Reading the signs

The role of genomic signatures in guiding treatment decisions

Anna Wagstaff

The concept of a genomic signature that can define any tumour according to its unique biology has been central to the emerging paradigm of personalised therapy. But how helpful are multi-gene assays in guiding treatment decisions in clinical practice, and what role can we expect them to play in the future?

Read the signature, choose the treatment. This was the enticing prospect that opened up with the advent of technologies capable of reading the gene expression profile of tens of thousands of genes in tumour tissue.

But more than a decade after these techniques became available, even the two best-known and tested multi-gene assays are not yet part of mainstream clinical practice in Europe. Although Genomic Health's Oncotype DX and Agendia's Mammaprint (both designed to sort high-risk from low-risk breast cancers), show that the technology has got better, quicker and considerably cheaper, the whole field is looking more complex than many had hoped.

It was only two years ago that the St Gallen conference, which has been setting out consensus guidelines for adjuvant treatment of early breast cancer since 1978, first mentioned a role for multi-gene assays, and then only for when traditional clinical and immunohistochemistry tests are inconclusive. As Europe's breast cancer community gathers for the 11th St Gallen conference in March 2011, the role of multi-gene assays in guiding adjuvant treatment will again be up for discussion. Their deliberations will be followed by specialists in other fields – colorectal, lung and prostate cancer – who are also looking for guidance on adjuvant therapy, and wonder how helpful the various multi-gene assays currently being proposed and trialled in their areas could be.

Not so simple

The idea that every tumour could be classified into clinically relevant categories based on the expression of a signature set of genes led to bonanza time for specialist biostatisticians in the 1990s. They found their services in huge demand as competing research groups scrambled to be the first to identify and validate signatory gene sets. But a head-to-head race between teams in Amsterdam and Rotterdam seeking a signature that would differentiate between breast cancers at high risk of recurrence and more indolent tumours gave the first indication things might not be quite so simple. Both signatures proved to be quite effective at predicting prognosis, but only three of the 76 genes used in the Rotterdam signature were also found among the 70 genes making up the Amsterdam signature.

Meanwhile, the general concept of the tumour signature was brought into increasing disrepute as the technology became more widely available and hundreds of disparate studies were conducted in populations where the recurrence rate within the studied population was too low and the numbers of genes being studied – and consequently the number of multiple comparisons – was too high. Commentators

like John Ioannidis from the University of Ioannina in Greece led calls for fewer, larger, better-designed studies, showing that the huge data sets that were being gathered from relatively small numbers of patients were open to almost any interpretation. As he commented at a media 'reality check' in 2007: "with 30,000 genes to choose from, anyone looking for a significant pattern is quite likely to find one," (*Cancer World* Jan/Feb 2008).

His remarks were echoed last year by Rachel Midgley who, as lead clinician in the QUASAR trials, has been examining the evidence for a multi-gene assay to guide treatment in colorectal cancer. She talked about the literature being "replete with rather unconvincing, small studies, which are underpowered, variable in their methodology and quality assurance and which are often linked to incomplete clinical data sets. Even meta-analysis of these sorts of studies is rather suspect … so perhaps it is not surprising that little progress has been made in defining a useful marker, which would help mainstream clinical decision making," (*Cancer Journal* 16:210–213).

Also undermining the concept of a single tumour signature was growing evidence that the biological make up can vary considerably from one region of a tumour to another. One area might test highly positive for amplified HER2 expression, for instance, while another area tests normal. Taken together with the evidence that biology changes over time and in response to treatment, and that the genetic signature of the primary tumour and metastasis can differ greatly, the question arises whether a tumour signature may only apply for a certain place and time. There is also a growing body of evidence that the biology of the non-cancerous host tissue surrounding the tumour plays a major role in determining prognosis and possibly also points to the need for different treatments. Perhaps we need host signatures as well as tumour signatures.

