



Molecular Classification of Breast Cancer

4

Jose Russo

4.1 Introduction

The rationale for a genomic classification is basically two: one is that knowing the gene expression profile of a tumor leads us to understand cancer behavior and second at the clinical level could help to identify genes that are associated with specific therapeutic target. The concept is that knowing the genomic signature of a given breast cancer it could provide the molecular pathway that help better decision in using available therapy that has been shown to be efficient from other primary tumors.

The first molecular classification came from the study performed by Perou and Sorlie [1]. In this study, they acquired samples of breast tissue from 42 individuals. Forty of these were mostly invasive ductal carcinomas, one was a fibroadenoma, and the last sample a normal breast tissue. In addition, contained in this study were 22 pairs of tumor samples from which 20 were paired before and after a chemotherapy regimen and two pairs from primary tumors paired with their respective lymph node metastasis. Using these samples, Perou and Sorlie [1] isolated the RNA and performed cDNA microarray. From these microarrays 8102 genes were initially identified, and a subset of these genes was selected

based on the variation in their expression using at least plus-or-minus four-times the median level of expression. Using these criteria, they finally selected 1753 genes for hierarchical cluster for their final classification. Perou and Sorlie [1] selected genes based on gene expressions that were similar in any sample taken from the sample tumor but varied between different tumors, and for that purpose they utilized the 22 paired tumor samples and identified 496 genes from the 1753 identified earlier based on the variation in their gene expression, with the added distinction of having greater variation between different tumors than from the same tumor. They called this new cluster the “Intrinsic Gene Subset.” From this new subset, Perou and Sorlie [1] were not only able to determine the expression levels for each sample but were also able to group them based off their expression within the two layers found in mammary gland structures (the lobules and ducts): the inner luminal epithelial cells and the surrounding basal myoepithelial cells. From these groupings, Perou and Sorlie [1] were able to further distinguish each group into gene clusters and identified one cluster for the luminal epithelial cells, the *luminal epithelial/estrogen cluster*. In this work [1], Perou and Sorlie [1] also identified three clusters for the basal epithelial cells: the ERBB2 overexpression cluster, the basal epithelial cell-associated cluster, and the cluster containing keratins 5 and 17 and the basal epithelial-cell-enriched gene [2]. The clusters identified were later refined into subtypes [3].

J. Russo (✉)
The Irma H Russo, MD-Breast Cancer Research
Laboratory, Fox Chase Cancer Center-Temple Health,
Philadelphia, PA, USA
e-mail: jose.russo@fccc.edu

The luminal cluster was separated into luminal subtype A, luminal subtype B, and luminal subtype C. The basal clusters were redefined: the ERBB2 overexpression cluster became simply the ERBB2+ subtype, the basal epithelial cell-associated cluster became the basal-like subtype, and the basal epithelial-cell-enriched gene cluster became the normal breast-like subtype.

From these data, the authors [1, 3] identified five subtypes and made the correlation with the clinical features indicated by each subtype. To accomplish this, they utilized data acquired from 49 breast cancer patients showing all five subtypes, who only had diseases local to the breast and with little-to-no metastasis present [3]. They specifically looked at the overall survival (survival months) and relapse-free survival (RFS) probabilities for each subtype over a 4-year period, in comparison to the other subtypes. In addition, they also looked at the outcomes when luminal subtype C and B were grouped with the other subtypes.

The final analysis between the correlation of molecular subtypes and clinical behavior indicated that the basal-like and ERBB2+ subtypes both had the lowest RFS and overall survival. Additionally, the luminal subtype C was shown to have the worst overall survival of all the luminal subtypes, and subtypes ERBB2+ and luminal B were shown to share certain genes associated with a poor prognosis.

4.2 Molecular Subtypes of Breast Cancer

Since the work of Perou and Sorlie [1, 3] defining of the Intrinsic subtypes, many studies have gone on to further refine and expand upon the initial

breast cancer classifications and redefining them as the *molecular subtypes of breast cancer*.

The modern, accepted subtypes are the luminal A, luminal B, basal-like, ERBB2+/HER2+, and the normal breast-like subtypes. There are several accepted means for distinguishing between the five subtypes. The primary method is the presence of absence of three different cellular receptors in the breast cancer tumors. The three receptors are the estrogen receptor (ER), the progesterone receptor (PR), and the human growth factor receptor 2 (HER2). Overexpression of these receptors has been observed in breast cancers, but has often only been looked at individually. In addition, different breast cancer tumors have been shown to have different expression levels of these receptors. Thus, by looking at all the receptors together and identifying which are overexpressed (or absent) in the tumor cells, there can be a clear classification used to distinguish between breast cancers.

