REVIEW

Breast Cancer Molecular Stratification

From Intrinsic Subtypes to Integrative Clusters

Hege G. Russnes,^{*,†} Ole Christian Lingjærde,^{*,‡} Anne-Lise Børresen-Dale,^{*,§} and Carlos Caldas[¶]

From the Department of Cancer Genetics,* Institute for Cancer Research, and the Department of Pathology,[†] Oslo University Hospital Radiumhospitalet, Oslo, Norway; the Departments of Computer Science[‡] and Medicine,[§] University of Oslo, Oslo, Norway; and the Department of Oncology,[¶] Cancer Research UK Cambridge Institute, Li Ka Shing Centre, University of Cambridge, Cambridge, United Kingdom

Accepted for publication
April 27, 2017.

Address correspondence to
Carlos Caldas, M.D.,
F.Med.Sci., Cancer Research
UK Cambridge Institute,
University of Cambridge, Li Ka
Shing Centre, Robinson Way,
Cambridge CB2 0RE, United
Kingdom. E-mail: carlos.caldas@cruk.cam.ac.uk.

Breast carcinomas can be stratified into different entities based on clinical behavior, histologic features, and/or by biological properties. A classification of breast cancer should be based on underlying biology, which we know must be determined by the somatic genomic landscape of mutations. Moreover, because the latest generations of anticancer agents are founded on biological mechanisms, a detailed molecular stratification is a requirement for appropriate clinical management. Such stratification, based on genomic drivers, will be important for selecting patients for clinical trials. It will also facilitate the discovery of novel drivers, the study of tumor evolution, and the identification of mechanisms of treatment resistance. Assays for risk stratification have focused mainly on response prediction to existing treatment regimens. Molecular stratification based on gene expression profiling revealed that breast cancers could be classified in so-called intrinsic subtypes (luminal A and B, HER2-enriched, basal-like, and normal-like), which mostly corresponded to hormone receptor and HER2 status, and further stratified luminal tumors based on proliferation. The realization that a significant proportion of the gene expression landscape is determined by the somatic copy number alterations that drive expression in *cis* led to the newer classification of breast cancers into integrative clusters. This stratification of breast cancers into integrative clusters reveals prototypical patterns of single-nucleotide variants and is associated with distinct clinical courses and response to therapy. (*Am J Pathol* 2017, 187: 2152–2162; <http://dx.doi.org/10.1016/j.ajpath.2017.04.022>)

Breast Cancer Classification and Patient Stratification

Grouping tumors into classes or entities is of importance for several reasons. In clinical management, categorization of tumors is a tool to decide or standardize treatment and patient care. Furthermore, a robust and objective classification is important when performing clinical trials in which response to therapy is evaluated. Likewise, robust subtypes are needed in epidemiologic and functional studies to learn more about mechanisms in tumor development and evolution during treatment with focus on response and resistance. A classification should have distinct entities recognized in an objective way, and single specimens should be assigned to predefined classes by reproducible methods. The

traditional way of constructing taxonomies in biology is using a tree-based approach in which major classes can have smaller subgroups, an approach suited for cancer classification as well. Breast cancer diagnostics are

Supported by the Cancer Research UK Cambridge Institute, a Cancer Research UK Major Cancer Centre award, the Cambridge Experimental Cancer Medicine Centre, and the National Institute for Health Research Cambridge Biomedical Research Centre (C.C.); and The Norwegian Cancer Society, The Norwegian Research Council, The Health Region South-East of Norway, The K. G. Jebsen Foundation, and Radiumhospitalets legater (A.-L.B.-D. and H.G.R.).

Disclosures: None declared.

This article is part of a review series on next-generation breast cancer omics.

Figure 1 Breast cancer diagnostics have several components for which clinical information and histopathologic analysis in near future will be accompanied by molecular-based classification. This will provide the basis for deciding standard treatment, planning for follow-up, selecting clinical trials, and strengthening focused translational research. Used with permission from Ellen Margrethe Tenstad (Science Shaped). ER, estrogen receptor; PgR, progesterone receptor.

multidisciplinary, with a molecular-based classification as one of the components (Figure 1).

By histopathologic analysis, microscopic examination of breast carcinomas reveals heterogeneity both at the cellular level and in the architectural structure. The cellular compositions can range from stroma-rich tumors with glandular structures of tumor cells with minimal atypia to tumors with large, highly atypical carcinoma cells growing as solid sheets and tumors with atypical cells intermingled with stroma, preinvasive tumor components, and normal breast glands. Using these histologic patterns, breast carcinomas can be classified according to the World Health Organization's recommendations.¹ The most frequent histologic type is invasive ductal carcinoma (invasive ductal carcinoma not otherwise specified). Invasive ductal carcinoma not otherwise specified constitutes a heterogeneous group of tumors that do not have sufficient characteristics of any of the special differentiation patterns. The remaining breast carcinomas are classified into special types based on the dominating growth pattern. With this approach, a tumor with a predominant (ie, >90% of the tumor) tubular differentiation will be recognized as a distinct entity, as will a tumor with apocrine, lobular, cribriform, mucinous, or medullary features. Of the special types, lobular carcinomas are most common (10% to 15%), whereas some of the others are extremely rare (<1%). Several of the rare subgroups have different clinical courses and outcomes.^{2,3}

Histologic grading is more important than the morphologic type for clinical management of patients with breast cancer. The most widely used system is Bloom and Richardson's⁴ histologic grade, modified by Elston and Ellis in 1991.⁵ Here, three features are assessed: the proportion of tubule formation, the mitotic count, and the degree of nuclear pleomorphism.

Breast cancer classification has, over the years, gradually shifted from being purely descriptive (ie, based on

morphologic findings) to be more integrative, taking into account both clinical features and tissue-based biomarkers. In most countries, this is designed as a tree-based model in which clinical information, for instance age and the extent of the disease, is taken into account with tumor biomarkers.

More than 75% of breast cancers will express estrogen receptor (ER) and/or progesterone receptor (PgR), and these markers have both predictive and prognostic value.^{6,7} The most widely used technique to measure ER and PgR protein expression is by immunohistochemistry (IHC). An advantage of the method is the visual evaluation, which prohibits normal glandular epithelium in the biopsy specimen, expressing these markers, causing false-positive results. In addition, the number and intensity of stained cells can be measured in a semiquantitative way. The major disadvantages are the specificity and sensitivity of different antibodies used, the detection systems, and protocols, causing interlaboratory differences. Hence, participation in quality assessment programs is of major importance. The interobserver variation is also a challenge. Approximately 10% to 15% of breast cancers will also express the receptor protein HER2/erbB2/neu, a receptor tyrosine kinase that is involved in regulation of cellular growth. HER2 is regarded as a prognostic and predictive factor, with the advent of anti-HER2 targeted therapy.⁸ Tumors with overexpression of HER2 (almost always attributable to increased copy numbers of the *HER2* gene) are called HER2⁺ tumors and are most often identified according to international guidelines,⁹ using a combination of IHC and *in situ* DNA hybridization techniques. In addition, most recent treatment guidelines recommend that assessing proliferation should be added, especially for ER⁺ tumors, because this will identify patients who can benefit from adjuvant chemotherapy treatment. The last consensus meeting in St. Gallen recommended using a multigene test, if available.¹⁰

