Molecular classification of breast cancer: What the pathologist needs to know

Emad A Rakha and Andrew R Green

Academic Pathology, Division of Cancer and Stem Cells, School of Medicine, University of Nottingham and Nottingham University Hospital NHS Trust, Nottingham, UK

Correspondence

Emad Rakha

Department of Histopathology, Division of Cancer and Stem Cells, School of Medicine, The University of Nottingham and Nottingham University Hospitals NHS Trust, Nottingham City Hospital, Nottingham, NG5 1PB, UK. Email: emadrakha@yahoo.com

Key words: breast cancer, molecular taxonomy, gene expression profiling, luminal, HER2, basal-like, next generation sequencing, molecular prognostic assays.

ABSTRACT

Breast cancer (BC) is a heterogeneous disease featuring distinct histological, molecular and clinical phenotypes. Although traditional classification systems utilising clinicopathological and few molecular markers are well-established and validated they remain insufficient to reflect the diverse biological and clinical heterogeneity of BC. Advancements in high-throughput molecular techniques and bioinformatics have contributed to the improved understanding of BC biology, refinement of molecular taxonomies and the development of novel prognostic and predictive molecular assays. Application of such technologies is already underway, and is expected to change the way we manage BC. Despite the enormous amount of work that has been carried out to develop and refine BC molecular prognostic and predictive assays, molecular testing is still in evolution. Pathologists should be aware of the new technology and be ready for the challenge. In this review, we provide an update on the application of molecular techniques with regard to BC diagnosis, prognosis and outcome prediction. Current contribution of the emerging technology to our understanding of BC is also highlighted.

INTRODUCTION

Historically breast cancer (BC) was classified based on clinicopathological features mainly tumour stage, and grade. Other morphological features such as histological type, proliferation status and lymphovascular invasion are also recognised as important morphological prognostic variables that reflect tumour biology (1, 2). Over time, knowledge about BC biology has significantly increased and led to the understanding that BC represents a heterogeneous group of tumours and that tumour behaviour and response to therapy is determined by the underlying biological features. The expression of oestrogen receptor (ER), progesterone receptor (PgR) and the human epidermal growth factor receptor 2 (HER2) that were originally identified as predictive of response to systemic therapy are now recognised to be the main determinants of BC biology and can be used to refine BC molecular and prognostic taxonomy. More recently, molecular data arising from a variety of high throughput techniques have been used to refine BC stratification and develop prognostic and predictive classification with the aim of individualised therapy.

Although molecular taxonomy of BC based on gene expression profiling, proteomics, DNA copy number alteration and chromosomal changes, mutation status, methylation and microRNAs has been expanding for many years and has increased our knowledge of BC biology, its clinical application remains limited. The introduction of next generation sequencing (NGS) or massively parallel sequencing (3) appears to have opened new avenues for decoding BC molecular complexity, refine molecular classification and identify new therapeutic targets. These molecular techniques hold promise for improving

diagnosis, prediction of outcome and behaviour, and in aiding selection of therapies for individual patients (4). However, its clinical utility is still under investigation (5).

Pathologists are currently using conventional and novel molecular techniques on routine practice to help diagnosis of morphologically challenging entities, to assess the expression of hormone receptors and HER2 status on every BC and help oncologists to refine the prognostic stratification of BC and complement the morphological variables with molecular biomarkers. Although immunohistochemistry (IHC) remains the most commonly used conventional molecular technique, other techniques are increasingly used in routine practice including *in situ* hybridisation (ISH), RT-PCR, and in some centres NGS and expression microarrays. In the research setting, several other molecular techniques are used including comparative genomic hybridisation (CGH), expanded immunohistochemistry with tissue microarrays and proteomics. In this review, the main applications of molecular techniques on BC are highlighted with emphasis on the practical applications which can be generally divided into three main categories; diagnosis, molecular prognostic and predictive taxonomies.

Using molecular biomarkers in the diagnosis of breast lesions

In addition to prognosis and treatment response prediction, molecular biomarkers are frequently used in the diagnosis of challenging breast lesions; to differentiate between benign and malignant entities, *in situ* and invasive tumours, subtyping of certain lesions and determination of the tissue of origin of less differentiated malignant tumours. The most frequent technique utilising in this aspect is IHC often using a panel of biomarkers (6, 7).

IHC plays a useful role in diagnosing spindle cell lesions, identification of myoepithelial cells, differentiate between ductal and lobular phenotype and between hyperplastic epithelial proliferative process from neoplastic clonal epithelial proliferation and in the classification of papillary lesions. Cytokeratins can be used to detect small nodal metastases or subtle invasive carcinomas such as invasive lobular carcinomas. IHC also is helpful in recognising metastases to the breast and mammary carcinomas metastasising to extramammary tissues. Different antibodies are useful for different tumours: PAX8 and WT1 for ovarian carcinoma; TTF1 for thyroid and pulmonary adenocarcinoma; melan-A, HMB45 and S100 for melanoma; and lymphoid markers for lymphoma. Specific genetic translocations are also helpful for diagnosis of certain breast lesions (see below) and for exclusion of specific soft tissue tumours when identified on a biopsy as a component of other mammary-specific lesions; for instance pure stromal component of a malignant phyllodes tumour to be differentiated from other soft tissue sarcomas that may have different management strategies (8).