The protein Ki-67, a known prognostic factor associated with proliferation, is utilized in a similar manner, looking at the low or high levels for subtype distinction. The grade of the tumor is another factor incorporated into identifying molecular subtypes, which looks at how the tumors appear in comparison to normal, well-differentiated breast tissue. High grades are described as “poor,” or not well-differentiated, while low grades are described as “good,” or well-differentiated (Table 4.1) [4, 5].

Additional classifications have been described to further distinguish between these accepted subtypes. This includes the further subtypes for luminal B: the HER2+ and HER- subtypes. The distinguishing factor between them is that Ki-67 levels are generally high in HER2+ luminal B

Table 4.1 Summary of the standard features for each of the five molecular subtypes [4, 5]

Molecular subtype	Estrogen (ER)	Progesterone (PR)	HER2	Ki-67	Tumor grade
Luminal A	+and/or	+	–	Low	Low
Luminal B	+and/or	+	+/-	High (HER2+)	High
ERBB2+/ HER2+	–	–	+	High	High
Basal-like (“triple negative”)	–	–	–	High	High
Normal breast-like	Normal	Normal	Normal	Normal	Low

breast cancers [4]. Another distinction is made between triple-negative breast cancers (TNBC) and basal-like breast cancers. Though TNBC (named for being negative for all three receptors) have traditionally been grouped with the basal-like subtype, they are not synonymous, and there is at most an 80% overlap between the two [5]. TNBCs have thus been separated into two further subtypes: basal-like and non-basal-like. The major distinction is the expression of cytokeratins 5 and 6 (CK5/6), as well as epidermal growth factor receptor (EGFR), for the basal-like subtype [6].

Other studies have described additional molecular subtypes, distinct from the accepted five. One proposed subtype is the claudin-low subtype. They are characterized with low expression of the claudin proteins (found within cellular junctions) and are associated with mammary stem cells [7]. Although similar to the basal-like subtype in being triple negative, the claudin-low subtype clinically shows to have a better prognosis than the basal-like subtype.

Based on immunocytochemical classification, the luminal A-subtype breast cancers suggest that these patients have a better prognosis compared with those with breast cancers of other subtypes. In a publication of Gao and Swain [8] raised the question that in these patients, chemotherapy could be omitted, and endocrine therapy alone could be sufficient based on the fact that the luminal A-subtype tumors are a unique subset that may have favorable tumor biology [8].

Despite the usefulness of the molecular subtype classifications of breast cancer, there are several limitations. One of the major limitations is the apparent lack of understanding of the variation in response to therapies specific to the subtype. Such variation has limited value in a clinical setting, where proper treatment is crucial to patient survival. Because of these limitations, several cancer research groups have sought out newer, more reliable studies for means of classifying breast cancers. Among them and one that is important to be discussed here are studies done by the Molecular Taxonomic Breast Cancer International Consortium (METABRIC) [9]. The methodology used by this consortium is the state

of the art by using sequencing technologies that identified the mutational patterns and genomic instabilities characteristic of different breast cancers. More importantly this new classification also integrates the classical classifications of breast cancer, describing features such as receptors and tumor grade, as well as direct comparisons with the molecular classifications. In the study reported by Dawson et al. [9], about 2000 breast tumors were analyzed, to acquire both their genomic and transcriptomic sequences, identifying where gene alterations had occurred. This included inherited variation to the genome, specifically single-nucleotide polymorphisms (SNPs) and copy number variants (CNVs), but also looked at acquired variation via single nucleotide variants (SNVs, aka mutations) and copy number aberrations (CNAs) [9].

In this chapter, I am summarizing their findings, and the reader is strongly encouraged to study this classification. From this seminal publication [9], ten novel subtypes of breast cancer were identified. The authors [9] designated Integrative Clusters (IntClust 1–10). Each cluster was primarily distinguished by the CNAs, which were identified to have the greatest variation, but also found to have overall differing gene expressions. The extent to which the clusters associated with the accepted intrinsic subtypes was analyzed for each cluster, as was the expression of the accepted prognostic receptors of estrogen, progesterone, and HER2. Further analysis identified the clinical implications for each cluster, such as the genomic instabilities, and distinguishing somatic mutations, but also more specific characteristics for each cluster including age of diagnosis and survivability probabilities.