During the last few decades, thousands of breast cancer samples have been extensively studied with a variety of molecular analyses. The 3 largest cohorts for which high-resolution molecular analyses at multiple levels have been performed are The Cancer Genome Atlas project (<https://cancergenome.nih.gov>, last accessed April 1, 2017), the International Cancer Genome Consortium (<https://icgc.org>, last accessed April 1, 2017), and the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC, www.cbioportal.org, last accessed April 1, 2017). The aim of these projects has been to improve the existing classification scheme by incorporating extensive molecular profiling and eventually leading to the identification of patients with breast cancer who can benefit from more novel therapeutic options. Information from molecular analyses has increased enormously, and we are now facing an era in which molecular assays can be implemented into routine breast cancer diagnostics aiming at the stratification needed for optimal treatment selection.¹¹ This review is focused on the two most validated molecular classifiers: the PAM50/intrinsic subtypes and the integrative clusters.^{12–14}

The Intrinsic Subtypes: Class Discovery Based on Gene Expression

The intrinsic subtypes emerged as a result of the work by Perou et al¹² and Sørlie et al¹³ more than a decade ago by analyzing gene expression in breast carcinoma samples taken before and after chemotherapy. The expression of approximately 9000 genes was measured by cDNA arrays, and genes that had more similar expression between two samples from the same patient versus those that varied between tumors from different patients were selected. With these simple criteria, a total of approximately 550 genes were identified and named the intrinsic gene list because these genes were presumably reflecting the phenotype of individual tumors. Hierarchical clustering based on the patterns of expression of these genes across all the samples revealed two main clusters (mostly stratifying according to ER status) with a total of five subclusters called intrinsic subtypes. These five intrinsic subtypes are also found in samples from several independent early-stage breast cancer cohorts.^{15–19} For most cohorts, the largest cluster is dominated by ER positive (ER⁺) tumors and has frequently two subgroups, both dominated by expression of genes normally expressed by luminal breast epithelial cells (hence called luminal breast cancers). The subgroup named luminal B generally has a higher expression of genes involved in mitosis and cell proliferation. The other main cluster is enriched for ER negative (ER[−]) tumors and often has two subgroups. One is dominated by ER[−], PgR[−], and HER2[−] tumors [triple-negative breast carcinoma (TNBC)], showing expression of genes typical of myoepithelial/basal epithelial cells, such as basal cytokeratins, and thus named basal-like tumors. The other group is often dominated by high expression of *erbB2/HER2*-related genes (called HER2-

enriched group). A fifth group of tumors has gene expression similar to patterns found in normal breast tissue samples and was thus called normal-like.

Although both clinical and biological features distinguish the five intrinsic subtypes, there is substantial variation within each group as well. Basal-like tumors share the fewest similarities with the other groups but have the greatest intrinsic diversity.^{20–23} These patients have a poor prognosis, and almost 40% of patients experience a relapse within 5 years after diagnosis.^{20,24} Focusing on TNBC tumors, Lehmann et al,^{25,26} by analyzing gene expression patterns, identified four distinct subtypes; two were basal-like but with differences in immune response (BL1 and BL2), one mesenchymal, and one luminal androgen receptor subtype. These overlapped with four TNBC subtypes proposed by Burstein et al²⁷ and Teschendorff and colleagues.^{22,28} Mutations of the *TP53* gene are frequent in this group, and the tumors display a variety of alterations with regard to copy number alterations (CNAs) and gene mutations.^{29–33}

The definition of the HER2-enriched subtype is debated. There is an incomplete overlap of this group and tumors defined as HER2⁺ by international guidelines. Hence, the intrinsic subtypes cannot be used as a predictor of HER2-targeted therapy. The tumors classified as HER2 enriched also display diversity, where ER status, CNAs, and mutation profiles seem to identify different subsets.^{34–38} Perez et al^{39–41} reported results from a retrospective study of almost 1400 tumors from the prospective North Central Cancer treatment Group (Alliance) N9831 trial of patients with early-stage HER2⁺ breast cancer. Although most samples were classified as HER2-enriched by PAM50 (72.1%), a substantial number were classified as other subtypes, and a significant association between intrinsic subtype and survival was seen.³⁹

Furthermore, rarer subtypes, such as claudin low and molecular apocrine, have been identified. If these subtypes are not represented in a class discovery data set, they will not be detected, and the signature can lack the ability to recognize such cases as separate entities.^{42–44}

Although the first definition of the intrinsic subtypes occurred almost two decades ago, an assay suited for diagnostic use was introduced just recently: the US Food and Drug Administration–approved test Prosigna. The test is based on the modified version of the original intrinsic subtype definition, the PAM50, and it assigns each sample to the luminal A, luminal B, HER2-enriched, or basal-like subtype.³⁴ In addition, the test will provide a numeric score that, for ER⁺ patients, predicts the risk of relapse, with the potential to be informative for identifying women who can benefit from adjuvant treatment.^{34,45}

Subtypes Defined by Patterns of DNA CNAs

The microarray technology paved the ground for high-throughput gene expression analysis, but the technology

also increased the resolution enormously for genome-wide copy number analyses. In parallel with the discovery of the gene expression defined subtypes, subgroups of tumors with similarities in CNAs were identified. Four different patterns of alterations were described by Hicks et al⁴⁶ using high-resolution comparative genome hybridization arrays in tumors from two breast cancer cohorts. Tumors with the simplex pattern have broad segments of duplications and deletions. Deletion of chromosomes 16q, 8p, and/or 22, as well as gain of chromosomes 1q, 8q, and/or 16p, is dominating, and the tumors are frequently ER⁺ and luminal subtype. The tumors called complex I have a sawtooth appearance, with narrow segments of deletions and duplications affecting more or less all chromosomes, and are frequently TNBC and basal-like. The tumors of complex II resemble the simplex type but have at least one localized region of clustered peaks of high-level gene amplifications with intermittent deletions, called firestorm. Such tumors are more frequently luminal B or HER2-enriched. The fourth pattern was called flat, defining profiles with no clear gains or losses except from copy number polymorphisms. The same patterns of DNA alterations have been identified in other data sets.^{47,48} A study by Chin et al,³⁰ also using comparative genome hybridization arrays, identified three subtypes of breast carcinomas that varied with respect to level of genomic instability. The groups had overlapping characteristics with the classes from the work of Hicks et al.⁴⁶ One group of tumors had few alterations and was dominated by chromosome 1q whole arm amplification and chromosome 16q whole arm deletion (the 1q/16q group), another group had more complex alterations (complex group), and the other had frequently high-level amplifications (mixed amplifier group). The whole arm gains and losses are attributable to translocations close to centromeres even in preinvasive components of the tumors.⁴⁹ Tumors with *BRCA1* mutation had similar pattern of changes as tumors in the complex/sawtooth group. Divergent definitions with regard to which genomic alterations characterize distinct subgroups of breast carcinomas have been published, but also older studies found chromosome 1q and 16q alterations to dominate in one type and multiple alterations on several arms to dominate in another type of breast cancer, even at the preinvasive stage of the disease.^{50–55} With regard to the intrinsic subtypes, selected subtype-specific alterations can to some degree be used as surrogate markers.^{29,30,48,56} Furthermore, by developing algorithms to detect the more complex CNAs, such as firestorm events, a subdivision of intrinsic subtypes attributable to the presence or absence of complex CNAs appears to have a clinical impact.^{48,57}

Recently, signatures based on DNA rearrangement patterns derived from whole genome sequencing data were published.⁵⁸ The information about type of rearrangement (deletions, tandem duplications, inversions, and translocations) as well as the size of the affected part and whether they were focally or genomically dispersed was used to

identify six different rearrangement signatures (RS1 to RS6). The main features of the signatures resemble some of the types of architectural alterations already found to characterize different types of breast cancer. RS1 and RS3 are characterized by tandem duplications (sawtooth profile), RS4 and RS6 by clustered rearrangements (firestorm pattern), RS5 by deletions, and RS2 by translocations (simplex pattern).⁵⁸ For most tumors, the genomic landscape of rearrangements is composed of combinations of these signatures.⁵⁹ For instance, RS3 is present in a substantial number of TNBCs and basal-like breast carcinomas, in line with earlier work finding TNBCs to have frequent tandem duplications.⁶⁰ Tandem duplications are attributable to deficiencies in homologous recombination, in line with breast cancer in women with germline *BRCA1* defects being most frequently the TNBC or basal-like type.

