Immunohistochemical markers as predictive tools for breast cancer

R A Walker

ABSTRACT
Breast cancer is the predominant malignancy where oncologists use predictive markers clinically to select treatment options, with steroid receptors having been used for many years. Immunohistochemistry has taken over as the major assay method used for assessing markers. Despite its extensive use there are still issues around tissue fixation, methodology, interpretation and quantification. Although many markers have been evaluated, the oestrogen receptor remains the most reliable and best example of a predictor of treatment response. It is of major importance clinically that those undertaking interpretation of predictive markers understand the technical pitfalls and are aware of how expression of a particular marker relates to breast cancer pathology. A false negative or a false positive result will impact on patient management.

Steroid receptors have been used for predicting outcome and response to therapy of breast cancer for many years. This has been the predominant cancer where oncologists have used such markers clinically to select treatment options. Assessment of receptors and other markers was by biochemical methods but practice has changed, with immunohistochemistry now being the major assay used. It has also taken over from other techniques such as flow cytometry and immunoassay. Despite its extensive use there are still issues around the methodology, interpretation and quantification that those assessing results and those applying the results must be aware of. These problems have been highlighted in a recent perspective and in recommendations from the Ad-Hoc Committee on Immunohistochemistry Standardization, USA. This review will consider general points that relate to these issues and are applicable to all markers, and then discuss those markers that are used either routinely or in a research setting for prediction. The important issue is that the markers will be used to determine therapy, so a false negative or a false positive result could impact on patient survival.

GENERAL ISSUES
Fixation
The type of fixative, delays and duration of fixation can be particularly important for the detection of certain antigens. Delayed fixation results in increased proteolytic degradation, which can lead to loss of immunoreactivity, particularly for the oestrogen receptor (ER). Formaldehyde fixation results in protein cross linking and hence better secondary structure for histology, but the cross linking is a slow process, needing 24–48 hours to be completed. If formalin fixation is shorter, the fixation process may be completed by coagulation fixation during tissue dehydration by alcohol. This can result in variations in immunostaining within a tissue section. Under-fixation has been found to have more of an effect on ER immunohistochemistry than over-fixation.

The problems relating to fixation have been recognised in various guidelines. The NHSBSP recommendations for the handling of surgically excised breast specimens are that they are received as soon as possible after surgery, and sliced to allow rapid and even penetration of the fixative. The ASCO/CAP HER2 guidelines recommend no less than 6 hours and no more than 48 hours fixation in sufficient buffered formalin, after slicing. Despite this, variation in fixation between laboratories is a major problem when trying to achieve standardisation of immunohistochemical assays.

When assessing predictive markers, if there are any concerns about fixation (as assessed by tissue morphology), then an alternative sample should be sought, or for HER2, an alternative method used. Caution about the significance of the result should be conveyed in the diagnostic report.

Samples
Needle core breast biopsies (NCB) are a standard method for non-operative diagnosis and should benefit from more rapid fixation. They are of particular value for marker assessment for patients receiving neo-adjuvant chemotherapy. Testing of NCB can result in marker data being available at multidisciplinary discussions for therapeutic planning. For ER, results between NCB and excised tumours are good, with results higher in the former. This may reflect a higher chance of sampling the tumour periphery, but could be due to better fixation. For HER2, crushing of tumour cells in core biopsies and edge artefact staining can cause problems in interpretation which has to be recognised.

Adjuvant therapy decisions will be made on the basis of findings in either NCBs or the excised primary tumours. The data available indicate that there is little difference between these tumour samples and local lymph node metastases. The issue arises as to whether there are changes between the primary and subsequent recurrent or metastatic disease, particularly with the increasing use of adjuvant treatment. Changes can occur in ER following the development of tamoxifen resistance, with 15–20% of cancers becoming negative. Reductions in the number of progesterone receptors (PgR) have been found following hormonal therapy. There is debate about the
frequency of differences in HER2 status between primary tumours and distant metastases, since studies are based on small numbers of cases.11-14 A recent report, in which 7.6% of cases were discordant, suggests that discrepancies relate to interpretational difficulties, heterogeneity and borderline amplification in the primary.15 If there are such issues about the results for the primary tumour, it is appropriate to retest distant metastases/recurrent disease if tissue samples are available.