4.3 Genomic Classification Based on the Normal Cell Subtype

Despite the benefits of the newer genomic classifications of breast cancer, alternative means of classifications still arise to confront new or unaddressed issues. Such issues were addressed in a study by Santagata et al. [10]. In their study, they

set out to produce a normal cell subtype-based classification system, where they focused on utilizing normal cell types found in normal breast tissue as references for breast cancer classifications. This method, they argue, has successfully been used before to characterize hematopoietic tumors (lymphomas, leukemias, etc.) by other research groups [11] but have rarely been emulated due to a poor understanding of cell-type diversity among tissues. They argue that their new classification, unlike previously produced ones, forms actual disease taxonomy for breast cancers. That is, previous classification systems have heavily relied on differing clinical results (based on different molecular platforms for analysis) to form categories based only on overall prognosis. These categories also vary greatly, and no new classification system is truly agreed upon in clinical settings, seeing them as unreliable for patient prognosis and treatments. Their new classification system aimed to provide such clinical reliability. This classification identified that breast cancers, being heterogeneous, can vary depending on their cellular origin: in the luminal epithelial layers or the myoepithelial layers [10]. Thus, they analyzed about 15,000 normal breast cells for cellular markers distinguishing between the two layers. They focused on identifying bimodal expression markers (which produced a clear negative/positive distinction) and utilizing these markers to distinguish between varying differentiation states of the cell populations [10]. Three of the major markers identified were hormone receptors of the luminal epithelia: the vitamin D receptor (VDR), the androgen receptor (AR), and the estrogen receptor (ER). Additional markers included different keratins, claudins, cluster of differentiation (CD) markers, and even Ki-67. Identifying the different expressions of these markers in the different cell populations allowed for the formation of eleven luminal layer subtypes (L1–11) and two myoepithelial layer subtypes (My1 and My2). Following these classified layers, the study focused on actually classifying human breast tumors based on normal cell types. Four unique subtypes, called “hormone receptor subtypes,” were identified: HR0, HR1, HR2, and HR3. Each subtype is based on the

expression of the three major hormone receptors (VDR, AR, and ER) and how many were expressed (0–3). The previously characterized luminal subtypes were then distinguished based on these novel subtypes. Next, the study looked to identify if breast tumors maintain the same expression patterns characteristic of the normal cell type, specifically the differentiation-state-specific patterns, which involved identifying the gene expression patterns among the luminal and basal markers (including the three major markers), as well as the specific marker of K5/K14 (found to be a reliable distinguisher between luminal layers). The expression of these markers was identified in ER+, HER2+, and triple-negative breast cancer (TNBC) tumors and compared with the expressions found in normal breast tissues with the same distinguished expressions. An example of this comparison can best be seen for the ER+ tumors, where they identified the ER+ tumors to co-express VDR in 93% of the tumors and AR in 59% of the tumors, and the K5/K14 were found to be negative in these tumors. When compared to the counterpart normal cells, there was found to be a near identical expression pattern: they both co-express VDR and AR to the same levels, and both rarely expressed K5 or K14.

Such identical expressions were seen in the HER2+ and TNBC tumors and their counterparts, verifying their normal cell subtype classification method [10]. Finally, the study aimed to identify the clinical significance behind the new hormone receptor subtypes. This involved acquiring tumor data from patients in a separately performed study by the Nurse’ Health Study, which had previously been classified as ER+, HER2+, and TNBC based on the presence of the classical receptors (ER, PR, and HER2) within their tumors. These classically assigned tumors were compared with the new HR subtypes, which showed that the HR subtypes provided more distinguished groups of tumors than the previous classification. In addition, the HR subtypes were clinically identified from each other by overall survival and relapse-free survival, which identified that the HR0 subtype had the worst prognosis and the HR3 subtype had the best prognosis [10].

4.4 Tests for Molecular Profiling of Breast Cancer

After the publication of the new genomic classification of breast cancer [12], several adaptation compiling few representative genes have been introduced in the practice of oncology. Among them are the Oncotype Dx that contain 21 gene expression signatures, and the first major trial was published in 2004 [13]. This study comprises women that are ER+ and LN negative breast cancer. This test was recommended by NCCN and ASCO. The Oncotype was developed from a FFPE compliant assay to predict distant recurrence of ER+ breast cancer and originally selected 250 candidate genes to test on NSABP B-14 and B-20 trials. At the end, they refined a 16 + 5 gene panel that could predict recurrence [13].

The MammaPrint contains 70 gene expression signatures, and the major trial was from 2002 comprising women <61 years, T1–T2, N0 disease [14]. The Prosigna from NanoString contains 50 gene expression signatures +5 control genes and represents the old PAM50 assay. The PAM or prediction analysis of microarrays recapitulated the microarray classifier using RT-PCR-based PAM50 assay in comparison to standard clinical molecular markers. The major trial was in 2013 using stage I–III cancer population and cleared by the FDA in 2013. Thirteen years after Perou’s paper [12], the PAM50 is entering the clinical arena. The PAM50 gene signature has been transferred to a novel and

robust method for mRNA quantification [15]. The method works well in FFPE, does not rely on amplification of nucleic acids, and is intended for kit use in local labs with the proper instruments. The PAM50 expression results are used to calculate a risk of recurrence score (ROR) and provide low-, intermediate-, and high-risk groups. The score is based on the intrinsic subtype and pathologic characteristics (T, N), with special weighting given to a set of proliferation associated genes. PAM50 correlates well the Oncotype and the use of the 4 IIC parameters (ER, PR, Her2, and ki67). Table 4.2 summarizes the major features of these three genomic tests in breast cancer.