The Integrative Clusters: Stratification Based on Genomic Copy Number Drivers Identified by Combining Gene Expression and DNA CNAs

METABRIC used a combined approach for class discovery. Genes whose expression across tumors was driven in *cis* by recurrent CNAs, which by definition enriches for genomic drivers (oncogenes, those whose overexpression is associated with copy number gains/amplifications, and tumor suppressor genes, those whose underexpression is associated with copy number losses), were first identified using a modified expression quantitative trait loci analysis. These *cis*-driven genes were then used to cluster the tumors into groups called integrative clusters.¹⁴ Rigorous statistical analyses revealed that the most parsimonious solution classified the tumors into 10 subtypes (clusters 1 through 10). An important advantage of the METABRIC study was the large sample sets used for class discovery ($n = 997$) and class validation ($n = 995$). The patients in the cohort had relatively consistent treatment regimens and long clinical follow-up (mean, 10 years), and the samples were characterized in detail with regard to CNAs and gene expression. The METABRIC study thus represented a unique data set for class discovery when initially published. The frequency of the integrative cluster subtypes, their copy number profiles, and the clinical outcome was validated in the additional 995 samples from the validation set. Recently, a gene expression surrogate profile for classification was described based on the expression level of 612 genes, with 10 centroids defined and used for class prediction.⁶¹ An extensive validation was performed for a total of 7500 samples, in which similar proportions of each of the integrative cluster subtypes was seen, and their molecular and clinical properties confirmed.

To date, the 10 integrative cluster subtypes represent the most extensive molecular-based taxonomy of breast cancer. This taxonomy partly captures subgroups defined by other approaches but importantly also groups tumors in more novel subtypes.

Table 1 Overview of the Integrative Cluster Subtypes and the Dominating Properties with Regard to Copy Number Driving Events, Biomarkers, Type of DNA Architecture,⁴⁶ Dominant PAM50 Subtype, and Clinical Outcome

Integrative cluster group	Copy number driver	Pathology biomarker class	DNA architecture	Dominant PAM50	Clinical characteristics (survival)
1	Chromosome 17/ chromosome 20	ER ⁺ (HER2 ⁺)	Simplex/firestorm (chromosome 17q)	Luminal B	Intermediate
2	Chromosome 11	ER ⁺	Firestorm (chromosome 11q)	Luminal A and B	Poor
3	Very few	ER ⁺	Simplex/flat	Luminal A	Good
4	Very few	ER ⁺ /ER ⁻	Sawtooth/flat	Luminal A (mixed)	Good (immune cells)
5	Chromosome 17 (<i>HER2</i> gene)	ER ⁻ (ER ⁺)/HER2 ⁺	Firestorm (chromosome 17q)	Luminal B and HER2	Extremely poor (in pre- Herceptin cohorts)
6	8p deletion	ER ⁺	Simplex/firestorm (chromosome 8p/ chromosome 11q)	Luminal B	Intermediate
7	Chromosome 16	ER ⁺	Simplex (chromosome 8q/chromosome 16q)	Luminal A	Good
8	Chromosome 1, Chromosome 16	ER ⁺	Simplex (chromosome 1q/chromosome 16q)	Luminal A	Good
9	Chromosome 8/ Chromosome 20	ER ⁺ (ER ⁻)	Simplex/firestorm (chromosome 8q/ chromosome 20q)	Luminal B (mixed)	Intermediate
10	Chromosome 5, Chromosome 8, Chromosome 10, Chromosome 12	TNBC	Complex/sawtooth	Basal-like	Poor 5-year, good long-term if survival

ER, estrogen receptor; TNBC, triple-negative breast carcinoma.

Six of the integrative cluster groups (clusters 1, 2, 3, 6, 7, and 8) are dominated by ER⁺ samples and the PAM50 subtypes luminal A and luminal B, but the groups have distinct genomic alterations (Table 1). Samples representing integrative cluster 3 have few genomic alterations but are dominated by whole arm gain of chromosomes 1q and 16p and loss of 16q, thus showing a simplex pattern of rearrangements. Integrative clusters 7 and 8 are also dominated by samples with whole arm gains and losses, but different chromosome arms are affected, probably reflecting tumorigenesis driven by centromere-close translocations that involve different chromosomes. Samples in integrative cluster 2 also have whole arm gains and losses that affect the same chromosomes but in addition high-level amplifications on chromosome 11q, which are narrow and reflect a typical firestorm event that affects the known driver genes *CCND1*, *EMSY*, and *PAK1*. This subtype has both luminal A and luminal B samples and represents an important subgroup of ER⁺ patients with a very poor prognosis. It also shows that the integrative cluster stratification is predictive independently of its prognostic value; despite these integrative cluster 2 tumors being aggressive clinically, their response to neoadjuvant chemotherapy is minimal.⁶¹

Tumors in integrative cluster 1 are characterized by amplifications of 17q, distal to *ERBB2*, and apparently targeting *RPS6KB1*, *PPM1D*, *PTRH2*, and *APPBP2*. Tumors in integrative cluster 6 are characterized by amplification of

8p12, with a minimal consensus amplicon dominated by a single gene, *ZNF703*, which has now been shown to be a novel breast cancer oncogene.^{62,63}

The samples in integrative cluster 5 are both ER⁺ and ER⁻, and the genomic alterations are dominated by the high-level amplification on 17q, centered and including the *HER2* gene. Interestingly, many of the samples in this group have whole arm gains and losses, in particularly affecting chromosomes 1 and 8. They are recognized by PAM50 as mainly luminal B or HER2 enriched, but a substantial number of luminal A and basal-like samples are also found within integrative cluster 5. In the METABRIC study, the patients in this subgroup had extremely poor prognosis, but none of these patients received adjuvant anti-HER2-targeted treatment.

The group dominated by ER⁻ samples is integrative cluster 10, encompassing tumors with a typical sawtooth pattern of alterations, with multiple CNAs affecting most of the chromosomes. The tumors only rarely have high-level amplifications, but deletions of chromosome 5q and gain of chromosomes 9p and 10p are significant events that identify tumors in this group and are known from previous studies to be a hallmark of basal-like tumors.⁴⁸ The samples are almost always of basal-like type and have a poor short-term (5 years) prognosis but a good prognosis for patients who survive the first 5 years after diagnosis. Consistently, a small number of ER⁺ breast cancers are stratified into integrative cluster 10, effectively identifying

a new class of basal-like ER⁺ breast cancers with poor prognosis.

