Molecular diagnosis in breast cancer

Fresia Pareja
Caterina Marchiò
Jorge S Reis-Filho

Abstract

Breast cancer is a complex and heterogeneous disease, encompassing a plethora of entities with distinct biological features and clinical behaviour. The advent of high throughput molecular methods has allowed a systematic characterization of the genomic landscape of breast cancer, revealing a profound heterogeneity in this disease. These methods are having a profound effect on the understanding of breast cancer. Some have already been incorporated in clinical practice, such as the prognostic 'gene signatures' that allow the tailoring of therapy in the subgroup of patients with oestrogen receptor (ER)-positive and HER2-negative breast cancer. In this review, we discuss the contribution of the main molecular methods in breast cancer research and how this information is changing our approaches to the diagnosis and management of this disease. We also address novel developments in the diagnosis and management of HER2-positive breast carcinomas and familial breast cancer.

Keywords *in situ* hybridization; liquid biopsies; molecular taxonomy; precision medicine; prediction; prognosis; prognostic signatures

Molecular pathology techniques have had a dramatic effect on the diagnosis of haematological malignancies and soft tissue sarcomas. This has led to a paradigm shift in the way entities are defined: from purely morphological, descriptive classification systems to a combined histopathological and molecular taxonomy. Many haematological malignancies are defined by specific recurrent chromosomal translocations and/or molecular aberrations.

The contribution of molecular pathology to the study of most types of carcinomas has been less profound. With the boom of high throughput technologies and increasingly coherent data on the molecular features of epithelial malignancies, molecular techniques are becoming an integral part of the armamentarium of surgical pathology laboratories.

Breast cancer has been more extensively studied with molecular methods than any other epithelial malignancy. Some of the 'molecular-era' breakthroughs have been translated into methods amenable to histopathologists (e.g. immunohistochemistry) and made 'prime time' in diagnostic practice (e.g. E-cadherin immunohistochemistry to differentiate between lobular neoplasia and low-grade solid ductal carcinoma *in situ*). Molecular data have also confirmed the concept that breast cancer is a heterogeneous disease, comprising several histological types with distinct biological features and different clinical behaviour. Understanding the molecular features of breast cancer may potentially provide additional diagnostic, prognostic and predictive information that may, in the not so distant future, facilitate the development of tailored therapy.

The main contribution of molecular methods for understanding breast cancer will be addressed, and the way this information is changing the management of breast cancer will be contextualized in this review.

Molecular classification of breast cancer

Microarray-based gene expression profiling has solidified the notion that breast cancer, rather than being a single disease, represents a group of entities with different molecular alterations and clinical behaviour. Seminal studies by the Stanford group led to the classification of breast cancer into four intrinsic subtypes: luminal A, luminal B, HER2-enriched and basal like.[1] Later on, additional molecular subtypes of breast cancer were identified, such as the claudin-low[2] and the molecular apocrine.[3]

The ER-positive group comprises the luminal A and luminal B tumours, which are characterized by the expression of ER, genes pertaining to the ER pathway, and other transcripts usually found in luminal epithelial cells. The prognosis of luminal tumours is largely determined by the expression of proliferation-related genes. Luminal B cancers display higher levels of genes pertaining to the proliferation cluster than luminal A tumours, and have a worse prognosis.[4,5] Luminal tumours show intrinsic heterogeneity. Along these lines, luminal A cancers can be further stratified into four subgroups, with different copy number alterations, somatic mutations profiles and clinical outcomes, including the copy number-high subgroup, which displays high genomic instability, recurrent *TP53* mutations and overactivation of Aurora kinases, and is associated with a worse clinical outcome.[6] The ER-negative cluster encompasses the HER2-enriched subgroup, characterized by high levels of expression of genes pertaining to the HER2 amplicon (17q11); the basal-like subgroup, characterized by the expression of genes expressed by basal/myoepithelial cells, such as basal cytokeratins; claudin-low tumors, which are enriched for genes related to cancer stem cells, epithelial to mesenchymal transition and immune response[2]; and the molecular apocrine subgroup, which shows increased androgen signaling and a molecular apocrine gene expression profile.[3] This molecular taxonomy of breast cancer has important clinical implications, as the different molecular subtypes display distinct biology, responses to therapy and clinical outcomes.[7]

Triple negative breast cancer (TNBC) shows a vast intertumour heterogeneity, and seven molecular subtypes have been put forward by Lehmann et al.,[8] including the basal-like 1 (BL1), basal-like 2 (BL2), mesenchymal (M), mesenchymal stem-like (MSL), immunomodulatory (IM), luminal androgen receptor (LAR), and unstable subgroups.[8] Nonetheless, it was later shown, by the same group, that the transcriptomic profiles of the

Fresia Pareja ***MD PhD*** *is an Instructor in Pathology at Memorial Sloan Kettering Cancer Center, New York, NY, USA. Conflicts of interest: none declared.*

Caterina Marchiò ***MD PhD*** *is Assistant Professor of Pathology at the University of Turin, Italy. Conflicts of interest: none declared.*

Jorge S Reis-Filho ***MD PhD FRCPath*** *is the Director of Experimental Pathology at Memorial Sloan Kettering Cancer Center, New York, NY, USA. Conflicts of interest: none declared.*

