Medical Research Archives

European Society of Medicine

OPEN ACCESS

Published: December 31, 2023

Citation: SharmaV 2023. The Current and Evolving Landscape of Breast Cancer Prediction and Prognosis, Medical Research Archives, [online] 11(12). https://doi.org/10.18103/mra.v11i12.4912

DOI
https://doi.org/10.18103/mra.v11i12.4912

ISSN: 2375-1924

RESEARCH ARTICLE

The Current and Evolving Landscape of Breast Cancer Prediction and Prognosis

Vijay Sharma[1]

[1] Cellular Pathology, Liverpool Clinical Laboratories, Royal Liverpool Hospital, School of Medicine and Department of Molecular and Clinical Cancer Medicine, Department of Molecular and Cancer Medicine, Institute of Systems, Molecular and Integrative Biology, University of Liverpool.

Email: vijay.sharma@liverpool.ac.uk

ABSTRACT

The biological behaviour of breast cancer is remarkably heterogeneous and it is essential to have tools which can provide the necessary risk stratification to plan clinical management. Breast cancer prediction and prognosis needs to be holistic, and account for multiple levels of organisation. The histological classification and grading of the tumour itself presents valuable predictive and prognostic information. Hormone receptor status remains a mainstay, but roles may emerge for assessment of the intrinsic molecular subtype, for a molecular subclassification of triple negative carcinomas, and for whole genome sequencing. The recent discovery that antibody drug conjugates are effective in patients with weak HER-2 protein expression has led to the definition of the HER-2 low group.

There has been a proliferation in predictive and prognostic models, numbering over 900, but the majority are at high risk of bias and tend to perform less well when applied to populations beyond the development cohort. The Nottingham Prognostic Index is a notable exception. Of the molecular risk stratification tools currently available, Oncotype Dx is the most widely recommended and used, but the question as to which test is superior remains unanswerable with current data. There is growing interest in omics-based approaches from which a number of biomarkers are being developed.

It is well established that the microenvironment of the tumour is key to the tumour's behaviour. Some components contain and destroy the cancer, whereas others are co-opted by the tumour and aid in its progression; the current evidence is reviewed, including the current status of tumour infiltrating lymphocyte assessment and immune checkpoint inhibition in breast cancer. The use of the liquid biopsy to achieve early detection of tumours and to manage tumour evolution is receiving intense attention; approaches include circulating tumour cells and circulating tumour DNA. Specific assessment of tumour giant cells may also provide the ability to anticipate tumour evolution. The influence of the gut microbiome on breast cancer is an intriguing development which requires further intensive study. There is a paucity of biomarkers in the setting of hereditary breast cancer. The use of polygenic risk scores in this setting is an interesting development requiring further study.

The greatest challenge of all is to pull from such complexity key decision nodes that are clear enough to guide treatment decisions without losing the depth and richness of the information that underlies them. Seeking and finding this balance has been and will continue to be the holy grail of all endeavours in this field.

Introduction

The biological behaviour of breast cancer is remarkably heterogeneous; there is a large degree of inter-individual variability in the short and long term prognosis of breast carcinomas as well as in their responses to therapy. In order to maximise the benefits of a personalised approach to therapy, it is essential to have tools which can provide the necessary risk stratification to plan clinical management. A number of tools are well established, including the routine assessment of the ER, PR and HER-2 receptor status, the Nottingham Prognostic Index, molecular predictors of chemotherapy response such as the Oncotype recurrence score reported by Genomic Health and bioinformatic predictors of response such as the Predict tool. However, there is a need for predictive and prognostic tools to develop further. Biomarkers typically fall into one of three categories: diagnostic biomarkers which aid in the subtyping of a tumour, prognostic biomarkers which are used to assess patient outcomes, and predictive biomarkers which are used to assess the expected response to therapies and guide treatment. In practice, the lines between predictive and prognostic biomarkers are often blurred.

A Brief History of Breast Cancer Prediction and Prognosis

Breast cancer has been known since the ancient world. The earliest written description of the disease comes from the *Edwin Smith Papyrus*, which is a copy produced in 1500 to 1600 BCE of an original ancient Egyptian medical document believed to date to 3000 BCE. It remained a recalcitrant disease to treat until key turning points in the 18th and 19th centuries [1]. The pioneering work of Virchow in the 19th century subjected tumours to microscopic analysis, which began the process of classifying tumours on their histological appearance, and introduced the concept of tumour grade. In parallel with the assessment of tumour grade came the assessment of stage. The first TNM staging for breast cancer appeared in 1958.

The first biomarker of breast cancer was the estrogen receptor (ER), the importance of which was established in the 1960's, with tamoxifen, the first selective estrogen receptor modulator, becoming available in 1971. This was followed by the progesterone receptor (PR). Most recently, there has been growing interest in the androgen receptor as a biomarker and treatment target, both in the context of ER and PR positive breast cancers, and in the context of the luminal androgen receptor subtype of triple negative breast cancer.

Ki67 was discovered in 1983 as a nuclear protein expressed in proliferating cells; it is widely used, and there is strong evidence that a high Ki67 proliferative index is associated with worse outcomes. The full scale deployment of this biomarker has been hampered by reproducibility problems which will likely be overcome by digital pathology. Ki67 also forms a component of a range of prognostic systems which were subsequently developed, including the PEPI score for assessment of response to neoadjuvant endocrine therapy, and the Oncotype recurrence score. An emerging trend is toward the use of paired Ki67 measurements, comparing pre- and post neoadjuvant therapy proliferation indices to assess the response to treatment and inform treatment decisions on a case by case basis, particularly in the setting of neoadjuvant endocrine therapy.

The next important biomarker was HER-2, first described in 1987, with the first drug directed against this receptor, Herceptin, becoming available in 1998. Interestingly, the initial trials of Herceptin showed no benefit of the drug, because the data were analysed looking at all breast cancers as a group. It was only when the subgroup of patients who showed HER-2 amplification were analysed that the benefits of HER-2 blockade in this subgroup were observed – a classic salutary tale of subgroup effects.