Another group capturing both ER⁺ and ER⁻ tumors is integrative cluster 9, which includes tumors with many copy number changes, again dominated by the whole arm changes of chromosomes 1, 8, and 16, but in addition almost all tumors have amplifications on 8q (including the oncogene *MYC*) and very frequent losses of 13q and 17p, where important tumor suppressor genes reside (*RB* and *TP53*, respectively). The samples in this group represent a mixture of PAM50 subtypes and have an intermediate-poor prognosis.

The integrative cluster 4 subtype is characterized by samples in which the CNAs are negligible, and the gene expression profile is actually dominated by immune-related genes. On the basis of PAM50, most samples are of luminal B and basal-like type, but the patients have a good-intermediate prognosis. Samples with nonaltered genomes might be explained by a combination of high infiltration of nonaberrant cells (ie, lymphocytes) with low-level CNAs. This subtype represents two distinct entities: one (integrative cluster 4-ER⁺) more related to the other good prognosis ER⁺ integrative cluster subtypes (integrative clusters 3, 7, and 8), and the other (integrative cluster 4-ER⁻) related to integrative cluster 10 but with a distinctive lymphocytic infiltration (mostly of T cells).

For diagnostic purposes, a classification in which both genome-wide copy number analyses and gene expression are needed would probably be too cumbersome and expensive for routine use. For single-sample prediction, the surrogate test based on the expression levels of the 620 identified genes that recapitulate the 10 integrative cluster classes (and thus identify presence of important CNAs) seems promising.⁶¹ Indeed, a test based on the same technology used for Prosigna is currently being developed.

Patterns of Gene Mutations in Breast Cancer

Massive parallel DNA sequencing has given insight into frequencies and distributions of a range of mutation types, including base substitutions, small insertions and deletions, as well as structural rearrangements and CNAs.^{14,33,59,64–66} These large-scale mutation data series have supported the notion that some genes seem to be important drivers in the evolution of different subsets of breast cancer and that there is an association between genomic drivers and expression-based subtypes. In one of the largest breast cancer data sets analyzed to date, the most frequent mutated genes were *TP53*, *PIK3CA*, *MYC*, *CCND1*, *PTEN*, *FGFR1*, *GATA3*, *RBI*, *ERBB2*, and *MAP3K1*.⁵⁹ In the METABRIC study, a selected gene panel was used and a ratiometric analysis identified 40 so-called Mut-driver genes.⁶⁶ Mutations of some of the genes were associated with clinical and pathologic features; for instance, *PIK3CA*, *GATA3*, *MAP3K1*, *KMT2C*, and *CBFB* were commonly mutated in ER⁺

tumors of low and intermediate histologic grade. Both mutation of *CDH1* and loss of allele of the gene are common findings in lobular carcinomas, and mutations in the *CDH1* and *CBFB* genes were associated with concurrent whole arm alterations of chromosomes 1 and 16, the feature most often seen in ER⁺ tumor carcinomas. Mutations in *TP53* were associated with higher grade in both ER⁺ and ER⁻ tumors, and mutations in *CDH1* and *HER2* were uncommon in ER⁻ cancer. The base position and the prognostic value of *TP53* mutation vary between the subtypes.⁶⁷ A subtype-dependent distribution is also seen for *PIK3CA* mutations; in the PAM50 subtypes, luminal A and luminal B tumors have the highest frequency of mutations in this gene.⁶⁴ A variation is found across the integrative cluster subtypes; the three groups with ER⁺ tumors and good prognosis have the highest number of cases with *PIK3CA* mutation (integrative clusters 3, 7, and 8), the ER⁺ groups with intermediate prognosis have lower numbers of mutated cases (integrative clusters 2, 6, and 1), and tumors in the remaining subtypes have low frequencies of mutated *PIK3CA*. One exception is integrative cluster 4, but when the tumors in this group are stratified by ER status, the ER⁺ cases have a mutation frequency within the range of the integrative clusters 3, 7, and 8 subgroups.⁶⁶ The same pattern, but with opposite distribution of frequencies, is also found for *TP53* mutations. Interestingly, the ER⁺ samples that by integrative cluster were classified into the integrative cluster 10, which is dominated by TNBCs and basal-like tumors, had a high frequency of *TP53* mutations, again reflecting the uniqueness of these ER⁺ tumors having a pathogenesis not driven by ER but by alterations disrupting DNA repair mechanisms.

HER2⁺ tumors show different combinations of mutations, again supporting the findings of important subclasses of HER2⁺ tumors. Functional mutations in *TP53* are more frequent in HER2⁺/ER⁻ tumors than in HER2⁺/ER⁺ tumors. Mutations typical for luminal tumors, such as *PIK3CA* and *GATA3*, are also found at substantial frequencies in HER2⁺ tumors, but the HER2⁺/ER⁺ tumors have more frequent *GATA3* mutations than the HER2⁺/ER⁻ tumors.^{64,66} *PIK3CA* mutations are found at a lower frequency in HER2⁺ tumors than in HER2⁻/ER⁺ tumors, but the frequency does not differ between HER2⁺/ER⁺ and HER2⁺/ER⁻ samples, which is of interest because the presence of *PIK3CA* mutated cells in HER2⁺ breast tumors seems to predict therapy resistance.⁶⁸

Tumors of the TNBC type are dominating the basal-like and the integrative cluster 10 subtypes. Multiple studies have found the most frequently mutated gene to be *TP53*,^{64,67} but the total number of mutated genes in such tumors has a substantial variation.^{33,59,64,69}

Single-base mutations in DNA occur not randomly but in a distinct context, depending on the mutagenic process, captured by the so-called mutation signatures.⁶⁹ This is based on the different types of base substitution types and information about the base immediately 5' and 3' to each of

the substitutions, found both in coding and noncoding parts of the genome. This is not a traditional classification scheme, but each tumor will have a proportion of a signature present or not. For instance, tumors with a defective BRCA1 function will have a particular signature, as will tumors induced by known carcinogens, such as UV light. At least 12 of these signatures are frequently present in breast cancer samples and have some correlation with subtypes defined by ER/PgR/HER2, gene expression, and structural genomic alterations⁵⁹

Finally, sequencing data can be used to infer the clonal architecture of tumors. One way to assess this is by analyzing the degree of intratumor heterogeneity of the mutant alleles. The more heterogeneous tumors are presumably the ones with a more complex clonal architecture. In the METABRIC data set, this analysis revealed two important findings: mutant allele intratumor heterogeneity and chromosomal instability. These findings were measured as the fraction of the genome altered by CNAs and appear to be directly correlated, and each integrative cluster appears to have a prototypical level of both intratumor heterogeneity and chromosomal instability.⁶⁶ This finding further supports the stratification of breast cancer into each of the integrative clusters as true distinct genome driver–based entities.

Multilevel Analysis to Improve Clinical Stratification

Regardless of which molecular level is analyzed, breast carcinomas can be stratified into biologically and clinically distinct subtypes. How many subgroups we need to define is probably dependent on the purpose of having a taxonomy of breast cancer. We foresee that molecular classification will give information alongside clinical examination, imaging analyses, and histopathologic examination (Figure 1). Information from

all these disciplines will help clinicians make treatment decisions and select subsets of patients for clinical trials and researchers perform more focused translational research. A taxonomy should be dynamic; as new knowledge, diagnostic procedures, or treatment modalities emerge, the taxonomy must be refined, for instance, by substratification of subtypes.