4.5 Conclusions

The advances in our understanding of the role of genomic changes in breast cancer have no doubt been spectacular in the last two decades from the first molecular classification described by Perou and Sorlie [1]. The molecular classification of breast cancer is a continuous quest that has created meaningful studies improving and giving a better understanding of the complexity of breast cancer. I have provided a brief summary that opens our understanding on the direction in which we are going and provides the technological response of diagnostic markers that help the oncologist to tie the complexity of the molecular pathway/s with diagnosis, prognosis, and therapeutic targeting.

Table 4.2 Major features of three genomic tests in breast cancer

	MammaPrint	Oncotype Dx	Prosigna
Input material	Fresh frozen FFPE	FFPE	FFPE
Platform	Microarray	qPCR	nCounter
#genes analyzed	70	21	50 + 5
Target patient population	Stage I–II	ER+ stage I–II	Stage I–II (stage I–III)
Regulatory	FDA cleared (frozen)	NCCN/ASCO Guidelines	FDA cleared
Performance site	Central	Central	Decentralized kit format
Features	Binary stratification; molecular subtypes; lots of data	Gold standard; ER+ cancer; now in DCIS	ROR compares well; innovative technology; <i>intrinsic subtypes</i>

Tips and Tricks

1. The rationale for a genomic classification are basically two: one is that knowing the gene expression profile of a tumor leads us to understand cancer behavior and second at the clinical level could help to identify genes that are associated with specific therapeutic target.
2. The modern, accepted molecular subtypes of breast cancer are the luminal A, luminal B, basal-like, ERBB2+/HER2+, and the normal breast-like subtypes. The luminal A-subtype breast cancers suggest that these patients have a better prognosis compared with those with breast cancers of other subtypes.
3. The molecular classification of breast cancer is a continuous quest that have created meaningful studies improving and giving a better understanding of the complexity of breast cancer.

References

1. Perou CM, Sorlie T, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747–52.
2. Fu J, Wu L, Fu W, Tan Y, Xu T, Hong Z, Wang F, Li S. How young is too young in breast cancer?—Young breast cancer is not a unique biological subtype. *Clin Breast Cancer*. 2018;18(1):e25–39. <https://doi.org/10.1016/j.clbc.2017.05.015>. ePub 2017 Jun 8
3. Sorlie T, Perou M, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001;98(5):10869–74.
4. Nishimura R, et al. Ki-67 as a prognostic marker according to breast cancer subtype and a predictor of recurrence time in primary breast cancer. *Exp Ther Med*. 2010;1(5):747–54.
5. Goldhirsch A, et al. Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol*. 2013;22(8):1736–47.
6. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, et al. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin Cancer Res*. 2004;10(16):5367–74.
7. Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. *Mol Oncol*. 2011;5(1):5–23.
8. Gao JJ, Swain SM. **Luminal A Breast Cancer and Molecular Assays: A Review**. (2018) *Oncologist*. Feb 22. pii: theoncologist.2017–0535. <https://doi.org/10.1634/theoncologist.2017–0535>. [Epub ahead of print].
9. Dawson SJ, Rueda OM, et al. A new genome-driven integrated classification of breast cancer and its implications. *EMBO J*. 2013;32(5):617–28.
10. Santagata S, Thakkar A, et al. Taxonomy of breast cancer based on normal cell phenotype predicts outcome. *J Clin Invest*. 2014;124(2):859–70.
11. Swerdlow SH. Lymphoma classification and the tools of our trade: an introduction to the 2012 USCAP Long Course. *Mod Pathol*. 2013;26(Suppl 1):S1–S14.
12. Tang G, Shak S, Paik S, Anderson SJ, Costantino JP, Geyer CE Jr, Mamounas EP, Wickerham DL, Wolmark N. Comparison of the prognostic and predictive utilities of the 21-gene recurrence score assay and adjuvant! For women with node-negative, ER-positive breast cancer: results from NSABP B-14 and NSABP B-20. *Breast Cancer Res Treat*. 2011;127(1):133–42.
13. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002;415(6871):530–6.
14. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R. A gene expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 2002;347(25):1999–2009.
15. Dowsett M, Sestak I, Lopez-Knowles E, Sidhu K, Dunbier AK, Cowens JW, Ferree S, Storhoff J, Schaper C, Cuzick J. 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(22):2783–90. <https://doi.org/10.1200/JCO.2012.46.1558>. Epub 2013 Jul.