profile as well as the ‘Rotterdam’ profile [146] provides prognostic information exceeding what is generated by lymph node status only.

The major danger of using such profiles selecting patients for therapy lies in the fact that we here, as for individual genes,
lack information whether it also predicts sensitivity to therapy. Thus, while the Amsterdam 70 gene signature [143] revealed prognostic information in both node-negative patients (in general not exposed to adjuvant treatment) as well as in node-positive patients (in general exposed to adjuvant treatment), it is not possible to infer from the data that the signature is not associated with sensitivity to therapy.

In contrasts, studies aiming at developing signatures predicting drug resistance have not been successful. While several studies have reported gene signatures associated with sensitivity to treatment with both taxanes [147, 148] and anthracycline-containing regimens [149, 150] as well as combinations containing both [149, 151–154], none of these signatures have come into general clinical use for a couple of reasons. First, these signatures have been developed in studies enrolling a limited number of patients only. Secondly, they consistently reported sensitivity as well as specificity predicting therapy outcome of ~70%–80%, which is below what should be required if a parameter is to be used to select patients for therapy.

Thirdly, while they report likelihood for a patient to respond to a certain treatment regimen, none of the studies at the same time explored the power of the signature predicting sensitivity to alternative treatment options. By defining a patient as chemoresistant we may avoid the toxic side-effects of a useless therapy; however, we may not allocate the patient to any alternative therapy without knowledge about its efficacy.

While the ‘21-gene’ signature developed by the NSABP was found associated with benefit from CMF in the adjuvant setting [155], its predictive value to taxane and anthracycline-containing chemotherapy in the neo-adjuvant setting did not exceed the power of other gene profiles [152].

Microarray data provide enormous amounts of information revealing gene expression of potentially all human genes in a single analysis. However, it is clear that such arrays may not identify all mechanisms of gene inactivation. This may be illustrated by a few examples (Figure 5). Assuming the probe on the array hybridises to the 5’ end of an mRNA transcript, pathological changes such as a single base mutation or even a deletion located downstream of the probing area would be overlooked. For single base substitutions within the area covered by the probe, the result is less predictable; the transcript may not bind the probe at all, but in other cases it may still bind partly or completely and, thus, the substitution may be overlooked. While arrays to explore splice variants have been constructed [1], as already mentioned, RNA splicing seems to be particularly extensive in cancer tissue [116, 119]. Thus, while in theory expression of each gene in the human genome could be quantified with an array containing 30 000 probes, to assess the full RNA splicing profile for each gene, we probably may need millions of probes on the arrays.

**Figure 4.** Splicing mechanism: splicing of exons is signalled by the consensus sequences: MAG|GTRAGT in the 3’ end of one exon and CAG|G in the 5’ end of the next exon. These sequences are brought together by the Spliceosome, looping the intron between out. Irregular splicing due to mutations in these consensus sequences are normally labelled mutations, and not alternative splicing. Alternative splicing is variation in which splice sites combine when all splice sites are intact.

**Figure 5.** Microarray signals are generated by hybridising a probe to the gene product [complementary DNA (cDNA)]. Alternative splicing, as well as different types of mutations may influence the result but may go undetected as well pending on the exact type of lesion as well as its location with respect to the area hybridising with the probe. The marks “X” and “-” indicate potential errors: “X” indicates wrongly detected transcript levels, while “-” signifies correct levels reported, but with other undetectable lesions present in the gene.
of rescue and survival is DNA repair, alternatively, lack of growth arrest/apoptosis caused by defects in the genes regulating these cellular functions.

The last two decades have provided enormous amounts of information regarding the mechanisms regulating growth control as well as the processes of apoptosis/senescence. Our challenge is how to implement this information to further our understanding of therapy failure. The idea of exploring disturbances in genes involved in apoptosis, as well as in DNA repair, such as the TP53 gene, as a cause of drug resistance in vivo, was on the basis of elegant preclinical experiments [157, 158]. While it is mandatory to take into account the results generated from experimental systems, the need for proper verification in vivo cannot be emphasized strongly enough. The story of the MDR1 gene, encoding the P-glycoprotein, as well as other genes coding for other membrane pumps, may depict this; while the theoretical concepts are well founded and data from experimental systems convincing [159, 160], in vivo studies correlating P-glycoprotein expression to drug sensitivity [161] have not supported the concept, and the results from clinical trials exploring pump inhibitors in general have been disappointing [162]. Thus, while the finding of gene signatures predicting resistance to different chemotherapy compounds across a panel of cell lines is interesting [163], the validity of these profiles to drug resistance in vivo remains to be confirmed.

Assessing tumour gene profiles using arrays detecting most of the genes in the human genome, biostatistical evaluation has indicated thousands of individual biological samples may be needed to generate robust gene signatures [164]. Suppose the average number of RNA splice variants for each gene to be five. Due to the significant sequence overlap between individual splice variants from a single tumour (Figure 5), this means that to evaluate expression of each out of five splice variants with confidence, we may probably need 10 probes, perhaps even more, for each gene on average. Thus, the number of spots on the array may increase above 300 000. Accordingly, the number of biological samples needed to generate robust gene expression signatures would in practical terms become unrealistic.