A major difference between the intrinsic subtypes/PAM50 and integrative cluster classifications, which are dealt with here, is the number of subgroups. Because of the benefit of integrating genomic events with gene expression alterations, integrative clusters define more groups. The two systems have overlapping findings, such as identifying a distinct subgroup of TNBCs (integrative cluster 10 and basal-like subtypes). In addition, integrative clusters have been able to separate out a group of TNBCs in which immune response seems to be of importance, and this division of basal-like tumors is probably in line with immunomodulatory subgroups defined by Lehmann et al²⁶ and Teschendorff et al.²² Integrative clusters also classify some basal-like samples into more luminal/ER⁺ dominated subgroups. Basal-like tumors with features more commonly found in luminal disease might be clinically important.⁷⁰

Furthermore, by integrative clusters, ER⁺ tumors dominate in six of the subgroups, whereas PAM50 only has two ER⁺-related subgroups. Interestingly, by looking at correlation values to all five PAM50 centroids, the diversity of luminal samples is also seen (Figure 2).

Intrinsic subtype class prediction is made by correlating the expression patterns of the selected genes from an individual sample to predefined centroids for each of the five main subtypes, and each sample is assigned to the class with the highest correlation value. The variation in the correlation level should be acknowledged; most samples have a high correlation to more than one of the centroids, and, on the other hand, some only have a weak correlation to all.¹⁷ The correlation values calculated for each sample for the five

Figure 2 Comparison of assignment of BASIS samples⁵⁹ to integrative cluster and PAM50 subtypes and the centroid correlation value to all five PAM50 centroids for each sample. An indication of the type of DNA architectural pattern frequently found within each integrative cluster subtypes⁴⁶ is given below as is an illustration of which subgroup is dominated by estrogen receptor (ER) and/or HER2 positivity (ER⁺ is blue, ER⁻ is red, HER2⁺ is purple). Basal, basal-like; HER2, HER2-enriched; LumA, luminal A; LumB, luminal B; Normal, normal-like. Used with permission from Ellen Margrethe Tenstad (Science Shaped).

centroids have some interesting variation (Figure 2). For instance, samples in integrative cluster 4+ and 3 with a luminal A subtype have frequently the second highest correlation to the normal-like centroid. This finding is in contrast to luminal A samples in integrative cluster 7 and 8 in which correlation to the luminal B centroid is frequently high. Furthermore, luminal B samples have a high correlation to the luminal B centroid; however, samples classified as integrative clusters 6, 1, and 9 also have a strong correlation to the HER2-enriched centroid, whereas luminal B samples from integrative cluster 4+, 3, and 7 frequently have a correlation to the luminal A centroid. This findings indicates that the gene-expression pattern of these 50 genes probably reflects a second level of intertumor diversity that might have importance for subclassification.

The consensus guidelines for breast cancer stratification recommend ER⁺ breast cancer to be considered a spectrum of diseases. By using PAM50, this can be solved by a predictor score for risk of relapse. Bartlett et al⁷¹ reported recently that for ER⁺ patients different multiparameter tests report broadly equivalent risk information for this population of patients, but a variation for the individual patients was substantial. Such tests are of some benefit in clinical practice of today, but generally they do not give information about which molecular processes are ongoing in a given tumor. This finding is in contrast to stratification into several subtypes of ER⁺/luminal type of breast cancer as seen by the seven integrative clusters dominated by ER⁺ cancers (1, 2, 3, 6, 7, 8, and 9) because genomic changes are significantly different between these groups.

Another example of a clinically challenging situation is how to identify patients who will benefit from anti-HER2—targeted therapy. The recommendation is to use a combination of protein analysis (IHC based) and *in situ* DNA copy number assessment of the gene as described in detailed guidelines.⁹ HER2⁺ tumors are found within all intrinsic subtypes,³⁸ and substratification of HER2⁺ cases by other molecular features seems to have an important clinical impact.³⁹ HER2⁺ tumors are enriched in several of the integrative cluster groups, in particular in 1, 5, 6, and 9. Interestingly, the type of CNAs of chromosome 17q, where *HER2* resides, varies among these 4 groups. Although integrative cluster 5 typically has a narrow high-level amplification that affects *HER2* and some neighboring genes, samples in integrative clusters 1, 6, and 9 have broader regions on chromosome 17q amplified. If *HER2* is involved, it is often with lower levels of gain than seen in samples from integrative cluster 5. This is probably reflecting different types of underlying biological disruption causing *HER2* copy number amplification, but it still remains to be seen whether this stratification can predict therapy response.

Challenges and Conclusion

There have been many different approaches to define molecular subtypes of breast cancer. Biological properties can

be described at many levels, and they can change during a tumor's lifetime because of phenotypic dynamics and genomic evolution.

Gene expression analyses have found that molecular subtypes share phenotypic traits with different types of breast epithelium cells. The hierarchy of development from stem cell to the lineage committed breast epithelial cells is only partly known, but breast cancer subtypes seem to reflect some features of distinct levels of development.⁷² Currently, the genes holding most mutations and high-level amplification driver events are probably known, but driver events attributable to low-level CNAs of large genomic regions or structural nonrecurrent events have been a challenge to identify. Because the integrative cluster approach defined driver events due to correlation between copy number changes and gene expression, a substantial number of tumors seem to have drivers of this type. This is probably the main reason that integrative clusters differ from intrinsic subtypes/PAM50 and can stratify breast carcinomas into more refined subtypes.

Whether tumors follow one path of progression or several or which alterations characterize the different levels of progression still remains to be defined. A challenge in class discovery is the need to study clinical samples. Breast cancers are diagnosed (and thus sampled) in patients at different stages of the disease and are rarely sampled at several time points during the development of the disease. Because a tumor can evolve and change phenotype and/or acquire additional genomic changes, some subtypes might represent more advanced stages of other subtypes, and this needs to be addressed in more detail.

Finally, it is important that clinical outcome, such as survival, is not a test for the validity of a classifier. The most important feature is robustness in biological properties, preferably by integrating data from multiple molecular levels. The last decades have provided us with enormous amounts of knowledge about molecular alterations in breast cancer. As discussed in this review, we do not face a situation with competing molecular classifications; classifiers are highly overlapping because of measuring different reflections of the main biological properties of the tumors. We should hope that during the next decade all disciplines in the breast cancer community will come to a consensus about a molecular classification scheme to be used in the best interest of the patients.

Acknowledgments

We appreciate the numerous discussions about breast cancer classifications with our colleagues and laboratory members over the years. We thank Ellen Margrethe Tenstad (Science Shaped) for assistance with the figures.