In the 1990's, the importance of BRCA was demonstrated. In the early 2000s, the first molecular risk stratification tools emerged, and with them the evidence supporting treatment de-escalation. Most recently CDK4/6 has emerged as a therapeutic target in ER and PR positive and HER-2 negative disease (see [2] for review). In the setting of disease progression, detection of PI3KCA mutation has emerged as an important biomarker.

The first choice of treatment of breast cancer is surgery. However, the psychological impact of radical surgery has driven the search for modalities to achieve breast conservation, leading to the use of neoadjuvant chemotherapy in breast cancer. The evidence shows that the combination of taxanes and anthracyclines is most effective in the neoadjuvant setting.

Response to NACT can be assessed clinically, radiologically and pathologically. The pathological response to neoadjuvant chemotherapy is assessed by the pathologist on the excision specimen. There are many grading systems described, including, but not limited to, the Miller-Payne system, RCB system, Sataloff system and the AJCC ypTNM staging [3]. Internationally, the Miller-Payne system is the most

widely used, but there is no overall consensus on the best system, and each has its advantages and drawbacks.

Neoadjuvant chemotherapy can increase the surgical opportunity for patients with advanced disease and improve the breast conservation rate for patients with early disease. It can also improve prognosis. However, there are some patients who do not respond, and it is essential to identify these patients as early as possible.

The Predictive and Prognostic Implications of the Histological Classification

The histological classification and grading of the tumour itself presents valuable predictive and prognostic information, and can be stratified into six prognostic groups (see [4] for review):

1. Very indolent: pure low grade adenosquamous carcinoma, pure fibromatosis-like metaplastic carcinoma, pure low grade mucoepidermoid, adenoid cystic and secretory carcinomas. Also includes encapsulated and solid papillary carcinomas, which are regarded as *in-situ* lesions despite lacking myoepithelial cells.
2. Excellent prognosis group: low metastatic potential with mainly lymph node metastasis: pure tubular and cribriform carcinoma (<3cm).
3. Good prognosis group: Grade 1 invasive breast carcinoma (NST), tubulolobular, mucinous and invasive papillary carcinoma.
4. Moderate prognosis group: Grade 2 invasive breast carcinoma (NST), classical invasive lobular carcinoma.
5. Poor prognosis group: Grade 3 invasive breast carcinoma (NST), pleomorphic lobular carcinoma, micropapillary carcinoma, Grade 3 matrix producing and squamous metaplastic carcinomas.
6. Very poor prognostic group: Grade 3 invasive breast carcinoma (NST) of large size, Grade 3 spindle cell metaplastic carcinoma, small cell carcinoma.

Current Developments in Immunohistochemical Prediction: HER-2 low

Previously, the assessment of HER-2 status by immunohistochemistry divided the expression of the receptor into three categories: negative (scored as 0 or 1+), borderline (2+) and positive (3+) based on the staining pattern. Borderline cases were tested for HER-2 amplification and cases showing the amplification assessed as positive. The recent discovery that antibody drug conjugates are effective in patients with weak HER-2 protein expression has led to the definition of the HER-2 low group, defined as a 1+ pattern of staining on immunohistochemistry or tumours with a 2+ staining which do not show HER-2 amplification. This paradigm regards HER-2 low tumours as being HER-2 equipped as opposed to patients with the amplification whose tumours are HER-2 driven. There are HER-2 low carcinomas present across hormone receptor positive and triple negative carcinomas, with a larger proportion being found in hormone receptor positive carcinomas (65%) as compared with triple negative carcinomas (35%). Current practice will therefore be to classify tumours into HER-2 negative, HER-2 low and HER-2 positive. Revised guidelines have been issued but there are discrepancies in the details which will need to be resolved. It is unclear whether HER-2 low identifies HER-2 equipped tumours amenable to antibody conjugate therapy, or whether there are distinct molecular subgroups within HER-2 low tumours (see [5] for review).

Predictive and Prognostic Models of Breast Cancer Using Clinical Predictors

There has been a proliferation in predictive and prognostic models. A recent systematic review of breast cancer prognostic models identified 58 prognostic models derived from pathological data which are reported in routine clinical practice, developed between 1982 and 2016 , 28 of which predict mortality, 23 of which predict recurrence and 7 of which predict both [6]. The most commonly used predictors in these models are nodal status, tumour size, tumour grade, age at diagnosis and the ER status. As might be expected, there is a tendency for most of these models to perform well in the development cohorts, but the performance becomes less accurate when the models are applied to other populations, particularly young patients, the elderly and high risk patients. The notable exception is the Nottingham Prognostic Index which provides consistent predicting ability across a range of independent populations; it remains a mainstay of prediction for this reason [6].

The Nottingham Prognostic index is based on the assessment of tumour histology (tumour size, pathological lymph node stage, grade). A further index, termed the NPI+, has further built on the original index by building in the immunohistochemical assessment of ER, PR, HER-2, CK 5/6, CK 7/8, EGFR, HER-3, HER-4, p53 and Mucin 1 into the prediction model [7].

Looking more broadly at all prediction models, another recent systematic review identified 922 models using 228 predictors published between 2010 and 2020; application of the Prediction model Risk of Bias Assessment Tool (PROBLAST) identified that the majority of the prediction models examined are at high risk of bias, mainly due to problems identified in the analysis domain of the tool [8]. Only 35 of the models were developed using the appropriate statistical methods. Given that prediction models are highly likely to require certification as medical devices, this is a major problem.

There is a growing interest in the subset of risk prediction models which have been developed using artificial intelligence (AI) and machine learning. These models do appear to perform slightly better than those derived by more traditional approaches. Of the currently described models, the most commonly employed machine learning method was the neural network method. However, as with prediction models as a whole, machine learning tools are also at high risk of bias, as revealed by the PROBLAST tool. (see [9] for review).

Although a high risk of bias does not mean that a model has no clinical value, and a low risk of bias is no guarantee that a model will have clinical value, it is an aspect of model development which requires more rigorous attention going forward. This will require close adherence to clinical reporting guidelines, and particular attention to appropriate statistical analysis.