Sections

There is stability of proteins in wax blocks, so cases can be analysed after a long period of storage. However, there is evidence of deterioration of protein reactivity once paraffin sections have been cut,16-17 which is a particular problem for nuclear antigens. For HER2, a time period of no more than 6 weeks between sectioning and staining has been recommended.8

Which antibody?

There is a wealth of antibodies commercially available that are directed against the common tumour markers. It is important that those undertaking marker interpretation understand the differences in specificity and sensitivity. How to select the best antibody for a specific antigen is complex, but is aided by comparisons with a “gold standard” and by the use of external quality assurance data such as that available from the UK National External Quality Scheme (NEQAS).

For ER, two monoclonal antibodies have been used widely, 1D5 and 6F11. Comparison between the two18 has found an overall high concordance rate, but with 6F11 giving stronger, cleaner staining. Results from UK NEQAS confirm that 6F11 has an overall more satisfactory performance.19 A new ER rabbit monoclonal antibody, 5P1, has been introduced and has been found to give favourable results; however, it was compared with 1D5,20 not 6F11, so it is difficult to conclude that it represents a new standard for ER assessment without further evaluation.21 In order to improve standardisation of methodology, PharmDx (Dako) has introduced kits for ER and PgR which have to be used with the Dako autoimmunostainer; that for ER has a cocktail of two monoclonal antibodies including 1D5. Laboratories have to be aware that new antibodies introduced may not be specific; the PgR rabbit monoclonal SP2 was found by UK NEQAS to give false positive results, which may partly have been due to it having been developed for use without antigen retrieval (K Miller, personal communication).

Press et al22 showed wide variations in both sensitivity and specificity between 38 HER2 antibodies, using HER2 gene amplification status as the “gold standard”. Recent UK NEQAS results confirm that when using cell lines of known HER2 gene amplification status, acceptable staining is achieved at a higher level with HercepTest kits rather than clone CB11 monoclonal antibody.23

It is important to understand that the affinities of different antibodies to the same protein can differ and so influence detection. This is evident for p5324-25 and could potentially make its clinical utility difficult.26

Methods

Most proteins in fixed tissue require some form of antigen retrieval. There are some antigens where enzymatic treatment is preferable,27 but most protocols relate to heating in a buffer, by pressure cooking, microwaving or water bath. Inter-laboratory comparisons have shown that insufficient antigen retrieval is a major contributory factor causing variations in extent of staining of test sections.28-30 Microwave antigen retrieval caused greater problems.31 Low level expressing cancers can be assessed as negative, which for ER could have an impact on patient management. Excess antigen retrieval can cause problems in the interpretation of HER2 immunostaining25 and could result in over-calling. The buffer used for antigen retrieval can also influence results. Tris-EDTA (pH 8.9)32 or borate buffer (pH 8.0)33 can give better results than citrate buffer (pH 6.0) for the ER clone 6F11, so if laboratories are experiencing problems with detection of ER in some or all cases when using the UK recommended method,7 it is worthwhile trying these alternative buffers.

Other methodological variables that could affect sensitivity include antibody dilutions and incubations and secondary detection systems. In one analysis of HER2, using automated quantitative analysis and varying dilutions of HER2 antibody, the concentration of the latter affected the apparent relationship between biomarker expression and outcome.34 This did not apply to ER. The use of automated immunohistochemical systems should overcome technical variability but will not compensate for commercial kit variability.