Companion diagnostics in breast cancer

The ability to predict an individual's response to a specific therapy is the main aim in modern precision medicine. A molecular diagnostic tool in the field of cancer therapy was first used in the 1970s to predict response of BC to the selective oestrogen receptor modulator, tamoxifen based on the expression of ER (9). Currently, several targeted cancer therapies are utilised in standard oncological care and this field is expanding. As a result, the concept of "companion diagnostics" has emerged which can be defined as a diagnostic test used as a companion to a therapeutic drug to determine its applicability to a specific patient. Currently, the US Food and Drug Administration (FDA)-approved companion

diagnostics are utilised in BC tests for the presence of HER2 protein overexpression or gene amplification. Despite not considered companion diagnostics by the FDA, ER and PgR testing is mandatory for effective hormone therapy decision making and can be considered as companion diagnostics in BC. Although prognostic multigene assays are not companion diagnostics *per se*, as they are not linked to a particular drug, they can result in changes in clinical decisions and treatment course based on their outcome predictions (Table 1).

HORMONE RECEPTOR TESTING: Hormone receptor status is determined by the tumour cells' expression of nuclear receptors for oestrogen (ER) and progesterone (PgR). Biochemical ligand-binding assays were initially used to detect ER and PgR, but they required fresh tissue and were technically challenging and therefore IHC assays have become routine. Different scoring methods are in use for determining the level of expression but the most widely used systems are the Allred scoring and the histochemical score (H-score) methods which both assess the proportion and intensity of staining that are summed to give an overall score. However, the currently agreed cut-off of positivity of ER and PgR for management purpose relies on proportion scoring and is 1% (10). Patients with BC showing any nuclear expression of hormone receptor in invasive tumour cells above the cut-off are likely to respond to hormone therapy and are therefore potential candidates for this therapy. However, for a diagnostic purpose, i.e. determination of a mammary origin of a metastatic carcinoma, a more stringent definition of positivity is often used based on the pathologist's discretion. Although current guidelines indicate that IHC is used for determination of hormone receptor status (10) in BC, ER and PgR are component genes of some multigene assays including Oncotype DX. Information

regarding hormone receptor status using these assays can be used as an additional quality measures for assessment methods. Discrepancy of results should trigger a reflex test.

HER2 TESTING: HER2 is overexpressed in 12% to 20% of BC most often because of *HER*2 gene amplification. Because of its predictive value, guideline recommendations for its assessment (11) and their updated versions (12, 13) have been published to provide guidance on HER2 testing in BC. Key aspects of these guidelines include a recommendation that all BC be tested for HER2 using IHC and subsequently with ISH in borderline positive IHC cases using a validated test. It should be recognised that both IHC and ISH represent an attempt to convert a continuous biological variable into a dichotomous category and borderline or equivocal cases exist and a reflex test is recommended to reduce the proportion of these cases. The use of the updated definition of positivity of HER2 has reduced the proportion of these borderline cases (12, 13).

Ki67 PROLIFERATION INDEX: The Ki67 proliferation index has been investigated as a BC prognostic and predictive factor in various settings (14). Ki67 is assessed in routine practice using IHC however, its analytic validity remains a matter of debate and formal inter-and intra-laboratory standardisation hampers its use in routine practice for management decision (15). Ki67 can be used in routine practice to i) determine the proliferation status in poorly fixed specimens, or 2) stratify grade 2 tumours into two prognostically distinct classes (16) akin to the molecular grade index (17). Ki67 is also used a component of some prognostic tools (18) however; the published 2016 American Society of Clinical Oncology (ASCO) clinical practice guideline on BC (15) recommends that Ki67 labelling index determined by IHC should not be used to guide choice on adjuvant

chemotherapy with intermediate quality of evidence base and moderate strength of recommendation.

GENETIC TESTS AND DIAGNOSIS: Some diagnostic microarray-based gene expression tests were developed for identification of cancer tissue of origin. These include the Pathwork Tissue of Origin Test that was developed using a 2000-gene classification model for identification of tumour tissue of origin with an overall accuracy in identifying the primary site of poorly differentiated tumours up to 90% (19) and the THEROS Cancer TYPE ID® which is a RT-PCR-based test using 92 genes and FFPE samples (20).

Importantly, some special type mammary carcinomas show specific translocations which characterise these tumours and can be used as diagnostic adjunct. Secretory carcinoma of the breast is characterised by a balanced translocation of genetic material between chromosomes 12 and 15 (t(12;15)) creating a new gene in which the 5' region of ETV6 is fused to the 3' region of NTRK3 producing ETV6-NTRK3 fusion gene (57). Mucoepidermoid carcinoma which is a rare type of metaplastic BC is characterised by a translocation between chromosome 11 and 19 (t(11;19)(q21;p13)) creating a novel fusion product between mucoepidermoid carcinoma translocated 1 (MECT1) and Mastermind-like gene family (MAML2); MECT1-MAML2 fusion gene (58). Adenoid cystic carcinomas as well as cylindroma show a specific translocation t(6;9)(q22-23;p23-24) creating MYB-NFIB fusion gene (59). In a study of breast adenoid cystic carcinoma mixed with a high grade triple negative BC components, the MYB-NFIB fusion gene was detected in both tumour subtypes and it was postulated that the progression from adenoid cystic carcinoma to high-grade triple-negative BC of no special type may involve the

selection of neoplastic clones and/or the acquisition of additional genetic alterations with enrichment of mutations affecting certain genes such as FGFR1 (21).

Prognostic and predictive taxonomies

Molecular classification of breast cancer

BC has been classified based on the expression of biomarkers using a variety of techniques, concepts and applications. Based on the expression of individual biomarkers, BC can be classified into ER positive and ER negative, HER2 positive and HER2 negative. Although this appears as a simplified molecular classification system, it remains as the most important and informative molecular BC taxonomy to date for clinical management in routine practice (15). These two markers with or without addition of other biomarkers; namely PgR and Ki67 can be used in combination to provide further important prognostic information (22, 23). For instance the response of ER positive HER2 negative tumours to hormone therapy is different to ER positive HER2 positive tumours. Despite the predictive and prognostic value of hormone receptors and HER2, complex molecular classifications based on multiple markers utilising high-throughput techniques have attracted attention as a novel method for molecular taxonomy. Molecular classification of BC was initially investigated using loss of heterozygosity analysis (LOH), karyotyping and CGH, which identified key genomic alterations including losses, gains and amplifications of genomic DNA (24-27). This provided the early framework for a molecular classification system that stratified BC into distinct classes. Global gene expression profiling (GEP) studies of BC using unsupervised clustering techniques have provided a more established molecular classification system and identified distinct clusters or intrinsic subtypes based on the quantitative expression of several genes (transcriptome profiles) (28, 29). Subsequent

class discovery studies have also reported an association between molecular intrinsic subtypes and patient outcome and that these classes are associated with distinct biological pathways making them potential candidates for targeted therapy.

