SECTION

1

BIOLOGY

Breast cancer subtypes

1.1

Stefania Morganti, Antonio Marra

Breast cancer is a highly heterogeneous disease that encompasses different biological tumors with a distinct prognosis and response to treatment. Historically, breast tumors have been classified into three major clinical subtypes – hormone receptor (HR) positive, human epidermal growth factor receptor 2 (HER2) positive, and triple negative (TN) – depending on the expression of estrogen receptor (ER), progesterone receptor (PR), and HER2 by cancer cells. These biomarkers not only reflect a different biology of tumor cells across subtypes, but are routinely assessed to define the best treatment strategy. Indeed, each of these subtypes is managed according to specific treatment algorithms that encompass both local management and systemic therapies.[1]

In the early 2000s, the advent of gene expression profiling allowed researchers to better characterize the biology of breast cancer. They identified five distinct "intrinsic" molecular subtypes (luminal A, luminal B, HER-2 enriched, basal-like and claudin-low) and a normal breast-like group.[2, 3] These subtypes showed a high correlation with both prognosis and treatment response, but only a partial overlap with clinical surrogate subtypes. Nevertheless, no studies have proved the clinical utility of molecular subtyping in a prospective randomized trial, so the approximation between intrinsic and surrogate subtypes is still valid for clinical purposes.[4]

Two assays are currently commercially available for molecular subtyping – the Prosigna and BluePrint assays – but neither of them is commonly used for clinical decision-making. Of note, most of the available evidence about breast cancer subtypes reported in this chapter has been derived from studies of gene expression profiling conducted in the early-stage setting.

• Intrinsic subtyping by gene expression profiling: a historical perspective

The biological behavior and prognosis of breast cancer can only partially be explained by traditional clinicopathologic factors, such as tumor stage, histologic subtypes, HR and HER2 status, tumor grading, or proliferation biomarkers. The limitations of these variables, in parallel with the progress in gene expression profiling techniques, have set the basis to invest in correlative studies to investigate the heterogeneity of breast cancer at gene expression level.

In 2000, Perou *et al.*[2] first uncovered the heterogeneity of breast cancer by identifying "molecular portraits" with different patterns of gene expression. They assumed that the phenotypic diversity of breast tumors might be related to different gene expression patterns that could be captured by complementary DNA (cDNA) microarrays. Indeed, the authors analyzed 496 genes in 84 normal or tumor samples from 42 individuals by cDNA microarrays and hierarchical clustering, and identified four different subtypes based on different gene expression profiles: basal-like, Erb-B2+, luminal epithelial/ER-positive, and normal-breast-like. They defined this list of genes by using a hierarchical clustering method to group genes based on the similarity of expression pattern across samples so that the selected genes were the ones showing high variance across tumors, but low difference among repeated samplings of the same tumor. Notably, these "molecular portraits" not only described similarities and differences among the tumors, but in many cases mirrored different biological pathways. The proliferation cluster, including genes whose expression level correlated with the cellular proliferation rate, was differentially expressed across samples and correlated with the mitotic in-

dex. Similarly, expression of the ESR1 gene (encoding ERα) and other genes expressed by breast luminal cells correlated with the ER status. Tumors with high expression of the "basal like" gene cluster showed instead staining for cytokeratin 5/6, commonly used to identify basal epithelial cells, and failed to express ER and most of the other genes usually co-expressed with it. Overexpression of the Erb-B2 oncogene was also associated with high expression of a specific subset of genes and low expression of genes associated with ER expression. Finally, the normal-breast-like subgroup showed the highest expression of genes expressed by adipose tissue and other nonepithelial cell types.[2] Of note, at that time the predictive and prognostic role of HER2 was largely unknown,[5] so the authors did provide the biological rationale to distinguish between two different subgroups of ER-negative tumors.

A year later, the same research group proved that intrinsic subtypes also correlate with relapse-free survival and overall survival (OS).[3] They profiled 78 tumor samples by using the intrinsic gene set of 456 cDNA clones by hierarchical clustering. Of note, the larger sample size allowed the researchers to identify three subgroups in the previously defined luminal/ER-positive subtype, namely luminal A, luminal B, and luminal C. The luminal A subtype demonstrated the highest expression of the ESR1 gene and related genes included in the ER cluster, whereas both the luminal B and C subtypes showed low-to-moderate expression of these genes. The luminal C subtype was further distinguished by the high expression of a novel cluster of genes with unknown coordinated function, although given the similarity with the luminal B subtype, these two subgroups were subsequently analyzed together.