For this reason, we believe there is a need to limit the number of genes to be examined on the basis of biological hypotheses generated upfront through experimental work. As mentioned, each individual factor found associated with chemoresistance so far has been involved in specific biological processes; either they represent targets for the drug, alternatively they are associated with processes such as growth arrest/apoptosis, or DNA repair. Most of these execute their function as part of a protein complex with other factors. Like other genes, they are activated by certain upstream genes (at the transcriptional or posttranslational level) and activate downstream genes in functional cascades. Thus, it is likely to postulate that inactivation of other factors in the same complex, or genes involved up- or downstream in the functional cascade, may exhibit similar effects.

The issue of TP53 mutations in anthracycline and mitomycin C resistance may be taken as an example. For reasons discussed elsewhere [8], to study resistance factors we decided to compare tumours expressing primary resistance to the combined group of tumours obtaining an objective response or a stable disease (less than a 25% reduction and less than a 25% increase in the sum of the products of two perpendicularly measured diameters of all measured lesions and the appearance of no new lesions). For both therapeutic compounds studied, we detected an association with TP53 mutations affecting the DNA-binding L2 and L3 domains and resistance to therapy (Table 1). To explain these results we need to make certain assumptions. Considering the group of tumours harbouring TP53 mutations in the L2 or L3 DNA-binding domains that are sensitive to therapy, we hypothesise that apoptosis could be achieved by one of two pathways, one of which contains TP53. Thus, tumours being sensitive to therapy despite their TP53 mutation may have the second pathway intact, while tumours harbouring TP53 mutations and being resistant may have this second pathway inactivated. Secondly, tumours that are resistant to therapy despite expressing wild-type TP53 may have a gene located up- or downstream of TP53 inactivated in addition to inactivation of the alternative pathway. The different scenarios are depicted in Figure 6.

A key problem applying such a model is that we do not know which genes are involved downstream of p53 in the apoptotic pathway. Considering in vitro experiments, p53 has been shown to activate several hundred genes, depending on the experimental situation [6, 7]. However, this represents a huge reduction in the number of candidate genes to explore compared with use of a general ‘hierarchical’ clustering or ‘supervised’ analysis, in which cases most genes of the genome are involved.

Exactly similar assumptions may be applied to other systems exploring causes of resistance to other compounds. Most DNA repair mechanisms involve protein complexes; thus, it is likely that mutations in different genes involved in the same complex may achieve similar biological outcomes. In addition, each gene is subject to transcriptional stimulation by upstream factors.

The technical problems associated with identifying gene defects were discussed above. On the other hand, it may be argued that the type of gene defect may not be important when applying microarray analysis. Provided the gene defect leads to a nonfunctional gene, the effect should be detected in

### Table 1. Correlations between TP53 mutations affecting the DNA-binding L2/L3 domain and lack of response (progressive disease; PD) in patients with locally advanced breast cancer treated with doxorubicin weekly monotherapy [59] or mitomycin C and 5-fluorouracil [68]

<table>
<thead>
<tr>
<th>Therapy</th>
<th>TP53 status</th>
<th>Response</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>PR/SD</td>
<td>PD</td>
</tr>
<tr>
<td>Doxorubicin*</td>
<td>TP53 wild type or mutated ‘non’</td>
<td>67</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>L2/L3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-FU and Mitomycin C*</td>
<td>TP53 wild type or mutated ‘non’</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>L2/L3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TP53 mutated L2/L3</td>
<td>22</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>TP53 mutated L2/L3</td>
<td>3</td>
<td>6</td>
</tr>
</tbody>
</table>

PR, partial response; SD, stable disease; PD, progressive disease.

*P = 0.008.

*P = 0.006.
as much as genes located downstream in the pathway will remain underexpressed. Thus, we believe gene arrays including genes involved in certain biological cascades may be a way forward. However, looking at the example presented in Figure 6, each tumour expressed different ‘gene profiles’ with respect to the genes involved in the apoptotic pathway. Curiously, the two tumours (B and C) expressing the largest similarity in gene profile had a different clinical outcome. Thus, we believe that arrays must be used in combination not only with RT-PCR to verify gene expression levels but also gene sequencing, MLPA and examination of epigenetic events, such as hypermethylation status, to explore the exact nature of any possible gene defect.

**depicting the future**

Understanding chemo-resistance should represent a great leap forward in our attempts to cure not only breast cancer but also cancers in general. So far, the results of microarray studies have been disappointing and the results from studies on individual
genes less optimistic than what we expected a few years ago. Yet, we believe there is great hope for the future.

First, our basic knowledge in the field of cell growth control, apoptosis and tumour biology in general has been extended over the last decade in a way we could only dream of some decades ago.

Secondly, we have at hand an armamentarium of methods allowing us to explore tumour biology in great detail. Techniques such as microarray offer novel opportunities; yet, we need to put these methods into a biological context.

Thirdly, while studies on factors predicting drug resistance has not so far provided the tools we need for clinical application, we see a merging picture of individual gene alterations linked to drug resistance. Most importantly, the data generated are not individual, single observations only; we are starting to see a biological picture, where the individual findings fit into a puzzle making more and more sense with respect to biological knowledge.

Studying genetic alterations causing drug resistance combining basic biological knowledge and select genes on the basis of functional hypothesis, together with application of different analytical methods, we believe the causes of chemo resistance may be identified within the next decade.

**references**

33. Knoop AS, Krudsen H, Balslev E et al. Retrospective analysis of topoisomerase IIa amplifications and deletions as predictive markers in