In the pioneer GEP study by Perou and colleagues in 2000 (28) using the expression of a subset (n=496) of differentially expressed genes termed the 'intrinsic' gene set, it was demonstrated that BC at the transcriptome level is not a single disease. Despite the fact that each individual tumour features a unique GEP related to its specific biological features and genetic abnormalities, tumours clustered together to produce distinct reproducible classes based on transcriptomic profiles with common overlapping features. In Perou's study (28) two main clusters were identified and appeared to be related to ER expression. The ER+ cluster was enriched with ER, ER-related genes and other genes characteristic of the luminal epithelial cells and this class was termed as Luminal to indicate its molecular similarity to them. The other major class contained ER-negative tumours and showed three distinct subclasses termed HER2-positive, basal-like and normal breast-like. The HER2 subgroup was characterised by overexpression of HER2 and other genes pertaining to the HER2 amplicon. The basal-like class was largely characterised by the lack of expression of ER and HER2 and by positive expression of genes characteristic of basal-like cells of the breast and by high proliferative activity. The normal breast-like class displayed a triple-negative phenotype but did not cluster with the basal-like centroid and was characterised by expression profiles similar to those found in normal breast tissue.

Subsequent GEP studies indicated that the luminal class, which comprises the majority of BC is heterogeneous with respect to the expression of other genes and outcome (30). The

Luminal cluster was further stratified into subclasses with at least two distinct subclasses reported in many studies; luminal A and luminal B subtypes. Most studies indicated that luminal B tumours are associated with a worse prognosis than tumours of the luminal A class however, the molecular definition was variable and not reproducible. In general it was characterised by ER expression but with higher proliferation rates and/or HER2 expression and low or absent PR expression (31). Other luminal subclasses have been described including luminal C (32) and luminal N (33) but the classification into luminal A and B remains the most validated sub-classification despite the limitations described above. Similar to the luminal class, some studies have classified basal-like tumours into several subgroups. In a previous study of 587 triple negative BC, Lehman and colleagues (34) reported six subtypes displaying unique GEP. These include basal-like I, basal-like II, an immunomodulatory, a mesenchymal, a mesenchymal stem-like, and a luminal androgen receptor (AR) subtype. Other authors have reported four subtypes of triple negative BC (luminal AR, mesenchymal, basal-like immune-suppressed, and basal-like immune-activated) (35) and we have split them into those that have high or low p53 expression (33). In addition to their clinical relevance, the molecular intrinsic subtypes showed distinct pattern of genomic alterations, emphasising the divergent biological characteristics of tumours from these classes. For instance, the luminal A class has the greatest number and diversity of significantly mutated genes, with *PIK3CA* at 45% being the most frequent. Luminal B cancers showed mutations affecting both *TP53* and *PIK3CA*. A high proportion of tumours in the HER2 class show a high frequency of *TP53* and *PIK3CA* mutations. HER2-luminal-like tumours had higher expression of genes such as *GATA3*, *BCL2* and *ESR1* and higher frequency of *GATA3* mutations. *TP53* mutation is most frequent in the basal-like cancers, with most of the significantly mutated genes in

luminal tumours being absent. Whilst *TP53* mutations are present in the basal-like and HER2 tumours, the type of mutation in this gene differs between subtypes (36).

In recent years several consortia were launched, with researchers from around the world collaborating to map the genomes of BC and other cancer. These consortia and other research groups started using the 'multi-omic' approaches involving different technology platforms: genomic DNA copy number arrays, DNA methylation arrays, exome sequencing, mRNA arrays, microRNA sequencing and reverse-phase protein arrays to develop a more global and integrated 'picture' of BC. An example of this genomics-driven classification of BC based on an integrative analysis of GEP and genome-wide copy number alterations (CNAs) was reported by the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) (37). This study of 2,000 BC reported that the number of molecular subtypes is likely to be 10, which are called "integrative clusters" and that these subtypes showed distinct clinical behaviour. The Cancer Genome Atlas (TCGA) network (38) have analysed 466 BC using five platforms; genomic DNA copy number arrays, DNA methylation, exome sequencing, mRNA arrays, microRNA sequencing and reverse phase protein arrays. The integrated information across platforms demonstrated the existence of four main BC classes, identified two novel protein expression defined subgroups related to stroma / microenvironment's elements and provided key insights into previously-defined gene expression subtypes. They have also identified specific signalling pathways dominant in each molecular subtype and hypothesised that much of the clinically observable heterogeneity and plasticity occurs within, and not across, these major molecular subtypes of BC (38).