Survival analyses conducted on a subgroup of 49 patients uniformly treated in a prospective study showed significantly different outcomes for the distinct subgroups, with the basal-like subtype having the worst prognosis. Of note, the two ER-positive groups (luminal A vs. luminal B+C) also had different outcomes, suggesting that the luminal B+C might represent a clinically distinct group with a worse prognosis.[3]

Subsequent studies have refined the intrinsic gene list,[6, 7] and have validated its prognostic significance on a large independent test set.[7] Notably, the intrinsic subtypes have been shown to increase outcome prediction when added to a multivariate model including standard clinical variables (ER status, node status, grade, and tumor size).[7]

PAM50 and the Prosigna assay

Because of the potential clinical value of molecular subtyping, the next step was to develop a clinical assay that could be applied in clinical practice. Thus, Parker *et al.*[8] developed the PAM50 assay to predict intrinsic subtypes from clinical tumor samples. In this validation study, the authors first developed the 50-gene classifier to identify subtypes, and then tested the prognostic power of PAM50 through a risk-of-relapse (ROR) score that considered subtype alone (ROR-S) or with the tumor size (ROR-C). They found a clear improvement in both prediction of relapse and response to neoadjuvant chemotherapy when adding ROR-S to clinical variables. The authors also compared clinical subtypes assessed by ER and HER2 status with the intrinsic subtype and observed only a partial overlap, suggesting that the latter cannot be inferred by receptor status only.[8]

It is noteworthy that PAM50 is based on real-time quantitative reverse transcription polymerase chain reaction (qRT-PCR) and assessed on formalin-fixed paraffin-embedded (FFPE) tissues, whereas previous studies had used microarray assays run on fresh frozen tissues. Because of its easy applicability to archival samples, PAM50 became the assay of reference to investigate tumor subtypes in the context of correlative studies. Hence, this 50-gene classifier was run on archival samples from many different trials, proving its prognostic power across clinical subtypes and treatment regimens. Indeed, PAM50 proved not to be prognostic and predictive of the benefit of chemotherapy and anti-HER2 and endocrine therapy (ET) across retrospective studies in different settings.[9]

To date, PAM50 is the most frequently used assay for transcriptomic subtyping in correlative studies. Nevertheless, its algorithm needs to be change with "platform- and cohort-adjusted" adjusted for each new dataset to avoid bias in subtype classification. Indeed, its applicability in clinical practice for subtyping single samples is limited. The Prosigna assay is an alternative PAM50-based subtype classifier specifically developed to standardize tumor subtyping and allow for decentralized testing in clinical laboratories.[9] The Prosigna assay is based on the NanoString nCounter Dx Analysis System, which provides more accurate measures of messenger RNA (mRNA) expression levels in FFPE tissue when compared with PCR. Additionally, its algorithm does not require additional cohort or platform normalization, providing a standard algorithm that can be applied to single patient samples or biased populations with both accurate and stable results. The Prosigna assay has been trained, tested, and validated to accurately identify intrinsic subtypes (94% concordance with PCR-based PAM50) and predicts risk of recurrence (ROR) in HR-positive breast cancer.[9]

The 80-gene BluePrint assay

BluePrint is an 80-gene molecular subtyping profile assay that classifies breast tumors into three molecular subtypes: luminal-type, HER2-type, and basal-type.[10] Differently than PAM50, the BluePrint assay is not based on hierarchical clustering. Rather, it was developed by using concordance between ER, PR, and HER2 status assessed by immunohistochemistry (IHC) and mRNA expression using the microarray-based TargetPrint assay. A cohort of 200 samples with concordant IHC status was used as a training cohort to identify the genes that best discriminated the three subtypes with a 3-fold cross validation procedure. Eighty genes provided the best performance and were selected for further validation on four independent cohorts.[10] Of note, this assay does not discriminate between luminal A and B subtypes, but luminal subtyping has been done using the 70-gene MammaPrint assay to stratify between low-risk (equivalent to luminal A) and high-risk (equivalent to luminal B) breast cancer.[10] Nevertheless, this subtyping did not significantly increase prognostic stratification.[10] Initially developed on frozen tissue, BluePrint is now available for FFPE tissue analyses, and there is good concordance between the two tissue preparations.[11] In addition, molecular subtyping through BluePrint predicts the benefit of neoadjuvant therapy and long term-outcomes in early-stage breast cancer.[12]