Some specialists also question the value of the information given by the gene signature. Ideally, they argue, targeted therapies should be directed at a functional pathway to which tumour survival is addicted; however, gene status does not reliably correlate with downstream protein status or with pathway function. Thus, the fact that a tumour shows amplified HER2 expression, for instance, does not mean that HER2 is necessarily the driving force, which is why not all patients with HER2-positive tumours benefit from HER2 inhibitors such as transtuzumab (Herceptin). The point is well illustrated by a comparison of the pathways involved in the Amsterdam and Rotterdam gene signatures, which demonstrate that, despite the minimal overlap between the genes, they had 21 pathways in common.

"But it works"

Steve Shak, chief medical officer at Genomic Health, is well-equipped to argue the finer points of targets and functional mechanisms, having come from Genentech, where he led Herceptin through the approval process. But his answer to those who question the usefulness of Oncotype DX is a simple one: it does what it says on the tin – it assists

ILLUSTRATION: FRED VAN DEELEN, WWW.ORGANISART.CO.UK

patients and doctors to decide whether to opt for chemotherapy in cases of early ER-positive breast cancers, by indicating the likelihood of recurrence and of response to treatment.

Oncotype DX uses a technique called real-time polymerase chain reaction (RT-PCR) to measure levels of mRNA expression from 21 genes in samples of tumour tissue taken from a standard paraffin bloc. Using these results, it then calculates a 'Recurrence Score' of between 0 and 100, which corresponds to the likelihood of the cancer returning within 10 years. Scores between 0 and 17 are defined as 'low risk'; 18 to 30 as 'intermediate risk' and 31 and above as 'high risk'.

Shak says that the Oncotype DX breast cancer assay has successfully come through the kind of close scrutiny that he and his colleagues welcome. "As a physician or patient I want a test that has demonstrated it really works; that there are multiple studies to show it is fit for the specific purpose for which I am going to be using it," he says. "We worked with leaders in oncology throughout the world to do multiple studies that not only identified the best genes but also validated their use in multiple well-defined rigorous clinical studies. We've now done 14 studies in more than 4000 patients." The use of Oncotype DX to guide adjuvant treatment decisions in early ER-positive breast cancers is now included in the published ASCO clinical guidelines.

Shak readily concedes that the track-record of research into molecular diagnostics in general has not been a glorious one. "When we started Genomic Health in 2000, we looked across a landscape of biomarker research and development, and it was easy at that time to get very depressed. There were tens of thousands of articles on this marker and that marker, and very few if any of those ever made it into clinical practice."

But every new technology comes with its own learning curve, and there are signs now that the field is maturing. In the US, the FDA-led MACQ project is trying to bring some level of quality control to the whole field of microarray, RT-PCR and other 'next-generation sequencing technologies' and promote their proper application in discovery and development. Crucially this includes providing guidelines on the capabilities and limitations of various data analysis methods in developing and validating predictive models, and reaching a consensus on 'best practice'. Mammaprint (which came out of the Amsterdam research) has become the first multi-gene assay to have its prognostic powers recognised by the FDA, although approval by the FDA (or EMA, its European counterpart), is not a requirement for these tests.