References

1. Lakhani S, Ellis IO, Schnitt SJ, Tan PH, Van de Vijver MJ: WHO Classification of Tumors of the Breast, ed 4. Geneva, Switzerland, World Health Organization, 2012

2. Weigelt B, Horlings HM, Kreike B, Hayes MM, Hauptmann M, Wessels LFA, de Jong D, Van de Vijver MJ, Van't Veer LJ, Peterse JL: Refinement of breast cancer classification by molecular characterization of histological special types. *J Pathol* 2008, 216: 141–150
3. Horlings HM, Weigelt B, Anderson EM, Lambros MB, Mackay A, Natrajan R, Ng CKY, Geyer FC, van de Vijver MJ, Reis-Filho JS: Genomic profiling of histological special types of breast cancer. *Breast Cancer Res Treat* 2013, 142:257–269
4. Bloom HJ, Richardson WW: Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. *Br J Cancer* 1957, 11:359–377
5. Elston CW, Ellis IO: Pathological prognostic factors in breast cancer, I: the value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* 1991, 19: 403–410
6. Early Breast Cancer Trialists' Collaborative Group (EBCTCG): Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 2005, 365:1687–1717
7. Fisher B, Redmond CK, Fisher ER: Evolution of knowledge related to breast cancer heterogeneity: a 25-year retrospective. *J Clin Oncol* 2008, 26:2068–2071
8. Loibl S, Gianni L: HER2-positive breast cancer. *Lancet* 2017, 389: 2415–2429
9. Wolff AC, Hammond MEH, Hicks DG, Dowsett M, McShane LM, Allison KH, Allred DC, Bartlett JMS, Bilous M, Fitzgibbons P, Hanna W, Jenkins RB, Mangu PB, Paik S, Perez EA, Press MF, Spears PA, Vance GH, Viale G, Hayes DF; American Society of Clinical Oncology; College of American Pathologists: 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
10. Coates AS, Winer EP, Goldhirsch A, Gelber RD, Gnant M, Piccart-Gebhart M, Thürlimann B, Senn HJ: Tailoring therapies—improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. *Ann Oncol* 2015, 26:1533–1546
11. Rakha EA, Green AR: Molecular classification of breast cancer: what the pathologist needs to know. *Pathology* 2017, 49:111–119
12. Perou CM, Sørli T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO, Botstein D: Molecular portraits of human breast tumours. *Nature* 2000, 406:747–752
13. Sørli T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Lønning PE, Børresen-Dale AL: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001, 98:10869–10874
14. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S; METABRIC Group, Langerød A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowitz F, Murphy L, Ellis I, Purushotham A, Børresen-Dale A-L, Brenton JD, Tavaré S, Caldas C, Aparicio S: The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012, 486:346–352
15. Sørli T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lønning PE, Brown PO, Børresen-Dale A-L, Botstein D: Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 2003, 100:8418–8423
16. Langerød A, Zhao H, Borgan Ø, Nesland JM, Bukholm IRK, Ikdahl T, Kåresen R, Børresen-Dale A-L, Jeffrey SS: TP53 mutation status and gene expression profiles are powerful prognostic markers of breast cancer. *Breast Cancer Res* 2007, 9:R30
17. Naume B, Zhao X, Synnæstvedt M, Borgen E, Russnes HG, Lingjærde OC, Strømberg M, Wiedswang G, Kvalheim G, Kåresen R, Nesland JM, Børresen-Dale A-L, Sørli T: Presence of bone marrow micrometastasis is associated with different recurrence risk within molecular subtypes of breast cancer. *Mol Oncol* 2007, 1: 160–171
18. Calza S, Hall P, Auer G, Bjöhle J, Klaar S, Kronenwett U, Liu ET, Miller L, Ploner A, Smeds J, Bergh J, Pawitan Y: Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients. *Breast Cancer Res* 2006, 8:R34
19. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, Livasy C, Carey LA, Reynolds E, Dressler L, Nobel A, Parker J, Ewend MG, Sawyer LR, Wu J, Liu Y, Nanda R, Tretiakova M, Ruiz Orrico A, Dreher D, Palazzo JP, Perreard L, Nelson E, Mone M, Hansen H, Mullins M, Quackenbush JF, Ellis MJ, Olopade OI, Bernard PS, Perou CM: The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* 2006, 7:96
20. Badve S, Dabbs DJ, Schnitt SJ, Baehner FL, Decker T, Eusebi V, Fox SB, Ichihara S, Jacquemier J, Lakhani SR, Palacios J, Rakha EA, Richardson AL, Schmitt FC, Tan P-H, Tse GM, Weigelt B, Ellis IO, Reis-Filho JS: Basal-like and triple-negative breast cancers: a critical review with an emphasis on the implications for pathologists and oncologists. *Mod Pathol* 2011, 24:157–167
21. Kreike B, van Kouwenhove M, Horlings H, Weigelt B, Peterse H, Bartelink H, van de Vijver MJ: Gene expression profiling and histopathological characterization of triple-negative/basal-like breast carcinomas. *Breast Cancer Res* 2007, 9:R65
22. 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
23. Chin SF, Teschendorff AE, Marioni JC, Wang Y, Barbosa-Morais NL, Thorne NP, Costa JL, Pinder SE, van de Wiel MA, Green AR, Ellis IO, Porter PL, Tavaré S, Brenton JD, Ylstra B, Caldas C: High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. *Genome Biol* 2007, 8:R215
24. Fulford LG, Reis-Filho JS, Ryder K, Jones C, Gillett CE, Hanby A, Easton D, Lakhani SR: Basal-like grade III invasive ductal carcinoma of the breast: patterns of metastasis and long-term survival. *Breast Cancer Res* 2007, 9:R4
25. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, Pietenpol JA: Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest* 2011, 121:2750–2767
26. Lehmann BD, Jovanović B, Chen X, Estrada MV, Johnson KN, Shyr Y, Moses HL, Sanders ME, Pietenpol JA: Refinement of triple-negative breast cancer molecular subtypes: implications for neo-adjuvant chemotherapy selection. *PLoS One* 2016, 11:e0157368
27. Burstein MD, Tsimelzon A, Poage GM, Covington KR, Contreras A, Fuqua SAW, Savage MI, Osborne CK, Hilsenbeck SG, Chang JC, Mills GB, Lau CC, Brown PH: Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. *Clin Cancer Res* 2015, 21:1688–1698
28. Teschendorff AE, Caldas C: A robust classifier of high predictive value to identify good prognosis patients in ER-negative breast cancer. *Breast Cancer Res* 2008, 10:R73
29. Bergamaschi A, Hjortland GO, Triulzi T, Sørli T, Johnsen H, Ree AH, Russnes HG, Trønnes S, Mælandsmo GM, Fodstad O, Børresen-Dale A-L, Engebraaten O: Molecular profiling and characterization of luminal-like and basal-like in vivo breast cancer xenograft models. *Mol Oncol* 2009, 3:469–482
30. Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo W-L, Lapuk A, Neve DM, Qian Z, Ryder T, Chen F, Feiler H, Tokuyasu T, Kingsley C, Dairkee S, Meng Z, Chew K, Pinkel D,