There are only nine common genes between PAM50 and BluePrint, although an overall concordance of 92% was observed after removing normal-like samples. In the OPTIMA study, comparing Prosigna to BluePrint in 295 postmenopausal patients with ER-positive/HER2-negative early breast cancer, a moderate concordance between the two assays was shown. A few studies have applied BluePrint in the neoadjuvant or adjuvant setting, showing a better prediction of pathologic complete response (pCR) and outcomes than clinical subtypes alone, although the evidence is not as strong as what has been proved for PAM50, and virtually absent in the metastatic setting.[13]

• Molecular features of intrinsic subtypes

In 2012, the molecular characterization of early breast cancer done by The Cancer Genome Atlas (TCGA) network further corroborated the importance of intrinsic subtyping in breast cancer.[14] This project profiled 510 primary breast cancer using six platforms: 1) gene expression DNA microarrays; 2) DNA methylation arrays; 3) micro-RNA (miRNA) sequencing; 4) Affymetrix single nucleotide polymorphism (SNP) arrays for DNA copy number analysis, 5) exome sequencing, and 6) reverse-phase protein arrays (RPPA). It is noteworthy that the combination of data from five different genomic/proteomic platforms (*i.e.*, all except for exome sequencing) showed that breast cancer can be grouped into four major subtypes, and that these "consensus clusters" correlate significantly with PAM50 mRNA subtypes.[14]

Also in 2012, Curtis *et al.*[15] published the result of the METABRIC (Molecular Taxon-

omy of Breast Cancer International Consortium) project, a similar effort that applied an integrated genomic/transcriptomic analysis to identify breast cancer subtypes with different biology and outcomes. The authors used discovery and validation sets of 997 and 995 primary tumors, respectively, and identified 10 novel molecular subtypes called "IntClust." Interestingly, differences across these subtypes were mainly driven by CNA aberrations, which accounted for the greatest variability in gene expression. These clusters further dissected the heterogeneity of the previously identified intrinsic subtypes, identifying subgroups with distinct biology and outcomes.[15, 16]

Luminal subtype

Luminal breast cancer is the most heterogeneous subtype in terms of both genomic alterations (*i.e.*, mutational profile and copy number variations) and gene expression.[17] As outlined above, two major types of luminal cancer can be identified: luminal A and luminal B. At the gene and protein expression levels, these subtypes are distinguished by differential expression of the luminal and proliferation signatures. Luminal A tumors have higher expression of the luminal expression signature, which includes ESR1, GATA3, FOXA1, XBP1, and MYB. Luminal B tumors have a similar expression of the ER gene, lower expression of other luminal-related genes (*i.e.*, PGR and FOXA1), but higher expression of proliferation/cell cycle-related genes or proteins (e.g., MKI67 and AURKA).[14, 18, 19] In terms of the mutational profile, PIK3CA/PTEN, TP53, and RB1 represent the most frequently altered pathways in luminal tumors. PIK3CA mutations are more frequent in the luminal A than luminal B subtype (49% vs. 32%), whereas PTEN mutation/loss are more common in luminal B tumors (24% vs. 13%). Interestingly, the TCGA RPPA analysis showed a discrepancy between PIK3CA mutations and biomarkers of pathway activation in luminal breast cancer, with the absence of elevation of pAKT, pS6, and p4EBP1 despite the presence of PIK3CA mutations.[14] On the contrary, other studies have shown a correlation between PIK3CA mutations and activation of the phosphoinositide 3-kinase (PI3K) pathway.[20]

Regarding the p53 pathway, both TP53 mutations and MDM2 amplifications are more frequent in luminal B (29%) than luminal A (12%) tumors, and gene expression analysis has confirmed that the p53 pathway is generally conserved in luminal A tumors but often altered in luminal B tumors. Similarly, the RB pathway is more frequently inactivated in luminal B than luminal A tumors, with more cyclin D1 (encoded by CCND1) amplifications (58% vs. 29%), CDK4 gains (25% vs. 14%), and higher activity signatures.[14]