Lessons have been learned too about

Multi-gene assays – the story so far

- Gene signatures, or multi-gene assays, can be captured in a variety of ways. Agendia uses the Agilent microarray platform to generate the Mammaprint, which classifies early ER+ breast cancers into high risk and low risk, based on the 70-gene signature originally identified by the Amsterdam team. Alternative microarray platforms include Affymetrix, which was used for the 76-gene Rotterdam assay and has also been used to generate a putative 23-gene signature to predict for recurrence in colon cancer (*JCO* 27:1564 –1571). This technology requires frozen tumour specimens, although new techniques are being developed to enable it to be used with paraffin-embedded specimens as well.
- Genomic Health uses an alternative technology, real-time polymerase chain reaction (RT-PCR), to read the 21-gene signature that forms the basis of the Oncotype DX multi-gene assay for breast cancer. This assay generates a Recurrence Score for each tumour of between 1 and 100. This can be done using paraffin-embedded tissue.
- Agendia argues that Mammaprint is more helpful, in that it rates tumours as either 'high risk' or 'low risk', thus avoiding the chance that readings may come back as inconclusive. Genomic Health counters that all the evidence shows the biology of ER+ breast cancers to be a continuous variable, and that giving risk according to a continuous recurrence score is therefore more accurate.
- There is disagreement also over which of the two assays can claim greater validity. Mammaprint is the only assay whose powers to predict outcome have been approved by the FDA. On the strength of this and other evidence, it is now reimbursed by Medicare in the US. However, Oncotype DX is the only assay to be included in the ASCO clinical guidelines as a tool for deciding who will benefit from chemotherapy (*JCO* 25:5287–5312), and it too is reimbursed by Medicare as well as Medicaid and the major US health insurers. The evidence levels behind the decision to include only Oncotype DX in the guidelines has been a point of debate (*JCO* 25:2057–2058; *JCO* 2058–2059; *JNCI* 101:1456–1452). More information about their relative merits will be available with the results of the large Tailor X and MINDACT trials, but these are not due to report final results for many years yet.

CORBIS/ED KASHI

Do I have to do this? Finding more accurate ways to predict who will benefit from chemotherapy and who will not would save countless cancer patients unnecessary suffering and cut healthcare costs

the need for a more collaborative approach to secure the patient numbers necessary to provide reliable results. Trials the size of MINDACT and TailorX, which seek to establish how effectively Mammaprint and Onctopye DX predict who will benefit from adjuvant chemotherapy, have enrolled many thousands of patients who will be followed for many years. When they were set up, it was hard to envisage studies of a similar scale outside of the highly organised and well-funded field of breast cancer. Yet we are now beginning to see sizeable trials for molecular markers and multi-gene assays being carried out in colorectal cancer by some of the big European and US collaborative groups, including a variant of Oncotype DX for predicting recurrence risk in colon cancer, being tested by the US NSABP (National Surgical Breast and Bowel Project) using tumour tissue from the QUASAR trials (see p 27, Beyond breast cancer).

The scientific rationale

The validity of these multi-gene assays rests on more than just the statistical evidence that they work, says Shak. The set of 21 genes used for the Oncotype DX assay – 16 'cancer genes' and five reference genes – make sense in terms of what is already known about the biology of breast cancer. "We picked the genes because when we looked across the studies they provided evidence that they really work. But the [cancer] genes actually turned out to be in four groups and then three independent genes." Those four groups, he explains, relate to proliferation, oestrogen receptors, HER2 and invasion. "It was very reassuring to us that, when the genes revealed themselves in terms of which ones worked, they actually did group themselves in a way that was very relevant to our understanding of breast cancer biology."

This remains true, argues Shak, even if you look at some of the most recent evidence about the biology of the host tissue. CD 68, a monocyte macrophage-related gene involved in the body's immune system is one of the 21 signature genes. "That gene relates to the host response to the tumour. So our test captures information about the tumour and host response and integrates it to give single score that is relevant for selection of the right treatment."

As for the argument that mRNA measures do not provide the most accurate reflection of the signalling pathways that are driving the tumour, Shak is pragmatic, describing himself as 'agnostic' on whether DNA, RNA or protein offers the most useful information. Genomic Health opted for measuring mRNA using the RT-PCR technique, he says, because with the technology currently available, that is what works best. "Proteins are complicated, they undergo clipping and post-translational modifications, the assays to look at them are complicated and challenging and I think the issue is not one of which is more important, but which one can be practically harnessed to serve patients."

Indeed, the practicality of the Oncotype DX test is widely regarded as one if its big advantages, because it can be done using paraffin-embedded tumour blocs that are routinely collected for pathology reports, it can be delivered by any overnight courier service and Genomic Health prides itself on a high level of quality control and can point to an excellent track record of reproducibility and accuracy.