Assessment

This will be discussed in relation to the specific markers, since systems vary. Problems common to all are: there are no external quality assurance schemes for assessing the ability of those undertaking evaluation of immunohistochemistry; the assessment schemes often include intensity as well as extent, but unless there are good controls (e.g. range of ER staining) and each run is compared to the results for controls, evaluation will be variable; defining what is a cut-off, since these vary between different studies, so what is positive and what is negative can vary for the same antigen. These issues and more have been recognised for some time. In 1997 the EORTC-CCCG issued a consensus report on a scoring system for immunohistochemical staining which considered staining patterns, area of assessment, counting methods and defining cut-off values.35 The need for quality assurance schemes is highlighted by a study in Germany of 172 pathologists, where 24% of ER interpretation resulted in a false negative assessment.26

Automated analysis

Interpretation of immunohistochemistry is usually done manually and is, therefore, dependent on the experience and ability of the interpreter. Computerised image analysis systems have been used since the late 1980s and were shown to provide a more accurate means of quantification of ER.36-37 However, cost and technical issues restricted their use. Image analysis systems require a linear relationship between the amount of antigen and the staining intensity detected; if diaminobenzidine is used as the chromagen, this relationship only occurs at low levels of staining intensity.38 Recent approaches have used antibody-conjugated fluorophores and fluorescent microscopy systems, for example AQUA,39-40 or in-house systems,41 but the protocols are complex and not suitable for a diagnostic service. There have been reports that assessment of HER2 immunohistochemistry by image analysis such as ACIS (ChromaVision) improves accuracy and reproducibility, but this still requires an operator to select regions to be quantified on the scanned slide.41-42 Further systems have been developed, e.g. ARIOL and PERIO, but all are expensive and currently are more suited for assessment of immunohistochemistry in a research setting.
OESTROGEN AND PROGESTERONE RECEPTORS

The oestrogen receptor was first identified in the 1960s. The analyses of oestrogen and progesterone receptor in breast cancers\(^5\)\(^6\) quickly provided the evidence that they could aid the identification of those cancers that were more likely to respond to endocrine treatment. The assays were dependent on the homogenisation of frozen tumour tissue with the preparation of a cytosol for subsequent ligand binding. The most widely used method was the dextran-coated charcoal assay (DCC), with results being expressed as fmol/mg cytosol protein, i.e. the receptors could be quantified. Response data showed that not only was the presence of ER important, but also the amount in aiding prediction. The presence of PgR, which is induced by oestrogen, is also a predictor of response. The DCC assay had the advantage of providing a quantifiable level of receptor, but it required fresh tissue and the level of receptor could be influenced by the presence of large amounts of normal breast and/or stroma. Such factors led the drive for histological based methods and the development of monoclonal antibodies to ER and PgR that could be used in fixed tissue and applied routinely. For the methods to be clinically valuable they have to have the same predictive power as the original biochemical assay.

Studies from large centres have shown that immunohistochemistry for ER is more sensitive than DCC for cancers from premenopausal women,\(^49\) and immunohistochemistry for ER\(^50\)\(^51\) and PgR\(^52\) has a high sensitivity and specificity in comparison to biochemical determination. All stress the importance of rigorous quality control of methodology. ER immunohistochemistry gave superior results to the biochemical assay in relation to type and duration of response of metastatic breast cancer to first line tamoxifen treatment.\(^64\) Harvey \textit{et al.}\(^65\) found ER immunohistochemistry to be superior to the ligand binding assay for predicting response to adjuvant endocrine therapy, even though the samples tested had been frozen for the biochemical assay years before and had only been fixed and processed to allow biochemical comparison. A similar approach was used by Cheang \textit{et al.}\(^59\) who found lower positive rates with immunohistochemistry to DCC and poorer prediction of response, but did acknowledge the problems caused by freezing prior to fixation. The International Breast Cancer Study Group has recently re-evaluated cancers entered into two trials of adjuvant endocrine therapy.\(^50\)\(^62\) These were originally tested for ER by ligand binding assay; standard fixed tissue was assessed for ER and PgR by immunohistochemistry in a central laboratory. There was good concordance between assays with similar outcomes, but for premenopausal patients immunohistochemical PgR could predict response, unlike the biochemical assay.

Overall, there is good data to show that immunohistochemical determination of ER and PgR can be of similar predictive value for response to endocrine therapy as the original biochemical assays, but optimal fixation and a high standard of quality assurance is needed.