Human epidermal growth factor 2-enriched subtype

The HER2-enriched (HER2-E) subtype is defined by the expression of the HER2-amplicon-associated genes (e.g., ERBB2/HER2 and GRB7), along with a high proliferation signature but a low basal signature. At the DNA level, it is characterized by elevated genomic instability with a high number of mutations across the genome, with frequent TP53 (75%) and PIK3CA (42%) mutations. Additionally, the HER2-E subtype is typically enriched for the APOBEC mutation signature, which is a known mechanism of mutagenesis in many tumors. At the protein level, HER2-E tumors present high protein and phosphoprotein expression of EGFR and HER2 (the EGFR/pEGFR/HER2/pHER2 signature).[14]

Although the HER2-E subtype is the most prevalent among clinical HER2-positive tumors, a different gene expression profile can be detected in up to 50% of HER2-positive breast cancer, with luminal subtypes being the most frequent. These HER2-positive/luminal tumors have higher expression of genes belonging to the luminal cluster, such as GATA3, BCL2, and ESR1, and a different mutational profile, with generally fewer TP53 and more GATA3 mutations.[14, 21]

Basal-like subtype

The basal-like signature is characterized by the expression of cytokeratin 5, 6, and 17 and high proliferation-related genes (e.g., MKI67),

with very low expression of luminal-related genes.[14] These tumors have the highest frequency of TP53 mutations (80%) and high p53 pathway activity, suggesting that this pathway is likely to be altered in all basal-like tumors. RB1 mutations/losses (20%) and BRCA1/BRCA2 mutations (20%) are also typical of basal-like tumors. PI3KCA mutations are instead less common (7%), although other genes involved in the PI3K pathway can be altered (PTEN mutation/loss 35%, INPP4B loss 30%).

At the protein level, basal-like tumors are characterized by high expression of DNA repair proteins, and the PTEN and INPP4B loss signature (pAKT). Interestingly, basal-like tumors were found to share many genomic alterations with serous ovarian carcinomas (BRCA1 inactivation, high AKT3 expression, MYC overexpression, RB1 loss and CCNE1 amplification, and TP53 mutations), suggesting a similar process of carcinogenesis.[14] In terms of pathologic features, most basal-like are typically negative for ER, PR, and HER2. Nevertheless, only 75-90% of TN breast cancer tumors are basal-like.[22]

Claudin-low

In 2007, an additional subtype defined claudin-low was described in both human and murine breast cancer, characterized by the low gene expression of claudin 3, 4, and 7 (tight junction proteins) and E-cadherin (a calcium-dependent cell-cell adhesion glycoprotein).[23] Moreover, this subtype was found to be exclusively enriched for the tumor initiating cell (TIC) genomic signature derived from CD44-positive/CD24-negative/low-sorted cells and mammospheres. Clinically, this subtype typically presents as a TN invasive ductal carcinomas with a high frequency of metaplastic and medullary differentiation and is mainly characterized by a poor prognosis.[24]

• Clinical subtypes

In clinical practice, breast cancer classification and subtyping rely on the expression of ER, PR, and HER2.[1, 25] Three major clinical subtypes have been identified for both stratification of prognosis and treatment purposes, namely the HR-positive/HER2-negative, HER2-positive, and TN subtypes. Many studies have shown that there is only a partial overlap between clinical and intrinsic subtypes, with a significantly higher accuracy and reproducibility of gene expression profiling assays.[26] Nevertheless, to date there are no prospective studies proving the clinical utility of tumor intrinsic subtyping, and the approximation between intrinsic and surrogate subtypes is still valid for clinical purposes. In 2011, the St. Gallen Consensus Panel first adopted the definition of "surrogate" subtypes for recommending adjuvant systemic therapy in the early setting, distinguishing between luminal A-like, luminal B-like, HER2-E, and TN.[4]

Hormone receptor-positive breast cancer

According to the latest American Society of Clinical Oncology (ASCO)/College of American Pathologists (CAP) guidelines, breast tumors with >1% of tumor nuclei positive for ER or PR are classified as HR positive.[27] This definition includes most breast cancer (70-80%) and groups a wide range of different tumors, not only in terms of ER/PR staining intensity, but also for histologic grade, proliferation markers (Ki67/MIB-1), genomic alterations, and gene expression profiles.[17] Differences in these pathological and genomic features translate into different clinical behaviors, response to therapies, and prognosis.

In terms of tumor histology, around 85% of HR-positive tumors are invasive ductal carcinomas. Nevertheless, ductal carcinomas are also the most common histology for TN and HER2-positive tumors. Lobular carcinomas account for ~15% of HR-positive tumors, whereas very rare histologies that are typically HR-positive are cribriform and tubular carcinomas.