- Jain A, Ljung BM, Esserman L, Albertson DG, Waldman FM, Gray JW: Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* 2006, 10:529–541
31. Fulford LG, Easton DF, Reis-Filho JS, Sofronis A, Gillett CE, Lakhani SR, Hanby A: Specific morphological features predictive for the basal phenotype in grade 3 invasive ductal carcinoma of breast. *Histopathology* 2006, 49:22–34
 32. Livasy CA, Karaca G, Nanda R, Tretiakova MS, Olopade OI, Moore DT, Perou CM: Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. *Mod Pathol* 2006, 19:264–271
 33. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al: The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* 2012, 486:395–399
 34. Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, Palazzo J, Marron JS, Nobel AB, Mardis E, Nielsen TO, Ellis MJ, Perou CM, Bernard PS: Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 2009, 27:1160–1167
 35. Rouzier R, Perou CM, Symmans WF, Ibrahim N, Cristofanilli M, Anderson K, Hess KR, Stec J, Ayers M, Wagner P, Morandi P, Fan C, Rabiul I, Ross JS, Hortobagyi GN, Pusztai L: Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clin Cancer Res* 2005, 11:5678–5685
 36. Staaf J, Ringnér M, Vallon-Christersson J, Jönsson G, Bendahl P-O, Holm K, Arason A, Gunnarsson H, Hegardt C, Agnarsson BA, Luts L, Grabau D, Fernö M, Malmström P-O, Johannsson OT, Loman N, Barkardottir RB, Borg A: Identification of subtypes in human epidermal growth factor receptor 2-positive breast cancer reveals a gene signature prognostic of outcome. *J Clin Oncol* 2010, 28:1813–1820
 37. Ferrari A, Vincent-Salomon A, Pivot X, Sertier A-S, Thomas E, Tonon L, et al: A whole-genome sequence and transcriptome perspective on HER2-positive breast cancers. *Nat Commun* 2016, 7:12222
 38. Carey LA, Berry DA, Cirincione CT, Barry WT, Pitcher BN, Harris LN, Ollila DW, Krop IE, Henry NL, Weckstein DJ, Anders CK, Singh B, Hoadley KA, Iglesia M, Cheang MCU, Perou CM, Winer EP, Hudis CA: Molecular heterogeneity and response to neoadjuvant human epidermal growth factor receptor 2 targeting in CALGB 40601, a randomized phase III trial of paclitaxel plus trastuzumab with or without lapatinib. *J Clin Oncol* 2016, 34:542–549
 39. Perez EA, Ballman KV, Mashadi-Hosseini A, Tenner KS, Kachergus JM, Norton N, Necela BM, Carr JM, Ferree S, Perou CM, Baehner F, Cheang MCU, Thompson EA: Intrinsic subtype and therapeutic response among HER2-positive breast tumors from the NCCTG (Alliance) N9831 Trial. *J Natl Cancer Inst* 2017, 109:djw207
 40. Perez EA, Romond EH, Suman VJ, Jeong J-H, Sledge G, Geyer CE, Martino S, Rastogi P, Gralow J, Swain SM, Winer EP, Colon-Otero G, Davidson NE, Mamounas E, Zujewski JA, Wolmark N: Trastuzumab plus adjuvant chemotherapy for human epidermal growth factor receptor 2-positive breast cancer: planned joint analysis of overall survival from NSABP B-31 and NCCTG N9831. *J Clin Oncol* 2014, 32:3744–3752
 41. Perez EA, Romond EH, Suman VJ, Jeong J-H, Davidson NE, Geyer CE, Martino S, Mamounas EP, Kaufman PA, Wolmark N: Four-year follow-up of trastuzumab plus adjuvant chemotherapy for operable human epidermal growth factor receptor 2-positive breast cancer: joint analysis of data from NCCTG N9831 and NSABP B-31. *J Clin Oncol* 2011, 29:3366–3373
 42. Weigelt B, Geyer FC, Reis-Filho JS: Histological types of breast cancer: how special are they? *Mol Oncol* 2010, 4:192–208
 43. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JJ, He X, Perou CM: Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res* 2010, 12:R68
 44. Farmer P, Bonnefoi H, Becette V, Tubiana-Hulin M, Fumoleau P, Larsimont D, Macgrogan G, Bergh J, Cameron D, Goldstein D, Duss S, Nicolaz A-L, Brisken C, Fiche M, Delorenzi M, Iggo R: Identification of molecular apocrine breast tumours by microarray analysis. *Oncogene* 2005, 24:4660–4671
 45. 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, 31:2783–2790
 46. Hicks J, Krasnitz A, Lakshmi B, Navin NE, Riggs M, Leibu E, Esposito D, Alexander J, Troge J, Grubor V, Yoon S, Wigler M, Ye K, Børresen-Dale A-L, Naume B, Schlichting E, Norton L, Hägerström T, Skoog L, Auer G, Månér S, Lundin P, Zetterberg A: Novel patterns of genome rearrangement and their association with survival in breast cancer. *Genome Res* 2006, 16:1465–1479
 47. Natrajan R, Lambros MB, Rodríguez-Pinilla SM, Moreno-Bueno G, Tan DSP, Marchio C, Vatcheva R, Rayter S, Mahler-Araujo B, Fulford LG, Hungermann D, Mackay A, Grigoriadis A, Fenwick K, Tamber N, Hardisson D, Tutt A, Palacios J, Lord CJ, Buerger H, Ashworth A, Reis-Filho JS: Tiling path genomic profiling of grade 3 invasive ductal breast cancers. *Clin Cancer Res* 2009, 15:2711–2722
 48. Russnes HG, Vollan H-KM, Lingjærde OC, Krasnitz A, Lundin P, Naume B, Sørli T, Borgen E, Rye IH, Langerød A, Chin S-F, Teschendorff AE, Stephens PJ, Månér S, Schlichting E, Baumbusch LO, Kåresen R, Stratton MP, Wigler M, Caldas C, Zetterberg A, Hicks J, Børresen-Dale A-L: Genomic architecture characterizes tumor progression paths and fate in breast cancer patients. *Sci Transl Med* 2010, 2:38ra47
 49. Rye IH, Lundin P, Månér S, Fjellidal R, Naume B, Wigler M, Hicks J, Børresen-Dale A-L, Zetterberg A, Russnes HG: Quantitative multi-gene FISH on breast carcinomas identifies der(1;16)(q10;p10) as an early event in luminal A tumors. *Genes Chromosomes Cancer* 2015, 54:235–248
 50. Tirkkonen M, Tanner M, Karhu R, Kallioniemi A, Isola J, Kallioniemi OP: Molecular cytogenetics of primary breast cancer by CGH. *Genes Chromosomes Cancer* 1998, 21:177–184
 51. Teixeira MR, Pandis N, Heim S: Cytogenetic clues to breast carcinogenesis. *Genes Chromosomes Cancer* 2002, 33:1–16
 52. Rennstam K, Ahlstedt-Soini M, Baldetorp B, Bendahl P-O, Borg A, Karhu R, Tanner M, Tirkkonen M, Isola J: Patterns of chromosomal imbalances defines subgroups of breast cancer with distinct clinical features and prognosis: a study of 305 tumors by comparative genomic hybridization. *Cancer Res* 2003, 63:8861–8868
 53. Baudis M: Genomic imbalances in 5918 malignant epithelial tumors: an explorative meta-analysis of chromosomal CGH data. *BMC Cancer* 2007, 7:226
 54. Korsching E, Packeisen J, Helms MW, Kersting C, Voss R, van Diest PJ, Brandt B, van der Wall E, Boecker W, Bürger H: Deciphering a subgroup of breast carcinomas with putative progression of grade during carcinogenesis revealed by comparative genomic hybridisation (CGH) and immunohistochemistry. *Br J Cancer* 2004, 90:1422–1428
 55. André F, Job B, Dessen P, Tordai A, Michiels S, Liedtke C, Richon C, Yan K, Wang B, Vassal G, Delaloge S, Hortobagyi GN, Symmans WF, Lazar V, Pusztai L: Molecular characterization of breast cancer with high-resolution oligonucleotide comparative genomic hybridization array. *Clin Cancer Res* 2009, 15:441–451
 56. Natrajan R, Weigelt B, Mackay A, Geyer FC, Grigoriadis A, Tan DSP, Jones C, Lord CJ, Vatcheva R, Rodríguez-Pinilla SM, Palacios J, Ashworth A, Reis-Filho JS: An integrative genomic and transcriptomic analysis reveals molecular pathways and networks regulated by copy number aberrations in basal-like, HER2 and luminal cancers. *Breast Cancer Res Treat* 2010, 121:575–589
 57. Vollan H-KM, Rueda OM, Chin S-F, Curtis C, Turashvili G, Shah S, Lingjærde OC, Yuan Y, Ng CK, Dunning MJ, Dicks E,