We are beginning to see sizeable trials for multi-gene assays carried out in colorectal cancer

The issue becomes one of how much additional information the multi-gene assay can supply

View from the clinic

Catherine Oakman, a medical oncologist with a special interest in breast cancer based at the Sandro Pitigliani Medical Oncology Unit, in Prato Hospital, Italy, will be among the many making the journey to St Gallen this March to learn more about how evidence that has emerged since March 2009 might impact on guidelines for clinical practice.

Lead author of an overview piece on breast cancer assessment tools and optimising adjuvant therapy, published at the end of last year in *Nature Reviews Clinical Oncology* (vol 7, pp 725–732), Oakman follows the recommendations of the St Gallen 2009 consensus, which pose the role of multi-gene assays as a last rather than a first step in consideration of adjuvant chemotherapy for a woman with ER-positive disease.

Mammaprint and Oncotype DX can both provide a molecular picture of biological features, says Oakman, but tumours are already assessed by clinical features – tumour size and lymph node involvement – and by histopathological features – morphology, and ER, PgR, HER2 and Ki-67 (proliferation) status. These are all captured by the traditional St Gallen criteria, she says, so the issue becomes one of how much additional information the multi-gene assay can supply.

"St Gallen criteria are very practical because they incorporate tumour parameters which are readily assessed in daily clinical practice. With that information you can make a therapy decision for most patients. For individuals with ER-positive disease in whom you are still unclear about additional benefit of chemotherapy over endocrine therapy alone, the St Gallen consensus advises consideration of a genomic test. The genomic test comes at the end and only if you are unsure."

She does concede that the reliability of immunohistochemical markers has been a problem. In trials where local pathology reports are quality controlled by a central laboratory, discrepancies of up to 20% have been found for ER, PgR and HER2. All the tests involve some level of reader evaluation, and until recently there was no standard agreement about where to set the diagnostic thresholds.

Improvements may come, says Oakman, from close working relationships between oncologists and pathologists, and standardisation of pathology testing – something she believes is already happening in a number of ways.

Greater involvement in multidisciplinary teams means pathologists are increasingly aware of how the results they report are used to guide treatment decisions, and this gives them an enormous incentive to get the most accurate results they can. There are also new step-by-step guidelines, published by ASCO in conjunction with the US College of American Pathologists (CAP), for testing ER, PgR and HER2. They can be implemented in any hospital laboratory, and though they are not binding, they set a quality standard by which any hospital can be evaluated.

ASCO-CAP guidelines have now set the threshold for ER and PgR positivity as at least 1% of tumour nuclei staining positive. The Ki-67 proliferation marker, however, remains problematic both in terms of an agreed threshold and standards for testing. Oakman accepts that genomic markers are currently superior in robustly and reproducibly measuring proliferation, but argues that given time, standards and guidelines may bring more reliability to this marker as well.

For some patients, treatment decisions are fairly clear. Evidence shows that, in general, tumours with a very high ER and PgR status are sensitive to endocrine therapy but derive little additional benefit from chemotherapy. Individuals with HER2-positive tumours are likely to derive benefit from trastuzumab, and this is always given alongside chemotherapy. For triple negative disease, chemotherapy is the only currently available systemic option.

The problem arises where tumours stain positive for ER, but show some signs of a more aggressive cancer, for instance low PgR, grade 2, intermediate Ki-67 and/or some nodal involvement. This represents a sizable group of women, as about one third of the 75% of patients with an ER+ tumour fall into this uncertain category, and the question of whether these patients might derive additional benefit from chemotherapy over endocrine therapy alone is unclear. "This is the group of patients where genomic tests might help," says Oakman, "That is of course if, firstly, the patient can afford or has insurance to cover the substantial cost, and secondly, if she would be agreeable to chemotherapy if the results showed her genomic risk to be high."