IM and the MSL subgroups might not derive from tumour cells, but rather stem from tumour infiltrating lymphocytes and stromal cells, respectively,[9] indicating that the most parsimonious number of TNBC molecular subtypes is four. The clinical implication of this classification was confirmed by studies showing that the different TNBC subgroups significantly differ in terms of their response to neoadjuvant chemotherapy.[9] The rate of pathologic complete response (pCR) for BL1 tumours is much higher than the one for BL2 and LAR tumours.[9]

An alternative molecular classification of breast cancer, based on the integrative analysis of copy number alterations and gene expression, was put forward by the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC), which categorized breast cancer into ten integrative clusters (IntClusts).[10] A gene expression method for the classification of breast cancer into the different IntClusts was later developed,[11] and provided an independent validation of the clinical relevance of this classification, as breast cancers corresponding to the different IntClusts displayed varying responses to neoadjuvant chemotherapy and different clinical outcomes.[11]

Prognostic gene signatures in breast cancer

The identification of patients who will benefit from adjuvant chemotherapy remains challenging. Multigene prognostic tests have become useful tools in the determination of the risk of recurrence and in the decision making of whether or not chemotherapy should be spared for some patients.[12] Whilst first generation prognostic assays, such as Mammaprint[13] and Oncotype Dx[14] have a better predictive power for recurrences within the first five years, more recent tests, such as Prosigna,[15] Endopredict,[16] and the Breast Cancer Index (BCI)[17] have good predictive power both for early and late recurrences (Table 1).[18]

The utility of multigene assays is limited in ER-negative breast cancer, as most cases are classified as "high-risk" due to their elevated proliferative rates,[5] restricting the prognostic value of multigene assays to ER-positive disease. Of note, all these five tests may be performed using formalin-fixed paraffin-embedded (FFPE) samples, facilitating their widespread use, and whilst Mammaprint, Oncotype Dx and BCI should be performed by central laboratories, Prosigna and EndoPredict may be set up in local pathology laboratories.[18]

Oncotype Dx is a reverse transcriptase-PCR (RT-PCR) assay which measures the relative expression of 21 genes, including 16 cancer-related genes and five reference genes, and computes a recurrence score (RS) from zero to 100, assigning individual patients into the low- (RS $<$ 18), intermediate- (RS 18–30) and high-risk (RS $\geq$ 31) categories,[14] which determines the risk of distant recurrence at 10 years and the benefit of the addition of chemotherapy in ER-positive, HER2-negative, node-negative breast cancer patients treated with tamoxifen. Interestingly, a recurrence score predicted by the integration of morphologic and immunohistochemical parameters including histologic grade, receptor status, tumour size and Ki67 expression can predict the Oncotype Dx RS with relative accuracy.[19] The clinical utility of Oncotype DX was validated by the initial results of the TAILORx study.[20] MammaPrint is a DNA microarray-based prognostic assay for patients younger than 61 years old with stage I or II ER-positive node-negative breast cancers.[13] This assay entails the evaluation of the expression of 70 genes, enabling the stratification of patients into low-risk and high-risk categories, and its utility was validated by the prospective randomized phase III MINDACT trial.[13] Prosigna is an RT-PCR based assay which, using the NanoString technology, measures the expression of 50 classifier genes from the PAM50 molecular classification algorithm and of 5 control genes, and computes a risk of recurrence (ROR) score, placing patients into the low-, intermediate- or high-risk categories, depending on their 10 year-risk of distant recurrence, which correlates with the intrinsic subtype of the case.[15] Its use was approved for the prediction of distant recurrence-free survival in postmenopausal women with stage I and stage II ER-positive breast cancer treated with adjuvant hormone therapy. EndoPredict is an RT-PCR based assay which calculates a risk score based on the expression of eight cancer-related genes and three reference genes, allowing the stratification of patients with early ER-positive breast cancer treated with adjuvant endocrine therapy alone, into high-risk and low-risk groups for 10 year-recurrence.[16] The integration of EndoPredict score with tumour size and nodal status allows the computing of EPclin, a comprehensive risk score, which has been validated in the ABCSG-6 and ABCSG-8 randomyzed phase III trials.[16] Lastly, BCI is an RT-PCR based assay which quantifies the expression ratio of *HOXB13* and *IL17BR*,[21] and integrates it with the molecular grade index (MGI), which assesses the expression of five genes, related to tumour grade and proliferative status.[17] It was designed for the identification of patients with early ER-positive, node-negative breast cancer receiving adjuvant hormone therapy at a high risk of recurrence. Its prognostic utility was validated in postmenopausal patients with early ER-positive breast cancer from the Stockholm trial.[22]

List of commercially available prognostic gene signature assays that are clinically useful in the context of ER+/HER2- disease

	Mammaprint	Oncotype Dx	Prosigna	EndoPredict	Breast Cancer Index
Method	Microarrays	qRT-PCR	NanoString	qRT-PCR	qRT-PCR
Feasibiity on FFPE samples	Yes	Yes	Yes	Yes	Yes
Type of assessment	Central laboratory	Central laboratory	Local laboratory	Local laboratory	Central laboratory
Level I evidence	Yes, IA	Yes, IA	Yes, IB	Yes, IB	Yes, IB
Information regarding the molecular subtype	No	No	Yes	No	No

Table 1