Assessment

As already discussed there are several issues around assessment of immunohistochemistry. The biochemical assays for ER and PgR are quantifiable, being expressed as fmol/mg cytosol protein. Various methods have been and are used for scoring ER and PgR. The H score\(^51\)\(^52\) involves assessment of the percentage of cells stained as weak, moderate or strong, which are summed to give an overall maximum score of 300. The original authors suggested a cut-off point of 100 to distinguish positive and negative.\(^55\) The Nottingham group chose an arbitrary score of <50 as negative in early studies,\(^55\) but subsequently have used a cut-off of 20, based on the point of the trough in staining.\(^54\) The quick score\(^55\) is based on the percentage range of cells staining from 1 to 4 and overall intensity as 1 to 3, which are then added to give a maximum of 7. This has generally now been replaced by the Allred score,\(^56\) which has expanded the lower end of the percentage of cells staining, giving a range of 1 to 5 and a maximum of 8. ER positive is defined as score >2. Other systems include assessment of percentage of positive cells, irrespective of intensity, with any staining considered positive.\(^50\)

There are other critical factors in assessment which relate to interpretation. Each assay should include a control comprising a high staining tumour, a low–moderate staining tumour and a negative case, and assessment of test cases, particularly for intensity, should be in relation to this. ER and PgR are present in normal breast and form a useful internal positive control. If there is no staining, problems with the assay are indicated, but the age of the patient should be taken into consideration since levels in young premenopausal women can be very low.\(^58\) The frequency of positivity is higher in grade I invasive cancers and screen detected cancers,\(^57\) so if there is no or low level staining, repeat assay should be undertaken. In assessing ER and PgR staining, only invasive carcinoma and nuclear reactivity are considered. Cyttoplasmic staining can be due to excess antigen retrieval, and can also be seen in apocrine differentiation, although such cells express androgen receptors rather than ER.

Although there have been many publications about ER immunohistochemistry, there is still debate about quantification and what is required clinically. Fisher \textit{et al.}\(^60\) compared various methods of scoring ER and PgR, involving percentage ranges and intensity, both summated and as a product, and concluded that the “any-or-none” method was just as good at prediction, and simpler. However, Barnes \textit{et al.}\(^47\) in a very thorough comparison of scoring methods, showed that there was a correlation between greater extent of staining and likelihood of favourable response. In neoadjuvant endocrine treatment the Allred score has been of value in identifying those cases more likely to respond.\(^39\) The BIG 1–98 trial of adjuvant endocrine therapy in postmenopausal women identified differences in outcomes between those cases that were ER negative and had 1–9% of positive cells and those with >10% positive cells, indicating the importance of detecting low levels of receptor.\(^56\) Schnitt, in a “Comments and Controversies”,\(^61\) has highlighted that variation in pre-analytical factors and assays will affect attempts to standardise quantification of ER by immunohistochemistry, but has also suggested that the highly sensitive antibodies and detection systems cannot identify differences in amounts in the higher staining tumours. Rimm \textit{et al.}\(^40\) have questioned whether this is due to the technique or interpretation, and consider that image analysis systems can give better quantitative discrimination. From a clinical perspective, having sensitive techniques that can detect low levels of ER is important. The Allred score concentrates on this low end, is easy to use and is recommended by the author.

Those undertaking interpretation of ER and PgR should be aware of all of these many factors. Table 1 gives recommendations for staining and assessment.

HER2

The clinical importance of amplification of human epidermal growth factor receptor 2 (HER2) (also knows as HER-2/neu/ c-erb B2) in breast cancer was recognised in 1987.\(^63\) Numerous subsequent studies found that either HER2 gene amplification
or protein expression predicted for poor prognosis. Following the development of a humanised monoclonal antibody against HER2 (trastuzumab), the reasons for establishing the HER2 status of breast cancers changed, since it is a prerequisite for trastuzumab’s clinical use. Trastuzumab was originally licensed for the treatment of patients with metastatic disease who had HER2 positive cancers. More recently several prospective randomised trials have shown that adjuvant trastuzumab reduces the risk of recurrence and mortality in patients with HER2 positive early stage breast cancer. This resulted in it being licensed for adjuvant use and being endorsed by the UK National Institute for Clinical Excellence (NICE).