- Provenzano E, Sammut S, McKinney S, Ellis IO, Pinder S, Purushotham A, Murphy LC, Kristensen VN, Brenton JD, Pharoah PDP, Børresen-Dale A-L, Aparicio S, Caldas C: A tumor DNA complex aberration index is an independent predictor of survival in breast and ovarian cancer. *Mol Oncol* 2015, 9:115–127
58. Morganelle S, Alexandrov LB, Glodzik D, Zou X, Davies H, Staaf J, Sieuwerts AM, Brinkman AB, Martin S, Ramakrishna M, Butler A, Kim H-Y, Borg A, Sotiriou C, Futreal PA, Campbell PJ, Span PN, Van Laere S, Lakhani SR, Eyfjord JE, Thompson AM, Stunnenberg HG, van de Vijver MJ, Martens JWM, Børresen-Dale A-L, Richardson AL, Kong G, Thomas G, Sale J, Rada C, Stratton MR, Birney E, Nik-Zainal S: The topography of mutational processes in breast cancer genomes. *Nat Commun* 2016, 7:11383
59. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al: Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 2016, 534:47–54
60. Stephens PJ, McBride DJ, Lin M-L, Varela I, Pleasance ED, Simpson JT, Stebbings LA, Leroy C, Edkins S, Mudie LJ, Greenman CD, Jia M, Latimer C, Teague JW, Lau KW, Burton J, Quail MA, Swerdlow H, Churcher C, Natrajan R, Sieuwerts AM, Martens JWM, Silver DP, Langerød A, Russnes HEG, Foekens JA, Reis-Filho JS, van 't Veer L, Richardson AL, Børresen-Dale A-L, Campbell PJ, Futreal PA, Stratton MR: Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* 2009, 462:1005–1010
61. Ali HR, Rueda OM, Chin S-F, Curtis C, Dunning MJ, Aparicio SA, Caldas C: Genome-driven integrated classification of breast cancer validated in over 7,500 samples. *Genome Biol* 2014, 15:431
62. Holland DG, Burleigh A, Git A, Goldgraben MA, Perez-Mancera PA, Chin S-F, Hurtado A, Bruna A, Ali HR, Greenwood W, Dunning MJ, Samarajiwa S, Menon S, Rueda OM, Lynch AG, McKinney S, Ellis IO, Eaves CJ, Carroll JS, Curtis C, Aparicio S, Caldas C: ZNF703 is a common luminal B breast cancer oncogene that differentially regulates luminal and basal progenitors in human mammary epithelium. *EMBO Mol Med* 2011, 3:167–180
63. Sircoulomb F, Nicolas N, Ferrari A, Finetti P, Bekhouche I, Rousselet E, Lonigro A, Adélaïde J, Baudet E, Esteyriès S, Wicinski J, Audebert S, Charafe-Jauffret E, Jacquemier J, Lopez M, Borg J-P, Sotiriou C, Popovici C, Bertucci F, Birnbaum D, Chaffanet M, Ginestier C: ZNF703 gene amplification at 8p12 specifies luminal B breast cancer. *EMBO Mol Med* 2011, 3:153–166
64. Cancer Genome Atlas Network: Comprehensive molecular portraits of human breast tumours. *Nature* 2012, 490:61–70
65. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, et al: The landscape of cancer genes and mutational processes in breast cancer. *Nature* 2012, 486:400–404
66. Pereira B, Chin S-F, Rueda OM, Vollen H-KM, Provenzano E, Bardwell HA, Pugh M, Jones L, Russell R, Sammut S-J, Tsui DWY, Liu B, Dawson S-J, Abraham J, Northen H, Peden JF, Mukherjee A, Turashvili G, Green AR, McKinney S, Oloumi A, Shah S, Rosenfeld N, Murphy L, Bentley DR, Ellis IO, Purushotham A, Pinder SE, Børresen-Dale A-L, Earl HM, Pharoah PD, Ross MT, Aparicio S, Caldas C: The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun* 2016, 7:11479
67. Silwal-Pandit L, Vollen H-KM, Chin S-F, Rueda OM, McKinney S, Osako T, Quigley DA, Kristensen VN, Aparicio S, Børresen-Dale A-L, Caldas C, Langerød A: TP53 mutation spectrum in breast cancer is subtype specific and has distinct prognostic relevance. *Clin Cancer Res* 2014, 20:3569–3580
68. Hanker AB, Pfefferle AD, Balko JM, Kuba MG, Young CD, Sánchez V, Sutton CR, Cheng H, Perou CM, Zhao JJ, Cook RS, Arteaga CL: Mutant PIK3CA accelerates HER2-driven transgenic mammary tumors and induces resistance to combinations of anti-HER2 therapies. *Proc Natl Acad Sci U S A* 2013, 110:14372–14377
69. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Børresen-Dale A-L, Boyault S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjord JE, Foekens JA, Greaves M, Hosoda F, Hutter B, Illic T, Imbeaud S, Imielinski M, Imielinski M, Jäger N, Jones DTW, Jones D, Knappskog S, Kool M, Lakhani SR, López-Otín C, Martin S, Munshi NC, Nakamura H, Northcott PA, Pajic M, Papaemmanuil E, Paradiso A, Pearson JV, Puente XS, Raine K, Ramakrishna M, Richardson AL, Richter J, Rosenstiel P, Schlesner M, Schumacher TN, Span PN, Teague JW, Totoki Y, Tutt ANJ, Valdés-Mas R, van Buuren MM, van 't Veer L, Vincent-Salomon A, Waddell N, Yates LR; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MML-Seq Consortium; ICGC PedBrain, Zucman-Rossi J, Futreal PA, McDermott U, Lichter P, Meyerson M, Grimmond SM, Siebert R, Campo E, Shibata T, Pfister SM, Campbell PJ, Stratton MR: Signatures of mutational processes in human cancer. *Nature* 2013, 500:415–421
70. Sternemalm J, Russnes HG, Zhao X, Risberg B, Nord S, Caldas C, Børresen-Dale AL, Stokke T, Patzke S: Nuclear CSPP1 expression defined subtypes of basal-like breast cancer. *Br J Cancer* 2014, 111:326–338
71. Bartlett JMS, Bayani J, Marshall A, Dunn JA, Campbell A, Cunningham C, Sobol MS, Hall PS, Poole CJ, Cameron DA, Earl HM, Rea DW, Macpherson IR, Canney P, Francis A, McCabe C, Pinder SE, Hughes-Davies L, Makris A, Stein RC; OPTIMA TMG: Comparing breast cancer multiparameter tests in the OPTIMA preliminary trial: no test is More equal than the others. *J Natl Cancer Inst* 2016, 108:djw050
72. Petersen OW, Polyak K: Stem cells in the human breast. *Cold Spring Harb Perspect Biol* 2010, 2:a003160