Molecular Risk Stratification Tools

The use of molecular risk stratification tools has focussed on the setting of women with early hormone receptor positive HER-2 negative breast cancers. It is known that some of these women will derive substantial benefit from adjuvant or neoadjuvant chemotherapy, whereas chemotherapy can be safely avoided in others. Identifying the latter group of patients in a safe and consistent way is therefore important. There are several validated commercially available molecular risk stratification tools available, which include Oncotype Dx, Mammaprint, Prosigna, Endopredict and the Breast Cancer Index. Each of these tools measures the expression level of a small subset of genes, some also incorporating clinical risk factors, to give a prognostic score.

Oncotype Dx uses the gene expression data of 21 genes to give a recurrence score (RS) from 0 to 100, which predicts the risk of recurrence over the next 9 years with endocrine therapy alone, adjusted for nodal status[10]. Of the 21 genes assessed, 16 are cancer related, and 5 are reference genes used for normalisation. The RSClin web tool allows the RS to be integrated with tumour size and grade[11]. Oncotype Dx has been prospectively validated for use in both pre- and postmenopausal patients with ER positive HER-2 negative disease which is either node negative or node positive with up to 3 positive nodes.

Mammaprint is a similar tool, validated for use in the setting of ER positive HER-2 negative disease, which uses the gene expression data of 70 genes to assign tumours to a low or high risk category [12]. Due to the low numbers of node-positive patients in the prospective trials of this tool, the evidence for this tool is confined to the setting of node negative disease[12]. However, the tool can be used in both ER positive and ER negative disease. The MINDACT trial demonstrated that this tool can identify a subgroup of patients who are at ultra-low risk of recurrence [13].

The Prosigna tool is based on the PAM50 tool and combines the tumour size with an assessment of the molecular subtype (luminal A, luminal B, HER-2 positive, basal-like) to generate a score which divides patients into low, intermediate and high risk groups[14]. The PAM50 measures the expression of 50 cancer related genes, 8 genes for normalisation, 6 for positive controls and 8 for negative controls.

Endopredict uses the gene expression data of 12 genes, 3 related to proliferation, 5 associated with hormone receptors, 3 reference genes for normalisation and one control gene, and combines these with clinical risk factors to divide tumours into low and high risk categories[15]. This tool is validated in the setting of postmenopausal patients with node negative or node positive disease, but is not validated for use in premenopausal patients.

The PEPI score, widely used in clinical trials, is a score used in the setting of neoadjuvant endocrine therapy and is derived from the Ki67 index, tumour size, lymph node status and ER expression. A PEPI score of 0 is associated with a low risk of recurrence without chemotherapy [16].

The Breast Cancer Index is an algorithmic signature based on two independent panels of biomarkers. The first, the Molecular Grade Index, assesses 5 proliferation genes. The second is an expression ratio of two genes which assesses estrogen signalling pathways[17]. The tool provides an overall risk of recurrence over 10 years and also a specific risk of late recurrence, defined as recurrence after 5 years. Some studies have suggested that this tool

can be used to predict the benefit of extended endocrine therapy after 5 years, although this predictive effect was not consistently validated in all studies[18].

Of the above tools, Oncotype Dx is the most widely recommended and used. Before the TAILORx trial, the National Surgical Adjuvant Breast and Bowel Project (NSABP) stratified patients into low (RS<18), intermediate (RS 18- <31) and high (RS ≥31) groups. Only the high risk group derived benefit from chemotherapy[19]. The benefit of chemotherapy in the intermediate risk patients was assessed by the TAILORx study, and showed that endocrine therapy alone was noninferior to combined adjuvant chemotherapy and endocrine therapy in patients with an RS of 11-25[20].

With so many commercially available tools, the question arises as to which is the superior test, or whether different tests perform better in certain settings. The OPTIMA preliminary study used five clinically validated tools, including Oncotype, MammaPrint and Prosigna, to compare how they performed against each other in risk stratification. They found a high level of disagreement between the tests. 60.6% of tumours were given a discordant risk category by at least one of the tools [21]. There may be many reasons for this. The thresholds used to define the risk categories are different, and the number of risk categories varies between 2 and 3 depending on the tool. The genes used in the tests are also different (Oncotype and MammaPrint cover 91 genes between them but share only three genes in common). The question as to which test is superior, or whether all of these tools will be superceded by other tools with a broader scope, remains unanswerable with current data. In the current landscape, the simultaneous use of multiple tools causes decreased rather than increased precision, and is therefore discouraged by current guidelines[20]. Of the current commercially available tools, Oncotype currently has the largest body of evidence from prospective trials.

Omics-Based Assessment

With the limitations of the current risk stratification tools, there is growing interest in omics-based approaches [22]. The hallmarks of cancer and their associated pathways present a vast array of complexity which can now be probed at multiple levels of organisation. Most studies of this kind integrate the main levels of the central dogma of molecular biology: genomics, transcriptomics and proteomics. Added to this are the epiomics of the central dogma: epigenetics (epigenetic DNA changes), epitranscriptomics (RNA modifications) and epiproteomics (post translational modifications). Another layer, given the metabolic changes that occur in cancer, is metabolomics. The proteomic assessment has been targeted in recent years to the tumour microenvironment, yielding, among many useful modalities, immunomics. Many biomarkers are in various stages of assessment based on these approaches. However, the integration of such exponentially expanding and vast information is a daunting task.

There is evidence across a range of cancers that artificial intelligence approaches perform well in the the integration of omics data. With regard to breast cancer, there are algorithms which show promise in the detection of breast cancer by the integration of cfDNA and proteomic data (cancerSEEK) and in the classification of breast cancer by the integration of mRNA expression proteomics and metabolomics (PROFILE) (see [23] for review).

Breast Cancer Heritability

Around 5-10% of breast cancers are familial, showing a strong family history of both breast and ovarian cancer. The most commonly encountered mutations in this group, accounting for 50% of familial breast cancers between them, are BRCA-1, BRCA-2, checkpoint kinase-2 (CHEK2) and PALB2[24]. Data on potential biomarkers in this group are sparse, and this is an area that requires more attention as these tumours have distinct behaviour. They occur earlier, show greater multifocality, and are more likely to be bilateral.