To overcome the problems of fresh tissue, the availability of microarray-based technology, cost and assay reproducibility, other techniques such as RT-PCR and IHC coupled with tissue microarrays using a smaller set of genes have been introduced to replicate this molecular taxonomy and to identify intrinsic subtypes in routine practice. Two main approaches have been identified. The first approach was based on identifying a minimum gene sets from microarray-based studies and used the minimum set of genes that can reliably identify the GEP defined classes. One successful example is the PAM50 (Prediction of Microarray using 50 classifier genes plus 5 reference genes) classifier (39) that categorises BC into four intrinsic subtypes; luminal A, luminal B, HER2-enriched, and basal-like. The other surrogate approach to identify intrinsic BC subtypes includes using tissue microarrays and IHC utilising a large panel of biologically relevant biomarkers and then applying unsupervised clustering techniques to identify molecular classes with and without comparison with GEP defined molecular subtypes. In a previous study we have applied 25 IHC biomarkers to 1,076 unselected BC series (40) and identified seven molecular classes primarily based on ER and HER2 expression. For a practical use in clinical routine practice, the number of biomarkers was reduced for to 10 biomarkers which produced comparable classification power (33). As the performance of clinicopathological factors varies among the molecular classes, the concept of refining the traditional Nottingham Prognostic Index (NPI) that utilises grade, size and lymph node status applied equally to BC cases regardless of the molecular features. NPI Plus (NPI+) was based on classifying BC into seven distinct molecular classes using the 10 biomarkers followed by incorporation of clinicopathological variables to identify distinct prognostic groups with each of the classes (33). Using the NPI+ formulae, through incorporating

molecular features and clinicopathological parameters, an improved patients' outcome stratification was achieved superior to the traditional NPI (33).

Although the identification of the intrinsic subtype-based molecular classification of BC has attracted attention, and improved our understanding of BC biology and increased hope in refinement of BC therapy prediction, their application in routine practice has been less successful. Targeted therapy of BC still relies of ER and HER2 regardless of the molecular class of the tumour; for instance HER2 positive BC patients are candidates for HER2 targeted therapy regardless of the intrinsic class whether HER2-enriched or luminal. Despite the limited clinical applicability, GEP has opened new avenues for refinement of BC molecular prognostication as it has led to the introduction of the molecular multigene assays that aim to identify subgroups of BC associated with outcome or specific response to therapy (41). This approach is based on identification of a set of genes (gene signature) that can be used collectively to identify tumours with specific biological or clinical features. The term "genomic signatures" was used to refer to the expression of a set of genes in a biologic sample using microarray technology while "metagene" refers to a single aggregate measure of the expression of a group of genes that usually show coordinated expression in a set of samples and defined by mathematical combination of the genes of interest. Most of these multigene assays were used in BC to stratify prognostically clinically relevant groups into low and high risk subgroups to guide further treatment. Although these molecular classification systems have provided fascinating new insights into BC biology and they may have provided more prognostic and better predictive power than conventional variables and complement them, we still have a long way to go in terms of delivering truly personalised medicine and further work is needed.

The first multigene prognostic assay was developed by *van't Veer* et al. (42) who used a class prediction approach utilising a 70-gene set associated with the likelihood of metastasis within 5 years. This 70-gene signature was validated in a subsequent study (42) and was later commercially marketed as the MammaPrint assay (Agilent, Amsterdam, the Netherlands). Other gene signatures have been developed based on prediction of outcome or response to specific therapy and used as prognostic and predictive signatures in the clinical context used in their development and validation. The most commonly available multigene assays include PAM50 risk of recurrence score (Prosigna kit) (43), Oncotype DX assay (44), Breast Cancer Index (BCI) (45), EndoPredict (46) and MammaPrint score [4]. Other studies have attempted to generate multigene predictors based on a hypothesis derived from *in vivo* or *in vitro* experiments or on genes characteristic of a biological process and then applied to BC samples (47, 48). Examples include genes associated with host immune responses, wound healing and other stromal gene signatures that carry prognostic value independent of ER status and proliferation and may represent candidate predictive markers for targeted therapies (49) (Table 2).

Despite the minimal overlap between various gene signatures, most of them show clinically significant risk stratification particularly in the clinically indeterminate group of ER-positive, HER2-negative and lymph node-negative or with low nodal burden disease (15, 50). These multigene prognostic assays are used to stratify BC into distinct prognostic groups; high risk and low risk groups with intermediate risk group in some tests. Patients in the low risk group can avoid chemotherapy while patients in the high risk group are considered as a candidate for chemotherapy. Current evidence indicates that these multigene prognostic assays have limited clinical application in ER negative, HER2 positive

and advanced stage tumours as patients with these tumours are typically offered chemotherapy(15).

Molecular classification of special breast tumour types

Most of the molecular profiling data of BC relate to ductal NST carcinomas which comprise approximately 75% of BC and their diagnosis is one of exclusion, when a tumour does not fit into a defined special subtype. Comprehensive molecular analysis of ILC (51) which is the most common special BC type revealed that besides E-cadherin loss which is the best known ILC genetic hallmark there are specific mutations targeting PTEN, TBX3, and FOXA1 as ILC enriched features. PTEN loss associated with increased AKT phosphorylation appeared to be highest in ILC among all breast cancer subtypes. Global gene expression profiling revealed the existence of 3 subtypes of ILC; reactive-like, immune-related and proliferative classes (51). These subtypes showed many significant genomic features at the mRNA and protein/phosphoprotein level with 1,277 genes differentially expressed between ILC subtypes. However, no difference between these ILC subtypes was identified in terms somatic mutations or DNA copy-number alterations. As expected the proliferative subtype was associated with the worst outcome whilst the reactive-like was associated with the best outcome (51). At the DNA ILC cases were significantly enriched for CDH1 mutations and mutations affecting TBX3 and FOXA1. GATA3 mutations appeared to be the second most discriminant event between ILC and ductal NST carcinoma after CDH1 mutations. In addition homozygous losses of the PTEN locus (10q23) and PTEN mutations were more frequent in ILC (51).