The principal testing methods used are immunohistochemistry and/or in situ hybridisation using either fluorescence (FISH) or a chromogen. In comparison to ER data on response, information is limited as to whether HER2 overexpression as detected by immunohistochemistry or HER2 gene amplification as detected by FISH is a better predictor. Data from the metastatic setting suggests that there is a higher overall response in patients with HER2 FISH positive than FISH negative cancers, but the overall response rate of the patients (all with HER2 positive breast cancers) to single agent trastuzumab was around 55%.

There are a variety of factors that can modify immunoreactivity for HER2 and therefore affect interpretation, which have been referred to above. These include: poor fixation, which can be a particular problem in excision specimens; crushing of tumour cells and edge artefact staining in NCB; batch variation of assay kits; excess antigen retrieval; and excess nuclear counterstain.

Table 1 Recommendations for staining and assessment of ER and PgR

<table>
<thead>
<tr>
<th>Critical factors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Staining</strong></td>
</tr>
<tr>
<td>Optimal fixation</td>
</tr>
<tr>
<td>Antigen retrieval—citrate pH 6.0, but test higher pH buffers if suboptimal staining</td>
</tr>
<tr>
<td>Antibody validated against biochemical assay</td>
</tr>
<tr>
<td>Positive control with range of staining; choose test tissue with normal breast included if possible</td>
</tr>
<tr>
<td>Quality assurance, internal and external</td>
</tr>
<tr>
<td><strong>Assessment</strong></td>
</tr>
<tr>
<td>Nuclear staining only; cytoplasmic staining may be due to excess antigen retrieval</td>
</tr>
<tr>
<td>Only invasive cancer assessed</td>
</tr>
<tr>
<td>Strong relationship with grade, so if grade I low/negative repeat</td>
</tr>
<tr>
<td>Use a recognised scoring system. Allred score easy to use and identifies low positive cases</td>
</tr>
</tbody>
</table>

Assessment

It is important that only invasive carcinoma is assessed. The scoring system used is the same whichever assay is employed and is shown in table 2.

Only membrane staining of invasive cells in considered. Cancers are categorised as negative if no staining is seen or membrane staining is <10% invasive cells; 1+ (and therefore negative) if there is faint membrane staining in >10% of cells). Equivocal or 2+ staining is weak to moderate complete membrane staining in >10% of cells or <30% with strong complete membrane staining. This requires further analysis by another system to check amplification status. A positive case is 3+ which is strong membrane staining in >30% of cells. There has been a change from 10% to 30% in the recent ASCO/CAP guidelines, which is being endorsed in the updated UK HER2 testing recommendations. HER2 overexpression is more likely to be present in a grade 2 or grade 3 invasive breast cancer. Unlike ER, staining should not be present in normal breast.

The main problems in interpretation and in intra- and interobserver variation arise with cases that are at the 1+/2+ borderline and the 2+/3+ borderline. It is these categories that can be affected by the technical issues outlined, so it is particularly important that those undertaking interpretation are aware of the impact of these issues and that regular audits are undertaken.

OTHER MARKERS

Epidermal growth factor receptor

Epidermal growth factor receptor (EGFR, also HER1) is a type 1 tyrosine kinase receptor that is expressed in normal breast. The frequency of detection by immunohistochemistry in breast cancers varies between different studies and can range from 15% to >60%. The reasons for this variation relate to differences in methodology, antibodies used, interpretation and the cancers studied. Unlike ER and HER2 there is no other recognised assay that antibodies and the immunohistochemical technique can be compared to. The EGFR PharmDx assay (Dako, UK) is licensed in the USA for testing colon cancer and is a similar assay to the HercepTest (Dako) with a scoring system of 0 to 3+. Reported studies vary from using an H score system to positive if any staining is present.