Inactivation of DNA repair pathways in these patients confers a worse prognosis. A key target which has significantly improved outcomes for this group is Poly ADP Ribose Polymerase (PARP). When PARP is inhibited, the base excision pair machinery for single-strand DNA break repair is not recruited, leading to a double strand break. If homologous recombination is intact, this defect is easily repaired. However, if homologous recombination is lost, and there is a BRCA-1 and BRCA-2 mutation, a pathway of genomic instability is triggered leading to cell death[24].

In addition to these highly penetrant autosomal dominant genes, there are a wider range of low penetrance gene variants which, when their influence is combined, explain up to 30% of breast cancer heritability. The majority are single nucleotide polymorphisms (SNPs) detected using Genome-Wide Association studies. The combined effect of these genes can be measured using Polygenic Risk Scores. A number of polygenic risk scores have been developed and described and there is evidence that these can be used to stratify

the risk of breast cancer development in their own right, and can improve risk stratification when combined with the established breast cancer risk factors. The latter context, which integrates the polygenic risk scores with clinical and lifestyle risk factors and mammographic density, is more likely to be clinically useful. They may therefore find a use in screening (see [25] for review). However, this would require a better understanding of the impact an assessment of high risk would have on the patient. Is there a risk reducing strategy that can be implemented? If so, what is it? Is it worth the psychological cost and the radiation exposure of a more intensive monitoring regime? A number of large scale studies are currently examining these questions.

The Tumour Microenvironment

It is well established that the microenvironment of the tumour is key to the tumour's behaviour. Current evidence supports the fact that a high infiltration of CD4 and CD8 positive tumour infiltrating lymphocytes, cytotoxic T-cells, FOXP3 positive regulatory T-cells before chemotherapy, type 1 tumour associated macrophages, B-cells, natural killer cells and dendritic cells are all associated with a better prognosis and a better response to NACT. FOXP3 positive regulatory T-cells after chemotherapy, type 2 tissue activated macrophages, bone-marrow derived suppressor cells, cancer stem cells, epithelial to mesenchymal transition, tumour infiltrating neutrophils, mast cells, adipocytes, cancer associated fibroblasts and neoplastic vessels are all associated with a worse prognosis and a worse response to NACT. This complexity reflects the role of the components of the microenvironment as a two-faced Janus in cancer. Some components contain and destroy the cancer, whereas others are co-opted by the tumour and aid in its progression (see [26] for review).

In the immune pathways, immune checkpoints PD-L1, CTLA-4 and the T-cell immunoglobulin and ITIM domain are associated with a better response to NACT. In contrast, T-cell immunoglobulin domain and mucin domain 3 and indole 2,3 oxygenase (which enhances the production and activity of some of the cancer-promoting immune cells) are associated with a worse response to NACT [26].

It is proposed that tumour infiltrating lymphocytes could be used to select patients for immunotherapy and also identify good prognostic subgroups who may benefit from de-escalation of chemotherapy. However, further studies and consistent refinement of the methods to assess TILs are needed. This includes better characterisation of the tumour microenvironment by identifying not only the lymphocyte subpopulations but also their spatial distribution, improved reproducibility of the assessment, and the use of machine learning approaches and AI to develop operator-independent methods of TIL assessment (see [27] for review).

PDL-1

There are well described immunomodulatory mechanisms which allow immunogenic tumours to achieve immune evasion and escape. Of the thousands of likely factors, the most well characterised is the PD-1 and PDL-1 interaction. PD-1 is strongly expressed by activated T cells, but binding of PDL-1 to PD-1 inhibits the normal activation of T-cells and ameliorates their tumour-killing actions. Inhibition of PDL-1 has emerged as a strategy to eliminate this pathway of immune evasion. This strategy is being increasingly used in a range of tumours. In the setting of breast cancer, its current use is in the setting of triple negative breast cancers. Atezolizumab in combination with nab-paclitaxel, and pembrolizumab combined with chemotherapy, are both currently being used in the setting of metastatic triple negative breast carcinoma [28].

The landscape of PDL-1 assessment as a biomarker is strangely muddled. Various antibodies acting on various epitopes are available, with expression detectable in both immune cells and tumour cells, depending on the antibody used. The scoring systems to define PDL-1 positivity are likewise variable, some focussing solely on the immune infiltrate with others requiring scoring of both immune cells and tumour cells. Because of the way the trials were conducted, each drug has been linked with its own companion diagnostic test (SP142 for atezolizumab and 22C3 for pembrolizumab; current additional drugs in development are linked with yet further different antibodies). This landscape has hampered robust biomarker development. There is a need for a clearly defined and standardised PDL-1 biomarker test to definitively assess the PDL-1 status in breast cancer. Such an approach would make more biological sense than the current situation, in which PDL-1 can be positive with one test but negative with another, restricting the choice of agent because of the linkage of the companion diagnostic to the drug. Furthermore, the latest evidence suggests that neoadjuvant pembrolizumab in combination with chemotherapy followed by adjuvant pembrolizumab improves event free survival for primary triple negative breast cancer regardless of the PDL-1 status. If further studies show similar findings, the need for the assessment of PDL-1 status may be called into question (see [28] for review).

The Intrinsic Molecular Subtypes of Breast Cancer

The subtypes of breast cancer are typically assessed using immunohistochemistry. However, these subtypes can also be identified by molecular analysis, and the PAM50 assay, which combines 50 standardised and reproducible gene assays, can identify the intrinsic subtype of breast cancer, dividing it into luminal A, luminal B, HER-2 positive and basal phenotype. Although there is a moderate correlation between the subtype assessment by immunohistochemistry and the intrinsic subtype, these are not as superimposable as is often assumed in clinical practice. It is this lack of precise correlation with immunohistochemistry that may make the PAM50 a useful test in some clinical settings, particularly when the behaviour of the tumour and response to therapy do not correlate with what the receptor status would predict (see [29] for review). For example, there is a subgroup of tumours which, despite showing high expression of ER and PR, show poor sensitivity to hormonal therapy, which the PAM50 can identify. It has been suggested that the PAM50 can identify HER-2 negative patients who are responsive to HER-2 therapy, although it remains to be seen whether, with the advent of the identification of HER-2 low patients, there is additional benefit of the PAM50 over immunohistochemistry [29].