Analysis of pure mucinous BC subtype indicated that they show a relatively low level of genetic instability and they tend to be homogeneously and preferentially clustered together, separately from ductal NST carcinomas. They less frequently harbour gains of 1q and 16p and losses of 16q and 22q than grade- and ER-matched ductal NST, and no pure mucinous carcinoma displayed concurrent 1q gain and 16q loss, a hallmark genetic feature of low-grade ductal NST (52). Pure invasive micropapillary carcinoma that has a characteristic morphological appearance with a so-called inside-out growth pattern shows specific copy number aberrations (53), high cyclin D1 expression, high proliferation rates, and MYC (8q24) amplification (54) compared to ER-matched and grade-matched ductal NST.

Special subtypes that belong to the basal-like subgroup include carcinomas with medullary features, as well as metaplastic carcinomas and salivary–gland-like tumours such as adenoid cystic carcinoma. Adenoid cystic carcinoma forms an interesting paradox as it sits within the basal-like group, which is generally regarded as of poor prognosis, yet its clinical behaviour is generally indolent. This underscores the astonishing heterogeneity that can occur even within individual intrinsic subtypes. In a previous study of acinic cell carcinomas (ACCs) of the breast using massively parallel sequencing (55), our group identified that the most frequently mutated gene is *TP*53 with a complex patterns of gains and losses similar to those of common forms of triple negative BC. Additional somatic mutations affecting breast cancer-related genes found in ACCs included *PIK*3CA, m*TOR*, *CTNNB*1, *BRCA*1, *ERBB*4, *ERBB*3, *INPP*4B, and *FGFR*2. Using NGS approach, our group also demonstrated that microglandular adenosis/atypical microglandular adenosis, particularly those associated with triple negative BC harboured at least one somatic non-

synonymous mutation with identical *TP*53 mutations and similar patterns of gene CNAs in microglandular adenosis and in the associated triple negative BC. Clonal shifts in the progression from microglandular adenosis to atypical microglandular adenosis and/or to triple negative BC were also observed. On the other hand pure microglandular adenosis lacked clonal non-synonymous somatic mutations and displayed limited copy number alterations (56). Importantly, these findings, in conjunction with others, underscore the significance for microglandular adenosis in clinical diagnosis. In another study of infiltrating epitheliosis using the same techniques (57), we demonstrated high prevalence of somatic mutations affecting PI3K pathway genes, suggesting that these lesions may be neoplastic rather than hyperplastic. The landscape of somatic genetic alterations found in infiltrating epitheliosis is similar to that of radial scars/complex sclerosing lesions, suggesting that they may represent one end of this spectrum of lesions.

There is also a strong evidence to indicate that the considerable molecular heterogeneity of BC is already present at the *pre-invasive* level with genomic, transcriptomic and phenotypic similarities found between ductal carcinoma *in situ* (DCIS) and coexisting invasive carcinoma (58). Similar to invasive BC, frequent genetic and genomic events have been reported in DCIS with several studies provided detailed descriptions of DCIS genomic, transcriptomic and proteomic profiling however, to date there are relatively little molecular data that can be used to predict the risk of progression to invasive tumour or risk of recurrence.

Next generation sequencing (NGS)

The introduction of NGS or massively parallel sequencing (MPS) has revolutionised BC genetics and genomics and is expected to assist in utilising for personalised treatment of BC patients. Common approaches to NGS include whole-genome sequencing (sequences the complete genome of a sample), whole-exome sequencing, targeted exome sequencing (target-enrichment methods to capture genes of interest), and hotspot (sequences selected regions/regions with recurrent mutations of selected genes of interest) sequencing. NGS has been used to characterise genomic alterations such as copy number changes, insertions/ deletions and mutations, facilitate sequencing at a greater depth (at the base-pair level) allowing the identification of subclonal mutations and help distinguishing the "driver" mutations that contribute to cancer development from the "passengers" mutations that do not appear to play a significant role in disease progression. In addition to providing information about the genomic landscape of BC, MPS has confirmed both inter-tumour and intra-tumour heterogeneity and showed that each BC is largely unique.

NGS have indicated that the mutation frequencies found in BC are lower than some other cancer such as lung squamous cell carcinoma or bladder urothelial carcinomas but are similar to those of ovarian and renal clear cell carcinomas. Somatic driver point mutations and/or copy number changes were identified in at least 40 cancer genes with a maximum of 6 mutated cancer genes in an individual BC though 28 cases of the 100 BC showed a single driver (59). Seven of those 40 cancer genes (*TP53, PIK3CA, MYC, ERBB2, FGFR1, CCND1 and GATA3*) were mutated in >10% of cases and these contributed 58% of driver mutations (59). Overall, BC was found to have a mean of 56.9 (range 5–374) somatic mutations per cancer. NGS has also demonstrated spatial and temporal intratumour heterogeneity of BC at a level beyond common expectations. Various degrees of intratumour

genetic heterogeneity have been demonstrated in BC even in the absence of overt histological phenotypic heterogeneity. Triple negative and basal-like tumours tend to have greater intra-tumoural heterogeneity than non-basal-like tumours. Mutations in common driver genes such as *TP53*, *PIK3CA,* and *PTEN* are usually found in high clonal frequencies and several somatic mutations are present in only a fraction of cancer cells. NGS also showed that the constellations of somatic mutations found between a primary BC and its metastases (temporal heterogeneity) and between distinct areas within the primary tumour (spatial heterogeneity) are not identical providing further evidence to indicate that BC evolve over the course of the disease. This clonal genetic heterogeneity may explain resistance of some BC to selective environmental pressures and therapy.

An increasing number of molecularly targeted drugs are available in the clinic as approved drugs or in the context of clinical trials (http://www.clinicaltrials.gov) and these include HER4, EGFR, VEGF, VEGFR, FGFR, KIT, BRAF, mTOR, PDGFR, MEK, TIE2, FLT3, SRC, RET, PD1, PDL1 and others (36). These drugs target specific molecular abnormalities, including mutated protein kinases and amplified or rearranged genes. BCs that carry any of these abnormalities particularly if they harbour the sensitising genomic abnormality is expected to respond to the corresponding targeted therapies. For example, the HER2 gene-amplified BC benefit from HER2-targeted therapies.