The presence of EGFR in breast cancers is associated with a lack of ER and poor prognostic features. There are other reasons why assessment of EGFR could be of value. Tyrosine kinase inhibitors of EGFR, such as gefitinib, are now available and are being tested in trials of advanced and early breast cancer. Although the results so far are not promising, if EGFR testing is to be used as a method of patient selection, there needs to be better standardisation of the assay.
Table 2  Immunohistochemical assessment of HER2

<table>
<thead>
<tr>
<th>Score to report</th>
<th>HER2 protein assessment</th>
<th>Staining pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Negative</td>
<td>No staining is seen or membrane staining is &lt;10% of invasive tumour cells</td>
</tr>
<tr>
<td>1+</td>
<td>Negative</td>
<td>Faint/barely perceptible membrane staining detected in &gt;10% of invasive tumour cells</td>
</tr>
<tr>
<td>2+</td>
<td>Equivocal</td>
<td>Weak to moderate complete membrane staining in &gt;10% of invasive tumour cells or &lt;30% with strong complete membrane staining</td>
</tr>
<tr>
<td>3+</td>
<td>Positive</td>
<td>Strong complete membrane staining in &gt;30% of invasive tumour cells</td>
</tr>
</tbody>
</table>

2+ cases should be assessed by FISH, as should other cases where there is heterogeneity, problems with immunohistochemistry interpretation and problems relating to fixation.

Basal markers

Gene expression profiling has identified different subgroups of breast cancers, that link to patient outcome. One subgroup that was associated with poor outcome expressed genes characteristic of basal or myoepithelial cells of normal breast. Most of these are high grade and lack ER, PgR and HER2 and have a higher risk of brain and lung metastases. However, there is no accepted consensus on the immunohistochemical profile that defines these basal like cancers. Most studies include cytokeratins 5/6 and/or 14, but Nielsen et al define them as lacking ER, PgR and HER2, expressing basal cytokeratins and EGFR and c-KIT. Matos et al consider them to express P-cadherin and p65 more frequently and recommend that cytokeratin 5, p65 and P-cadherin can be used to distinguish a basal like carcinoma. Rakha et al have proposed that basal cytokeratins (5/6 and 14) can be used to define basal like carcinomas irrespective of the expression of other markers.

The response of patients with basal like breast cancer to chemotherapy has been reported as both poor and good. There are similarities between basal like breast cancers and those cancers arising in women with BRCA1 mutations. Therapeutic approaches that have potential in BRCA1 deficiency, for example carboplatin and PARP inhibitors, could be of value in the management of basal like cancers, and clinical trials of the management of ER, PgR, HER2 (triple) negative cancers are being undertaken. EGFR and c-KIT are expressed at a high level in many basal like carcinomas and clinical response to neoadjuvant chemotherapy has been reported as both poor and good. A decrease in Ki-67 was found to predict good clinical response to neoadjuvant chemotherapy in a study using cytology, but a subsequent study using NCB was less conclusive, particularly for pathological response. A problem with these studies is that the pretreatment assessment has to be done on NCBs, i.e. relatively small samples. If there is intratumoural heterogeneity of Ki-67 expression, this will affect the counts obtained from small samples. One issue with the use of Ki-67/MIB1 is the lack of an agreed scoring method and the definition of high/low, positive/negative. For assessing changes, percentage of positive (any staining) cells (counting 1000–3000 cells) has been used, whereas others estimated the percentage of positive nuclei within the area of highest positivity. Assessment of whole tumour sections has been of 10–20 random fields at ×400 to give a percentage, but cut-off levels have varied from positive if >5% of cells staining with 20% as high, to <9.5% low, >9.5%–<15.5% intermediate and >15.5% high, when compared to histological grade.

Other proliferation markers that have been evaluated immunohistochemically in breast cancer include cyclin E, cyclin D1, p21 and p27, but there is no strong evidence for their use as predictive markers outside of clinical research.

Table 3  Markers and their value in prediction in breast cancer

<table>
<thead>
<tr>
<th>Established and in routine clinical use</th>
<th>Potential for clinical use; need refinement of scoring systems or antibodies</th>
<th>Research interest, less likely to be used clinically</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oestrogen receptor</td>
<td>Epidermal growth factor receptor</td>
<td>P53</td>
</tr>
<tr>
<td>Progesterone receptor</td>
<td>Ki-67 (MIB-1)</td>
<td>Cyclin E, cyclin D1, p21, p27</td>
</tr>
<tr>
<td>HER2</td>
<td>Topoisomerase II alpha</td>
<td>Bcl2, bax, bcl-x, survivin</td>
</tr>
</tbody>
</table>

Proliferation markers

The Ki-67 antigen is expressed in the nucleus of cells in all phases of the cell cycle and is a useful marker of cell proliferation. The MIB1 antibody is reactive against the antigen in fixed, embedded tissue and gives comparable results to the original Ki-67 antibody, which was only reactive with antigen in fixed tissue. Several studies have shown that both Ki-67 and MIB1 staining are of prognostic value.