Triple Negative Breast Cancer

As the most aggressive form of breast cancer, as well as the form for which the fewest new treatments have become available, there has been intense interest in biomarker discovery in this group of tumours. In 2011, expression profiling identified 6 subtypes of triple negative breast cancer. Revisions to these have since been proposed, most collapsing the classification into four rather than six subtypes, and there is no international consensus on the accepted classification to use. The six subtypes identified by the original analysis, and their potential treatment significance, are as follows (see [30] for review):

1. Basal-like 1: highly proliferative tumours with a high Ki67. Includes tumours with DNA response pathway aberrations. Likely to respond to PARP inhibitors and genotoxic agents.
2. Basal-like 2: aberrant growth factor signalling and aberrant myoepithelial marker expression. Growth factor and mTOR inhibitors are potential therapies.
3. Mesenchymal: this group incorporates the metaplastic carcinomas, showing aberrations of cell motility, cellular differentiation and extracellular receptor interaction. mTOR inhibitors are also a potential therapy in this group.
4. Mesenchymal stem-like: contains high levels of tumour-associated stromal cells; low expression of cell proliferation genes and high expression of stemness, angiogenesis and growth factor genes; PI3K inhibitors and antiangiogenic therapy may be treatment options in this group.
5. Immunomodulatory: characterised by high levels of tumour infiltrating lymphocytes. There is increased expression of immunological signalling genes, likely overexpressed in the immune infiltrate rather than the tumour. This group may be particularly responsive to immunotherapies such as PDL-1 inhibition, but further studies are required to probe this potential association.
6. Luminal Androgen Receptor: associated with active hormonal signalling, despite loss of ER and PR expression, with high expression of AR. Antiandrogen therapy is a potential therapy.

Because the expression patterns of the immunomodulatory and mesenchymal stem-like subgroups are likely derived from the tumour microenvironment rather than the tumour cells, some subsequent systems omitted these as specific subgroups, with the microenvironment-related changes captured in other ways (for example, dividing basal-like into immunosuppressive and immune activated subgroups) [30]. Further research is needed to achieve consensus on a molecular classification of triple negative breast cancer and most crucially to establish how such a classification can inform treatment.

There is intense interest in the information that whole genome sequencing of triple negative breast cancers can reveal beyond the molecular subclassification discussed above. Many of these studies have revealed similar genetic abnormalities to those highlighted by the molecular classification, namely TP53 mutations, immune response genes (particularly immunocheckpoints), PI3KCA mutations and DNA repair pathways (see [31] for review). However, novel abnormalities have also been revealed, including mutations in AURKA (for which a targeted therapy exists), MYC and JARID2. Whole genome sequencing is also beginning to address equity in the genetic assessment of these tumours. Triple negative breast cancers are twice as common in women of African or Hispanic descent than in white women. In women of African descent, EZH2 overexpression, BRCA1 alterations (including methylation) and BRCA20delaAAGA have emerged as specific signatures [31].

Circulating Tumour Cells (CTCs)

The traditional mainstay of breast cancer prediction has been the assessment of tissue biopsies. However, this poses challenges in the dynamic assessment of tumour evolution through the course of therapy because of the inability to account for tumour heterogeneity, the inaccessibility of metastatic lesions, and the difficulty of asking patients to repeatedly undergo invasive procedures to obtain tissue samples. For this reason, the use of liquid analytes detected in peripheral blood is receiving much attention. The earliest research focussed on the use of serum tumour markers, of which CEA, CA-15-3 and CA125 are the most widely used. Elevated levels of these biomarkers are associated with worse outcomes, but their utility in preoperative risk stratification remains controversial due to conflicting data in this setting. Furthermore, these biomarkers are unable to account for the effects of tumour heterogeneity and tumour evolution.

The discovery of circulating tumour cells (CTCs) has led to the potential of a liquid biopsy. Circulating tumour cells are living cells which break off from their tumour of origin or from a metastatic deposit and are shed into the bloodstream, where they can be detected. In addition to the amenability of CTCs to repeated sampling and treatment monitoring, the nature of CTCs as metastatic precursors make analysis of CTCs potentially superior to the focal snapshot of a tumour provided by a tissue biopsy (see [2] for review). Over 200 trials of CTCs in breast cancer are currently in various stages of progress.

The current landscape of CTCs is complex. There is strong evidence that the presence of higher numbers of CTCs is associated with a worse prognosis, particularly in the metastatic setting, but also in early disease. A recent meta analysis suggested that breast cancers can be divided into aggressive (≥5 CTCs) and indolent (<5 CTCs) subgroups [32]. The predictive capabilities of CTCs remain less clear. In early stage disease, the GeparQuattro and REMAGUS 02 trials did not show an association between CTCs and response in the main tumour. However, the presence of CTCs after therapy was shown, in the same studies, to predict early relapse [33,34].

In the metastatic setting, the results of current trials are equally mixed. Using the presence of persisting CTCs after one cycle of chemotherapy to guide switching therapy had no impact on overall survival [35]. Giving HER-2 directed therapy to patients with HER-2 negative primary tumours with newly emerging HER-2 positive CTCs showed only marginal benefits[36,37]. CTCs were more successful in detecting a subset of ER positive HER-2 negative tumours who would benefit from chemotherapy [38].

The major drawback of current technologies is the limited detection of CTCs due to their very low levels in peripheral blood (approximately one CTC per one billion red blood cells) and their short circulation time of 10 to 30 minutes. Developing a reliable and efficient method of CTC enrichment is essential. Various approaches which are being tried include implantable devices, cytapheresis (a method of cell fraction enrichment from large cell volumes) or sampling blood from the blood vessels directly draining the tumour (see [2] for review). Each of these approaches reintroduces a degree of invasiveness to the sampling, with the approach of sampling vessels draining the tumour requiring surgical access to the tumour. Another complicating factor is that CTC extravasation has a circadian rhythm, so the timing of sampling is critical [39]. There is evidence that molecular aspects of CTCs can be assessed and that it is feasible to do drug testing on CTCs [40].

CTCs remain a promising avenue of development and it is likely that the intense research attention this approach is receiving will overcome the technical hurdles of this technology.