Future perspectives: As a natural extension of the increasing application of the high-throughput sequencing technology, the list of cancer driver genes is growing, and a considerable number of these are potentially targetable. This may also help to understand the mechanisms underlying treatment failure. Furthermore, the identification of targets holds

great potential for monitoring clonal evolution in response to treatment and, hence, the early detection of treatment failure. Application of such technologies is already underway is expected to result in further refinement of BC prognostication and prediction of response to specific therapies.

In conclusion: Molecular testing has become increasingly important in the prevention, diagnosis, and treatment of BC. Despite the enormous amount of work that has been carried out to develop and refine BC molecular classification, it is still in evolution. With the increasing use of more sophisticated high-throughput techniques such as NGS, large amounts of data will continue to emerge, which could potentially lead to identification of novel therapeutic targets and allow more precise classification systems that can predict outcome and response to therapy.

REFERENCES

1. Rakha EA, Reis-Filho JS, Baehner F, *et al*. Breast cancer prognostic classification in the molecular era: the role of histological grade. Breast Cancer Res 2010;12:207.
2. Aleskandarany MA, Green AR, Benhasouna AA, *et al*. Prognostic value of proliferation assay in the luminal, HER2-positive, and triple-negative biologic classes of breast cancer. Breast Cancer Res 2012;14:R3.
3. Torres TT, Metta M, Ottenwalder B, Schlotterer C. Gene expression profiling by massively parallel sequencing. Genome Res 2008;18:172-7.
4. Ng CK, Schultheis AM, Bidard FC, Weigelt B, Reis-Filho JS. Breast cancer genomics from microarrays to massively parallel sequencing: paradigms and new insights. J Natl Cancer Inst 2015;107.
5. Alix-Panabieres C, Pantel K. Challenges in circulating tumour cell research. Nat Rev Cancer 2014;14:623-31.
6. Lee AH. Use of immunohistochemistry in the diagnosis of problematic breast lesions. J Clin Pathol 2013;66:471-7.
7. Rakha EA, Aleskandarany MA, Lee AH, Ellis IO. An approach to the diagnosis of spindle cell lesions of the breast. Histopathology 2016;68:33-44.
8. Fisher C. The diversity of soft tissue tumours with EWSR1 gene rearrangements: a review. Histopathology 2014;64:134-50.
9. Fisher B, Redmond C, Brown A, *et al*. Influence of tumor estrogen and progesterone receptor levels on the response to tamoxifen and chemotherapy in primary breast cancer. J Clin Oncol 1983;1:227-41.
10. Hammond ME, Hayes DF, Dowsett M, *et al*. American Society of Clinical Oncology/ College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. J Clin Oncol 2010;28:2784-95.
11. Harris L, Fritsche H, Mennel R, *et al*. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. J Clin Oncol 2007;25:5287-312.
12. Rakha EA, Pinder SE, Bartlett JM, *et al*. Updated UK Recommendations for HER2 assessment in breast cancer. J Clin Pathol 2015;68:93-9.
13. Wolff AC, Hammond ME, Hicks DG, *et al*. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. J Clin Oncol 2013;31:3997-4013.
14. Yerushalmi R, Woods R, Ravdin PM, Hayes MM, Gelmon KA. Ki67 in breast cancer: prognostic and predictive potential. Lancet Oncol 2010;11:174-83.
15. Harris LN, Ismaila N, McShane LM, Hayes DF. Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. J Oncol Oncol 2016;34:1134-50.
16. Aleskandarany MA, Rakha EA, Macmillan RD, *et al*. MIB1/Ki-67 labelling index can classify grade 2 breast cancer into two clinically distinct subgroups. Breast Cancer Res Treat 2011;127:591-9.
17. Ma XJ, Salunga R, Dahiya S, *et al*. A five-gene molecular grade index and HOXB13:IL17BR are complementary prognostic factors in early stage breast cancer. Clin Cancer Res 2008;14:2601-8.
18. Wishart GC, Rakha E, Green A, *et al*. Inclusion of KI67 significantly improves performance of the PREDICT prognostication and prediction model for early breast cancer. BMC Cancer 2014;14:908.
19. Monzon FA, Lyons-Weiler M, Buturovic LJ, *et al*. Multicenter validation of a 1,550-gene expression profile for identification of tumor tissue of origin. J Clin Oncol 2009;27:2503-8.