Changes in Ki-67 expression following preoperative endocrine treatment can predict long term outcome. The rationale is that endocrine treatments act by inhibiting tumour cell proliferation, so decreases in Ki-67 after short-term treatment indicate effective responses. Pretreatment assessment of Ki-67 has also been shown to predict response to preoperative chemotherapy. A decrease in Ki-67 was found to predict good clinical response to neoadjuvant chemotherapy in a study using cytology, but a subsequent study using NCB was less conclusive, particularly for pathological response. A problem with these studies is that the pretreatment assessment has to be done on NCBs, i.e. relatively small samples. If there is intratumoural heterogeneity of Ki-67 expression, this will affect the counts obtained from small samples. One issue with the use of Ki-67/MIB1 is the lack of an agreed scoring method and the definition of high/low, positive/negative. For assessing changes, percentage of positive (any staining) cells (counting 1000–3000 cells) has been used, whereas others estimated the percentage of positive nuclei within the area of highest positivity. Assessment of whole tumour sections has been of 10–20 random fields at ×400 to give a percentage, but cut-off levels have varied from positive if >5% of cells staining with 20% as high, to <9.5% low, >9.5%–<15.5% intermediate and >15.5% high, when compared to histological grade.

Other proliferation markers that have been evaluated immunohistochemically in breast cancer include cyclin E, cyclin D1, p21 and p27, but there is no strong evidence for their use as predictive markers outside of clinical research.

Apoptotic proteins and p53

As with proliferation markers, there have been many immunohistochemical studies evaluating expression of apoptotic proteins including bcl-2, bax, bcl-x and survivin, but for various reasons including availability of suitable antibodies, methods of evaluation and lack of strong evidence, these are not suitable as routine predictive markers.

p53 has been considered as a potential predictor of response of breast cancers to chemotherapy, but much of the data comes from mutation analysis. Immunohistochemistry detects stabilised p53 protein, which may reflect a mutation but will not detect protein truncation mutations; there are also problems in evaluation and defining what is positive.

Topoisomerase II alpha

Topoisomerase II alpha is a target of anthracycline action, a chemotherapeutic drug that is frequently used in the management of breast cancer. The gene encoding this is TOP2A which maps to 17q21 and can be co-amplified with HER2. There are conflicting reports as to whether TOP2A amplification can be used as a predictor of response to anthracycline based chemotherapy, although recent reports suggest that it could...
be a useful marker.\textsuperscript{13} This study found that there was a good correlation between amplification and immunohistochemical detection of the protein using the antibody Ki-S1 when >25% cells staining was used as the cut-off for defining overexpression.

THE FUTURE
Will immunohistochemistry remain the main method for assessing predictive markers? There has been debate for some time about its role in HER2 testing, with some centres preferring frontline FISH, to which TOP2A could be added. Real time (quantitative) PCR is being used to assess gene expression levels for ER\textsuperscript{61} and HER2.\textsuperscript{119} Commercial assays that cover expression of a range of genes, e.g. Oncotype DX are available. If molecular assays become automated on the scale of biochemical assays and cost per test becomes competitive, the use of immunohistochemistry, with its problems around quantification, could change.

CONCLUSIONS
Table 3 presents a summary of the various markers discussed and their potential roles in prediction. Despite evaluation of many markers, ER remains the most reliable and best example of a predictor of treatment response for breast cancer. HER2 is used as a marker to select patients for a specific form of treatment, trastuzumab, but there is insufficient data about response. Immunohistochemical determination of expression in breast cancer pathology.

Acknowledgements: I am grateful to Beverley Richardson for secretarial assistance.

Competing interests: None declared.

REFERENCES