Circulating Tumour DNA (ctDNA)

All cells release DNA into the bloodstream, including tumour cells, and the tumour cell-derived fraction of circulating DNA is referred to as circulating tumour DNA (ctDNA). Circulating tumour DNA holds the same attraction as CTCs for the same reasons. Various applications for ctDNA have been established, including early detection, diagnosis, prediction of response to neoadjuvant chemotherapy, disease monitoring and to detect resistance. There is emerging data from a range of trials that ctDNA can be used to risk stratify patients in the neoadjuvant setting, with detection and persistence of ctDNA being associated with resistance and metastatic recurrence (see [2] for review). However, most of the ctDNA in the blood does not show mutations. Possible solutions include a shift toward epigenetic assessment of ctDNA, and the assessment of the fragmentation patterns of cfDNA, which can yield useful information without the need for mutational information (see [41] for review). Finally, the omics based approaches discussed above can be applied in this setting to assess the circulome.

MicroRNAs

MicroRNAs (miRNAs) are small non-coding endogenous RNAs which inhibit translation by

binding to the UTR 3' of the target mRNA. They can be isolated in fixed tissues, blood, saliva and urine, making them attractive biomarkers. A broad range of miRNAs have been investigated in breast cancer, and show promise in the settings of both early detection of the primary tumour, and early detection of metastatic disease, with some potentially predicting the metastatic site [24]. Most of the research in this field is in the earlier phases of biomarker discovery, and it is likely that the clinical application would involve the use of panels of predictive miRNAs, as no single miRNA has emerged as a clear biomarker to focus on.

Gastrointestinal Microbiota

There is emerging evidence that the composition of the gut microbiota can influence the response to chemotherapy and that chemotherapy, in turn, can influence the composition of the gut microbiota. The human digestive system contains approximately 100 trillion bacteria. A recent systematic review of nine studies reveals some intriguing trends [42]. A high alpha diversity, which reflects a high diversity in the microbiome and is regarded as an indicator of good gut health, is associated with better responses to neoadjuvant chemotherapy, whereas low alpha diversity is associated with resistance. The beta diversity refers to differences in the composition of the microbiome between samples. Interestingly, one of the most consistent biomarkers to emerge from the limited evidence currently available is *Bacteroides*, which is associated with a worse response to neoadjuvant chemotherapy. A number of organisms have been found to be associated with better responses to neoadjuvant chemotherapy, and these have tended not to be consistent between studies. The broader range of these organisms likely reflects the finding that higher alpha diversity overall is favourable. Beta diversity was also used to stratify patients based on tumour size, grade, axillary lymph node metastasis and TNM stage [43].

An observational study of the use of antibiotics during cancer chemotherapy showed that patients who do not receive antibiotics have higher rates of pathological complete response, and reduced disease free and overall survival [44].

It is intriguing to note that each histological subtype of breast carcinoma is associated with a unique breast microbial profile [45], pointing to the role of breast microbiota in the tumour microenvironment and providing a potential mechanistic link between the composition of the gut microbiome and the behaviour of the tumour in the remote site.

Larger scale studies will be needed to harness the power of this promising avenue of exploration, in particular to identify the specific combinations of microbiota which are beneficial and harmful.

Polypoid Giant Cancer Cells

Polypoid giant cancer cells are the tumour-specific example of the phenomenon of polyploidy or whole-genome duplication, an adaptive mechanism which, despite some costs to cellular function, provides a mechanism for broad adaptations to occur to stressors. In the context of cancer, polypoid giant cancer cells can restructure the genetic and epigenetic landscape of the tumour cells as well as the tumour microenvironment. Measuring biomarkers in these cells provides a way to anticipate the evolution of the tumour. Many of the genes shown to be overexpressed in these cells in breast cancer are, unsurprisingly, genes in the proliferative and apoptotic pathways well known to be involved in breast cancer pathogenesis and progression. There are some genes which have been shown to induce the formation of polypoid giant cancer cells, which include CDC25C, and its upstream regulation by p38MAPK-ERK-JNK. The genes Aurora A and Aurora B are suppressors of polypoid giant cell formation, and inhibition of these genes has been shown to induce polypoid giant cell formation (see [46,47] for review). The key to developing these as biomarkers will be to find specific markers to identify polypoid giant cells which can be easily applied in the clinical setting.

Conclusion

Breast cancer prediction and prognosis is a rapidly advancing field, but still anchored by the established clinical risk factors, histological classification and assessment, and the use of immunohistochemistry and in-situ hybridisation. The Nottingham Prognostic Index remains the most reproducible and robust prognostic tool. Molecular risk stratification strategies are likely to evolve with the use of the liquid biopsy to respond dynamically to tumour evolution, and the development of liquid biopsy approaches will benefit from the omics approaches powered by AI. Omics approaches in tissue samples can also be used to reconcile the differences and limitations of the more focussed molecular stratification tools currently in use. The development of future risk stratification tools will need to take great care in using the proper statistical approaches and to avoid bias. It is likely that AI will play a key role in developing future algorithms, and there is evidence that AI-based tools perform slightly better.

There is potential for prediction and treatment of triple negative carcinomas to evolve with the use of a molecular subclassification, but further studies are

needed. The insights gained by whole genome sequencing in triple negative carcinomas set the stage for whole genome sequencing to come online for clinical use in these tumours, as has already happened for other tumour types in other tissues.

The literature abounds with potential biomarkers and targets, many of which may come to be of benefit to breast cancer patients. The tumour microenvironment will be a key focus of further work. PDL-1 inhibition has been the first application of this approach to enter the clinic but there will likely be many more. The intriguing role of the gut microbiota is a promising avenue of exploration which requires much more attention.

Breast cancer prediction and prognosis needs to be holistic, and account for multiple levels of organisation. The technology to gather this information and to integrate it is available, although technical and analytical challenges remain. The greatest challenge of all is to pull from such complexity key decision nodes that are clear enough to guide treatment decisions without losing the depth and richness of the information that underlies them. Seeking and finding this balance has been and will continue to be the holy grail of all endeavours in this field.