20. Ma XJ, Patel R, Wang X, *et al*. Molecular classification of human cancers using a 92-gene real-time quantitative polymerase chain reaction assay. Arch Pathol Lab Med 2006;130:465-73.
21. Fusco N, Geyer FC, De Filippo MR, *et al*. Genetic events in the progression of adenoid cystic carcinoma of the breast to high-grade triple-negative breast cancer. Mod Pathol 2016;29:1292-305.
22. Rakha EA, Reis-Filho JS, Ellis IO. Combinatorial biomarker expression in breast cancer. Breast Cancer Res Treat 2010;120:293-308.
23. Senkus E, Kyriakides S, Ohno S, *et al*. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2015;26 Suppl 5:v8-v30.
24. Reis-Filho JS, Simpson PT, Gale T, Lakhani SR. The molecular genetics of breast cancer: the contribution of comparative genomic hybridization. Pathol Res Pract 2005;201:713-25.
25. Roylance R, Gorman P, Hanby A, Tomlinson I. Allelic imbalance analysis of chromosome 16q shows that grade I and grade III invasive ductal breast cancers follow different genetic pathways. J Pathol 2002;196:32-6.
26. Jones C, Ford E, Gillett C, *et al*. Molecular cytogenetic identification of subgroups of grade III invasive ductal breast carcinomas with different clinical outcomes. Clin Cancer Res 2004;10:5988-97.
27. Yao J, Weremowicz S, Feng B, *et al*. Combined cDNA array comparative genomic hybridization and serial analysis of gene expression analysis of breast tumor progression. Cancer Res 2006;66:4065-78.
28. Perou CM, Sorlie T, Eisen MB, *et al*. Molecular portraits of human breast tumours. Nature 2000;406:747-52.
29. Sorlie T, Perou CM, Tibshirani R, *et al*. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869-74.
30. Sorlie T, Tibshirani R, Parker J, *et al*. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proceedings of the National Academy of Sciences of the United States of America 2003;100:8418-23.
31. Habashy HO, Powe DG, Abdel-Fatah TM, *et al*. A review of the biological and clinical characteristics of luminal-like oestrogen receptor-positive breast cancer. Histopathology 2012;60:854-63.
32. Sotiriou C, Neo SY, McShane LM, *et al*. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci U S A 2003;100:10393-8.
33. Rakha EA, Soria D, Green AR, *et al*. Nottingham Prognostic Index Plus (NPI+): a modern clinical decision making tool in breast cancer. Br J Cancer 2014;110:1688-97.
34. Lehmann BD, Bauer JA, Chen X, *et al*. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest 2011;121:2750-67.
35. Burstein MD, Tsimelzon A, Poage GM, *et al*. Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. Clin Cancer Res 2015;21:1688-98.
36. Santarpia L, Bottai G, Kelly CM, *et al*. Deciphering and Targeting Oncogenic Mutations and Pathways in Breast Cancer. Oncologist 2016.
37. Curtis C, Shah SP, Chin SF, *et al*. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012;486:346-52.
38. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature 2012;490:61-70.
39. Parker JS, Mullins M, Cheang MC, *et al*. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 2009;27:1160-7.
40. Abd El-Rehim DM, Ball G, Pinder SE, *et al*. High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses. Int J Cancer 2005;116:340-50.

41. van 't Veer LJ, Dai H, van de Vijver MJ, *et al*. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530-6.
42. van de Vijver MJ, He YD, van't Veer LJ, *et al*. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999-2009.
43. Dowsett M, Sestak I, Lopez-Knowles E, *et al*. Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. J Clin Oncol 2013;31:2783-90.
44. 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 2004;351:2817-26.
45. Sgroi DC, Sestak I, Cuzick J, *et al*. Prediction of late distant recurrence in patients with oestrogen-receptor-positive breast cancer: a prospective comparison of the breast-cancer index (BCI) assay, 21-gene recurrence score, and IHC4 in the TransATAC study population. Lancet Oncol 2013;14:1067-76.
46. Muller BM, Keil E, Lehmann A, *et al*. The EndoPredict Gene-Expression Assay in Clinical Practice - Performance and Impact on Clinical Decisions. PLoS One 2013;8:e68252.
47. Sotiriou C, Wirapati P, Loi S, *et al*. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst 2006;98:262-72.
48. Teschendorff AE, Miremadi A, Pinder SE, Ellis IO, Caldas C. An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. Genome Biol 2007;8:R157.
49. Winslow S, Leandersson K, Edsjo A, Larsson C. Prognostic stromal gene signatures in breast cancer. Breast Cancer Res 2015;17:23.
50. Kelly CM, Bernard PS, Krishnamurthy S, *et al*. Agreement in risk prediction between the 21-gene recurrence score assay (Oncotype DX(R)) and the PAM50 breast cancer intrinsic Classifier in early-stage estrogen receptor-positive breast cancer. Oncologist 2012;17:492-8.
51. Ciriello G, Gatza ML, Beck AH, *et al*. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. Cell 2015;163:506-19.
52. Lacroix-Triki M, Suarez PH, MacKay A, *et al*. Mucinous carcinoma of the breast is genomically distinct from invasive ductal carcinomas of no special type. The Journal of pathology 2010;222:282-98.
53. Thor AD, Eng C, Devries S, *et al*. Invasive micropapillary carcinoma of the breast is associated with chromosome 8 abnormalities detected by comparative genomic hybridization. Hum Pathol 2002;33:628-31.
54. Marchio C, Iravani M, Natrajan R, *et al*. Genomic and immunophenotypical characterization of pure micropapillary carcinomas of the breast. J Pathol 2008;215:398-410.
55. Guerini-Rocco E, Hodi Z, Piscuoglio S, *et al*. The repertoire of somatic genetic alterations of acinic cell carcinomas of the breast: an exploratory, hypothesis-generating study. J Pathol 2015;237:166-78.
56. Guerini-Rocco E, Piscuoglio S, Ng CK, *et al*. Microglandular adenosis associated with triple-negative breast cancer is a neoplastic lesion of triple-negative phenotype harbouring TP53 somatic mutations. The Journal of pathology 2016;238:677-88.
57. Eberle CA, Piscuoglio S, Rakha EA, *et al*. Infiltrating epitheliosis of the breast: characterization of histological features, immunophenotype and genomic profile. Histopathology 2016;68:1030-9.
58. Vincent-Salomon A, Lucchesi C, Gruel N, *et al*. Integrated genomic and transcriptomic analysis of ductal carcinoma in situ of the breast. Clinical cancer research : an official journal of the American Association for Cancer Research 2008;14:1956-65.
59. Stephens PJ, Tarpey PS, Davies H, *et al*. The landscape of cancer genes and mutational processes in breast cancer. Nature 2012;486:400-4.
60. Filipits M, Rudas M, Jakesz R, *et al*. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. Clin Cancer Res 2011;17:6012-20.

61. Wang Y, Klijn JG, Zhang Y, *et al*. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005;365:671-9.
62. Chang HY, Nuyten DS, Sneddon JB, *et al*. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. Proc Natl Acad Sci U S A 2005;102:3738-43.