Around 10% of HR-positive breast cancer is diagnosed in patients harboring a germline mutation in cancer susceptibility genes, such as BRCA1 (%), BRCA2 (2%), CHEK2 (1%), ATM (0.5-1%) and PALB2 (0.5-1%). All of them are more prevalent in HR-positive than in HR-negative breast cancer, except for BRCA1 that is instead more frequently mutated in TN tumors.[17]

In terms of gene expression profiles, most

HR-positive tumors belong to the luminal subtype, so that HR-positive tumors are usually referred to as luminal-like. Biological differences between luminal A and B usually reflect distinct pathological features, with luminal A-like tumors having high ER and PR expression, a low tumor grade (*i.e.*, well differentiated), and low Ki67/MIB-1 expression. On the contrary, luminal B-like tumors are more commonly high grade, have lower ER/PR expression, and higher Ki67. Additionally, some luminal B-like tumors show HER2 overexpression based on IHC or fluorescence in situ hybridization (FISH) analysis and are referred as triple positive.[14, 18, 19]

The distinction between luminal A-like and luminal B-like is especially relevant in the early-stage setting, because it was adopted after the 2011 Saint Gallen Consensus Conference to identify what patients with HR-positive breast cancer would have benefit from adjuvant chemotherapy.[4] The key discrimination factor was Ki67, with tumors expressing Ki67 higher than 14% classified as luminal B and tumors with lower Ki67 expression classified as luminal A. The importance of Ki67 has been reduced progressively since then, mainly because of its low reproducibility and the availability of genomic signatures. In the metastatic setting, the difference between luminal A and B is not relevant for clinical decision-making.

The large majority of HR-positive breast cancer has a luminal expression profile, although a non-luminal subtype can be identified in around 8% of early and 15% of metastatic HR-positive breast tumors.[28] In a combined dataset including 1,670 patients with HR-positive metastatic breast cancer, 2.3% of samples were classified as basal-like and 13.6% as HER2-E.[28] Regarding the IHC profile, non-luminal/HR-positive tumors have lower ER/PR expression and higher Ki67, albeit none of these biomarkers allow discriminating between luminal and non-luminal HR-positive breast tumors. In the early setting, tumors with non-luminal/HR-positive breast cancer have been shown to be less sensitive to ET and have worse outcomes.[28]

Many correlative studies of large, randomized trials have investigated the prognostic and predictive value of molecular subtyping in patients with HR-positive breast cancer, although only a few of them were conducted in the metastatic setting. PAM50 profiling of patients included in the PALOMA-2 and PALOMA-3 studies, investigating palbociclib plus ET versus ET alone in first- and second-line line therapy for HR-positive metastatic breast cancer, classified ~44-50% of tumors as luminal A, ~30% as luminal B, ~20% as HER2-E, ~1-3% as normal-like, and ~1-2% as basal-like tumors.[29, 30] In both studies, patients with luminal A and B breast cancer derived a significant benefit from the addition of palbociclib. In the PALOMA-3 trial, non-luminal tumors also derived a benefit from the combination therapy, whereas PALOMA-2 patients with HER2-E tumors had similar median progression-free survival (PFS) in both arms. Nevertheless, the small number of patients limits the meaningfulness of this analysis. In 2021, a pooled analysis of the MONALEESA trials, testing the addition of ribociclib to ET versus ET alone in slightly different metastatic settings, investigated the prognostic role of PAM50 on 1160 tumor samples.[31] The tumors were classified as follows: 46.7% luminal A, 24% luminal B, 14% normal-like, 12.7% HER2-E, and 2.6% basal-like. In both treatment arms, the intrinsic subtype was independently associated with PFS ($P<0.001$) after adjusting for clinicopathologic variables. All subtypes derived benefit from the addition of ribociclib except for basal-like, that also had a 3.96 times higher risk of disease progression than luminal A. Interestingly, the HER2-E subtype showed the worst prognosis with ET alone, but the greatest relative benefit from the combination therapy. The EGF3008[32] and the BOLERO-2[33] studies also assessed the PAM50 subtypes in patients with HR-positive metastatic breast cancer receiving letrozole ± lapatinib and exemestane ± everolimus, respectively. In both cases, non-luminal subtypes were associated with worse PFS and OS.

Of note, most of these studies assessed the PAM50 subtype based on primary samples. Paired analysis of primary and metastatic samples showed a shift in around 40% of luminal A toward more aggressive subtypes,