References

1. Elson NC, Lewis JD, Shaughnessy EA, Reyna C. Lessons from other fields of medicine, Part 1: Breast cancer. *Handb Clin Neurol*. 2023;192:101-118. doi:10.1016/B978-0-323-85538-9.00003-1
2. Holler A, Nguyen-Strauli BD, Frauchiger-Heuer H, Ring A. "Diagnostic and Prognostic Biomarkers of Luminal Breast Cancer: Where are We Now?". *Breast Cancer (Dove Med Press)*. 2023;15:525-540. doi:10.2147/BCTT.S340741
3. Members of Breast Cancer Expert Panel on C. [Expert panel consensus on pathological diagnosis of breast cancer with neoadjuvant therapy, the 2020 version]. *Zhonghua Bing Li Xue Za Zhi*. Apr 8 2020;49(4):296-304. doi:10.3760/cma.j.cn112151-20200102-00007
4. Rakha EA, Tse GM, Quinn CM. An update on the pathological classification of breast cancer. *Histopathology*. Jan 2023;82(1):5-16. doi:10.1111/his.14786
5. Popovic M, Silovski T, Krizic M, Dedic Plavetic N. HER2 Low Breast Cancer: A New Subtype or a Trojan for Cytotoxic Drug Delivery? *Int J Mol Sci*. May 4 2023;24(9)doi:10.3390/ijms24098206
6. Phung MT, Tin Tin S, Elwood JM. Prognostic models for breast cancer: a systematic review. *BMC Cancer*. Mar 14 2019;19(1):230. doi:10.1186/s12885-019-5442-6
7. Green AR, Soria D, Powe DG, et al. Nottingham prognostic index plus (NPI+) predicts risk of distant metastases in primary breast cancer. *Breast Cancer Res Treat*. May 2016;157(1):65-75. doi:10.1007/s10549-016-3804-1
8. Hueting TA, van Maaren MC, Hendriks MP, Koffijberg H, Siesling S. The majority of 922 prediction models supporting breast cancer decision-making are at high risk of bias. *J Clin Epidemiol*. Dec 2022;152:238-247. doi:10.1016/j.jclinepi.2022.10.016
9. Gao Y, Li S, Jin Y, et al. An Assessment of the Predictive Performance of Current Machine Learning-Based Breast Cancer Risk Prediction Models: Systematic Review. *JMIR Public Health Surveill*. Dec 29 2022;8(12):e35750. doi:10.2196/35750
10. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*. Dec 30 2004;351(27):2817-26. doi:10.1056/NEJMoa041588
11. Sparano JA, Crager MR, Tang G, Gray RJ, Stemmer SM, Shak S. Development and Validation of a Tool Integrating the 21-Gene Recurrence Score and Clinical-Pathological Features to Individualize Prognosis and Prediction of Chemotherapy Benefit in Early Breast Cancer. *J Clin Oncol*. Feb 20 2021;39(6):557-564. doi:10.1200/JCO.20.03007
12. Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *N Engl J Med*. Aug 25 2016;375(8):717-29. doi:10.1056/NEJMoa1602253
13. Lopes Cardozo JMN, Drukker CA, Rutgers EJT, et al. Outcome of Patients With an Ultralow-Risk 70-Gene Signature in the MINDACT Trial. *J Clin Oncol*. Apr 20 2022;40(12):1335-1345. doi:10.1200/JCO.21.02019
14. Wallden B, Storhoff J, Nielsen T, et al. Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. *BMC Med Genomics*. Aug 22 2015;8:54. doi:10.1186/s12920-015-0129-6
15. Dubsky P, Brase JC, Jakesz R, et al. The EndoPredict score provides prognostic information on late distant metastases in ER+/HER2- breast cancer patients. *Br J Cancer*. Dec 10 2013;109(12):2959-64. doi:10.1038/bjc.2013.671
16. Ellis MJ, Tao Y, Luo J, et al. Outcome prediction for estrogen receptor-positive breast cancer based on postneoadjuvant endocrine therapy tumor characteristics. *J Natl Cancer Inst*. Oct 1 2008;100(19):1380-8. doi:10.1093/jnci/djn309
17. Cardoso F, Kyriakides S, Ohno S, et al. Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. Oct 1 2019;30(10):1674. doi:10.1093/annonc/mdz189
18. Noordhoek I, Treuner K, Putter H, et al. Breast Cancer Index Predicts Extended Endocrine Benefit to Individualize Selection of Patients with HR(+) Early-stage Breast Cancer for 10 Years of Endocrine Therapy. *Clin Cancer Res*. Jan 1 2021;27(1):311-319. doi:10.1158/1078-0432.CCR-20-2737
19. Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol*. Aug 10 2006;24(23):3726-34. doi:10.1200/JCO.2005.04.7985
20. Sparano JA, Gray RJ, Makower DF, et al. Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. *N Engl J Med*. Jul 12 2018;379(2):111-121. doi:10.1056/NEJMoa1804710
21. Bartlett JM, Bayani J, Marshall A, et al. Comparing Breast Cancer Multiparameter Tests

in the OPTIMA Prelim Trial: No Test Is More Equal Than the Others. *J Natl Cancer Inst*. Sep 2016;108(9)doi:10.1093/jnci/djw050