Even then, there is no guarantee that the genomic test will provide the guidance doctor and patient are looking for. "If such a patient is assessed as high or low risk by Oncotype DX, this helps treatment decisions. However if such a patient is assessed as intermediate risk, I am still

unclear about the estimated additional chemotherapy benefit," says Oakman. Shak, from Genomic Health, cites a survey indicating that in 30% of cases, testing with Oncotype DX does in fact lead to a change in the initial decision on treatment. This is not something Oakman is able to confirm from her own experience.

Currently both Oncotype DX and Mammaprint are reimbursed in some European states, but so far only by a small minority of social insurance providers. Oakman suggests it would be best to wait for the results of the TailorX and MINDACT trials to get a clear picture of their powers to predict benefit from chemotherapy in this patient subgroup, before deciding whether public health insurances should cover the cost of genomic tests, at least when traditional markers are inconclusive.

Are genomic tests the future?

Tailoring treatments to patients is a strategic goal for oncology. For many people, this paradigm is all about tracking down targets and designing biological therapies that can block them, and this remains the great hope for the future.

For the moment, however, chemotherapy remains a mainstay of treatment for many cancers, and as early detection strategies ensure more and more cancers are caught at an early stage, the ability to sort the patients who will derive benefit from these toxic treatments from those who will not is by far the greatest 'personalisation' issue in terms of the numbers of patients affected.

Using current regimens in stage II colorectal cancer, for instance, curing an additional three to four patients requires putting one hundred patients through chemotherapy, 40% of whom will suffer significant toxicity. The future for genomic tests like Oncotype DX and Mammaprint will lie in how far they outperform traditional clinical and pathological criteria and how the costs stack up against the benefits for patients and the savings for health services.

However, even if the trials show beyond doubt that these genomic tests have clinical value, they still predict only 'general chemosensitivity' rather than specifying a treatment. The quest for predictive biomarkers for specific chemotherapies is another level of sophistication for the future (see also Treatment of Triple Negative Breast Cancer, p 13).

Beyond breast cancer

In January last year Genomic Health launched an Oncotype DX (RT-PCR) assay for predicting the risk of recurrence in patients with stage II colon cancer (defined as involvement of the bowel wall without lymph-node involvement). The 12-gene assay was shown to be significantly predictive in sorting those with a 'low' risk of recurrence (8%–10% within three years) from those with a 'high' risk (20%–25% risk (*JCO* 27 (15s): abstract 4000).

This January, results from a trial of Agendia's experimental 18-gene Coloprint microarray assay were published (*JCO* 29:17–24), showing it too was significantly predictive at sorting stage II and III patients at 'low' risk of recurrence (around 12.5% recurrence at five years) from those at 'high' risk (around 33% at five years).

Neither study claimed to provide statistical proof that the assays could predict who would benefit from chemotherapy. However, David Kerr, one of the leaders of the QUASAR trial that collaborated in the study of the Oncotype DX assay, said, "When you look at the totality of the data, we think that the new assay provides a clinically useful tool for [identifying] which stage II colon cancer patients should be selected for chemotherapy," (*Medscape* 19 May 2010).

Roberto Labianca, lead author of ESMO's clinical guidelines for diagnosis, adjuvant treatment and follow-up of primary colon cancer (*Ann Oncol* 21 Supp 5: v70–v77), insists, however, that while these data are interesting in the context of the wider search for biological markers that can predict response to treatment in this patient group, they are not sufficiently robust to merit being included in guidelines for adjuvant treatment. For the moment, he says, decisions on adjuvant treatment should continue to be guided by clinical and pathological markers which define stage II disease as high risk according to involvement of the bowel wall, invasion of vascular, neural or lymphatic vessels inside the tumour tissue, obstruction or perforation of the tumour during surgery, or where fewer than 12 nodes have been examined.

The search is also on for multi-gene assays that might predict recurrence and benefit from treatment in many other cancers. Tests for prostate cancer and non-small-cell lung cancer both feature in the Genomic Health pipeline.

Genomic tests are reimbursed in some European states, but only by a small minority of social health insurers