Table 1: Summary of the molecular assays commonly used in breast cancer diagnosis prognosis and therapy prediction in clinical practice

Assay	Analyte	Method	Applications
Diagnostic assays			
Diagnostic biomarker panels	Proteins	IHC	IHC biomarkers can be used for diagnosis of histologically challenging breast lesions, differentiate between mammary and non-mammary tumours, and between metaplastic spindle cell carcinomas from other benign spindle cell lesions. IHC can also be used to differentiate ADH/low grade DCIS from hyperplasia, *in situ* from invasive carcinomas and ductal from lobular tumours.
Chromosomal translocation and fusion genes	Tumour DNA	FISH or RT-PCR	Specific translocations are reported to be specific to certain breast lesions such as secretory and mucoepidermoid carcinoma of the breast
Predictive and prognostic assays			
ER and PgR expression	Protein	IHC	Prediction of hormone therapy response, prognostic and diagnostic markers. Also provide some information on response to chemotherapy; ER- tumours respond better than ER+ tumours
HER2 status	Protein	IHC	Protein overexpression and/or gene amplification predict response to antiHER2 targeted therapy. It also provides prognostic information and can be used to help in diagnosis (ie Paget's disease)
	Tumour DNA (*HER*2 gene copy number)	ISH; FISH, CISH, DDISH	
Germ line testing	Non tumour DNA	DNA sequencing	*BRCA*-germ line positivity predicts good response to PARP inhibitors and other synthetic lethality approach and to some extent chemotherapy. Identify genetic predisposition and indication for counsel patients and relatives in addition to guiding screening
Ki67 labelling index	Protein	IHC	It used in some centres to assess proliferation status as a prognostic variable. Also used as a component of some prognostic tools such as PREDICT. May be useful in poorly fixed tumours and in grade 2 cancers
Multigene assays	Tumour RNA	RT-PCR, Expression microarrays and nCounter technology platform	Prognostic in ER+, HER2- breast cancer and mainly used in patients with lymph node negative ER+ HER2- tumours treated with hormone therapy to determine the risk and benefits of using chemotherapy. Currently not recommended in metastatic, locally advanced or advanced stage tumours and have limited prognostic value in ER- and HER2 positive tumours
	Protein	IHC	

ER=oestrogene receptor, PgR= Pprogesterone receptor, IHC= immunohistochemistry, RT-PCR, reverse transcription-polymerase chain reaction, FISH, florescence in situ hybridisation.

Table 2: Summary of prognostic multigene assays that are currently available for early stage invasive breast cancer

Test Name	Component gene(s)	Intended Clinical Utility	Biological Material/ Technology	Classification
MammaPrint/ Amsterdam Signature (42)	70 genes (first prognostic gene signature to be identified)	Node negative ER+ or ER- Estimates the recurrence risk. Test may extend to node positive patients	Fresh or FFPE, Microarray	2 categories: low risk and high risk
Oncotype DX (recurrence score) (44)	21 genes (16 cancer-related and 5 controls)	ER+, HER2-, node-negative BC. Predicts the likelihood of chemotherapy benefit as well as recurrence in hormone therapy treated patients. Test may extend to node-positive patients	FFPE, RT-PCR	3 categories: low risk (RS<18), intermediate risk (RS 18–30), or high risk (RS>31)
EndoPredict (60)	11 genes (8 cancer-related and 3 controls)	Predicts distant and late recurrences in ER+ / HER2- node negative and positive patients treated with endocrine therapy alone	FFPE, RT-PCR	Two categories. Can be combined with tumour size and nodal status to produce clinical score (EPclin)
Prosigna (PAM50) Kit (risk of recurrence assay; ROR) (43)	50 genes (used in the PAM50 molecular classification assay) and 5 control genes	ER+ node negative and positive treated with hormone therapy. Evaluates distant recurrence-free survival at 10 years	FFPE, RT-PCR and by the nCounter Dx analysis system	3 categories: low, intermediate and high risk
Breast Cancer Index (BCI) (17)	A combination of MGI (5 genes) and the 2 gene ratio (H/I) (HOXB13:IL17BR)	Risk of distant recurrence in ER+ node negative BC. Risk of late distant metastasis and benefit from extended (>5 years) endocrine therapy	FFPE, RT-PCR	Two categories
The Rotterdam Signature (61)	76 genes (60 genes for ER+ and 16 genes for ER-patients)	Node negative patients. Predict recurrence in ER+ treated with tamoxifen	Fresh tissue, Microarray	Two categories
BluePrint® Molecular Signature	80-gene profile	Classifies tumours into intrinsic subtypes to suggest the potential effect of adjuvant therapy	FFPE, Microarray	

IHC and *in situ* hybridisation-based assays				
IHC4 score	4 genes (ER, PgR, Ki67 and HER2	ER+ patients treated with hormone therapy	FFPE and IHC	
Insight Dx Mammostrat Plus	9 genes (As above plus ER, PgR, KI67 and HER2)	As above, plus hormone receptor/HER2 status	FFPE, additional genes evaluated by IHC and FISH	2 categories
Assays based on signatures characteristic of a biological process				
Wound-response signature(62)	442 genes	Node negative and positive	Fresh, Microarray	2 categories
Immune signatures	14 genes related to immune function	Predictive for relapse in trastuzumab-treated HER2+ patients		2 categories
Invasiveness Gene Signature (IGS)	186 genes	Predict 10-year distant metastasis free survival in node negative patients	Fresh, Microarray	2 categories

FFPE, Formalin-fixed paraffin-embedded, ER, oestrogen receptor; PgR, progesterone receptor; RT-PCR, reverse transcription-polymerase chain reaction, FISH, florescence in situ hybridisation.