22. Neagu AN, Whitham D, Bruno P, Morrissiey H, Darie CA, Darie CC. Omics-Based Investigations of Breast Cancer. *Molecules*. Jun 14 2023;28(12)doi:10.3390/molecules28124768
23. He X, Liu X, Zuo F, Shi H, Jing J. Artificial intelligence-based multi-omics analysis fuels cancer precision medicine. *Semin Cancer Biol*. Jan 2023;88:187-200. doi:10.1016/j.semcancer.2022.12.009
24. Deb S, Chakrabarti A, Fox SB. Prognostic and Predictive Biomarkers in Familial Breast Cancer. *Cancers (Basel)*. Feb 20 2023;15(4)doi:10.3390/cancers15041346
25. Roberts E, Howell S, Evans DG. Polygenic risk scores and breast cancer risk prediction. *Breast*. Feb 2023;67:71-77. doi:10.1016/j.breast.2023.01.003
26. Zhang W, Xu K, Li Z, Wang L, Chen H. Tumor immune microenvironment components and the other markers can predict the efficacy of neoadjuvant chemotherapy for breast cancer. *Clin Transl Oncol*. Jun 2023;25(6):1579-1593. doi:10.1007/s12094-023-03075-y
27. 27. Valenza C, Taurelli Salimbeni B, Santoro C, Trapani D, Antonarelli G, Curigliano G. Tumor Infiltrating Lymphocytes across Breast Cancer Subtypes: Current Issues for Biomarker Assessment. *Cancers (Basel)*. Jan 26 2023;15(3)doi:10.3390/cancers15030767
28. Emens LA, Loi S. Immunotherapy Approaches for Breast Cancer Patients in 2023. *Cold Spring Harb Perspect Med*. Apr 3 2023;13(4)doi:10.1101/cshperspect.a041332
29. Falato C, Schettini F, Pascual T, Braso-Maristany F, Prat A. Clinical implications of the intrinsic molecular subtypes in hormone receptor-positive and HER2-negative metastatic breast cancer. *Cancer Treat Rev*. Jan 2023;112:102496. doi:10.1016/j.ctrv.2022.102496
30. Lu B, Natarajan E, Balaji Raghavendran HR, Markandan UD. Molecular Classification, Treatment, and Genetic Biomarkers in Triple-Negative Breast Cancer: A Review. *Technol Cancer Res Treat*. Jan-Dec 2023;22:15330338221145246. doi:10.1177/15330338221145246
31. Tierno D, Grassi G, Scomersi S, et al. Next-Generation Sequencing and Triple-Negative Breast Cancer: Insights and Applications. *Int J Mol Sci*. Jun 2 2023;24(11)doi:10.3390/ijms24119688
32. Cristofanilli M, Pierga JY, Reuben J, et al. The clinical use of circulating tumor cells (CTCs) enumeration for staging of metastatic breast cancer (MBC): International expert consensus paper. *Crit Rev Oncol Hematol*. Feb 2019;134:39-45. doi:10.1016/j.critrevonc.2018.12.004
33. Pierga JY, Bidard FC, Mathiot C, et al. Circulating tumor cell detection predicts early metastatic relapse after neoadjuvant chemotherapy in large operable and locally advanced breast cancer in a phase II randomized trial. *Clin Cancer Res*. Nov 1 2008;14(21):7004-10. doi:10.1158/1078-0432.CCR-08-0030
34. Bidard FC, Peeters DJ, Fehm T, et al. Clinical validity of circulating tumour cells in patients with metastatic breast cancer: a pooled analysis of individual patient data. *Lancet Oncol*. Apr 2014;15(4):406-14. doi:10.1016/S1470-2045(14)70069-5
35. Smerage JB, Barlow WE, Hortobagyi GN, et al. Circulating tumor cells and response to chemotherapy in metastatic breast cancer: SWOG S0500. *J Clin Oncol*. Nov 1 2014;32(31):3483-9. doi:10.1200/JCO.2014.56.2561
36. Jacot W, Cottu P, Berger F, et al. Actionability of HER2-amplified circulating tumor cells in HER2-negative metastatic breast cancer: the CirCe T-DM1 trial. *Breast Cancer Res*. Nov 14 2019;21(1):121. doi:10.1186/s13058-019-1215-z
37. Pestrin M, Bessi S, Puglisi F, et al. Final results of a multicenter phase II clinical trial evaluating the activity of single-agent lapatinib in patients with HER2-negative metastatic breast cancer and HER2-positive circulating tumor cells. A proof-of-concept study. *Breast Cancer Res Treat*. Jul 2012;134(1):283-9. doi:10.1007/s10549-012-2045-1
38. Bidard FC, Jacot W, Kiavue N, et al. Efficacy of Circulating Tumor Cell Count-Driven vs Clinician-Driven First-line Therapy Choice in Hormone Receptor-Positive, ERBB2-Negative Metastatic Breast Cancer: The STIC CTC Randomized Clinical Trial. *JAMA Oncol*. Jan 1 2021;7(1):34-41. doi:10.1001/jamaoncol.2020.5660
39. Diamantopoulou Z, Castro-Giner F, Schwab FD, et al. The metastatic spread of breast cancer accelerates during sleep. *Nature*. Jul 2022;607(7917):156-162. doi:10.1038/s41586-022-04875-y

40. Yu M, Bardia A, Aceto N, et al. Cancer therapy. Ex vivo culture of circulating breast tumor cells for individualized testing of drug susceptibility. *Science*. Jul 11 2014;345(6193):216-20. doi:10.1126/science.1253533
41. Gianni C, Palleschi M, Merloni F, et al. Cell-Free DNA Fragmentomics: A Promising Biomarker for Diagnosis, Prognosis and Prediction of Response in Breast Cancer. *Int J Mol Sci*. Nov 17 2022;23(22)doi:10.3390/ijms232214197
42. Csendes D, Gutlapalli SD, Prakash K, et al. Gastrointestinal Microbiota and Breast Cancer Chemotherapy Interactions: A Systematic Review. *Cureus*. Nov 2022;14(11):e31648. doi:10.7759/cureus.31648
43. Terrisse S, Derosa L, Iebba V, et al. Intestinal microbiota influences clinical outcome and side effects of early breast cancer treatment. *Cell Death Differ*. Sep 2021;28(9):2778-2796. doi:10.1038/s41418-021-00784-1
44. Zhang X, Yu L, Shi J, et al. Antibiotics modulate neoadjuvant therapy efficiency in patients with breast cancer: a pilot analysis. *Sci Rep*. Jul 7 2021;11(1):14024. doi:10.1038/s41598-021-93428-w
45. Tzeng A, Sangwan N, Jia M, et al. Human breast microbiome correlates with prognostic features and immunological signatures in breast cancer. *Genome Med*. Apr 16 2021;13(1):60. doi:10.1186/s13073-021-00874-2
46. Neves Rebello Alves L, Dummer Meira D, Poppe Merigueti L, et al. Biomarkers in Breast Cancer: An Old Story with a New End. *Genes (Basel)*. Jun 28 2023;14(7)doi:10.3390/genes14071364
47. Casotti MC, Meira DD, Zetum ASS, et al. Computational Biology Helps Understand How Polyploid Giant Cancer Cells Drive Tumor Success. *Genes (Basel)*. Mar 26 2023;14(4)doi:10.3390/genes14040801