Nature Reviews Cancer | AOP, published online 29 December 2008; doi:10.1038/nrc2571

PERSPECTIVES

OPINION

Do 'basal-like' breast cancers really exist?

Barry Gusterson

Abstract | It has been proposed that gene expression profiles will revolutionize the classification of breast cancer, eventually replacing histopathology with a more reproducible technology. These new approaches, combined with a better understanding of the cellular origins of breast cancer, should enable us to identify patient subgroups for more effective therapy. However, in such a rapidly advancing field it is essential that initial and thought-provoking results do not become established as 'facts' without question. This Opinion addresses some of the negatives and positives generated by the term 'basal-like' breast cancer, and questions its existence as an entity.

The classification of breast cancer can be approached from two main standpoints. The first is based on the histological appearance of the cancer in combination with a grading system. Grading combines proliferation rate, degree of differentiation (assessed by tubule formation) and nuclear pleomorphism. Thus, grading is a morphological surrogate for the multiple genetic and epigenetic changes that occur in cancer progression. The combination of grading, tumour size and lymph node status identifies prognostic groupings[1]. Although these broad categories provide limited information for individual patient management, several clearly definable and reproducible subtypes are evident (FIG. 1). The second approach is to classify breast cancers on the basis of a marker that can predict a probable response to a particular treatment, such as expression of hormone receptors and ERBB2. Both morphology-based and protein expression-based tests, when applied with appropriate quality control, are reproducible and have both prognostic and predictive power to direct patient management. Indeed, in November 2000 the National Institutes of Health, having reviewed all evidence-based parameters that might predict the use of adjuvant chemotherapy, produced a consensus statement that only patient age, axillary lymph node status, histological tumour type, standardized pathological grading and the expression status of hormone receptors had proven clinical relevance[2]. The 2003 St Gallen Consensus took this further and stressed the importance of quality control in assessing hormone responsiveness and of recognizing endocrine non-responsive invasive and *in situ* breast cancers[3].

At least 15% of breast cancers lack the expression of oestrogen receptors (ERs), progesterone receptors (PRs) and ERBB2, and are referred to as triple-negative cancers (TNCs). TNCs generally have a poor prognosis and are a heterogenous group of cancers including poorly differentiated adenocarcinomas (some of which are BRCA1-associated cancers[4]) and clearly defined rare types of breast cancer[5] that range from those with an excellent prognosis to aggressive metaplastic carcinomas. Given the heterogeneity of TNCs there is a clear clinical need to identify predictive markers and novel treatments that will improve patient treatment[4].

With the present limited ability to predict individual patient response to chemotherapy, a rational approach is to assess those genes expressed in individual tumours in order to identify cancer groupings related to treatment outcome. However, the ability to identify subgroups of cancers on the basis of gene expression profiling (GEP) and to demonstrate that they correlate with different survival rates has not, as yet, provided any significant advances over current clinico-pathological practice[6]. Indeed, a recent meta-analysis of publicly available breast cancer gene expression studies showed that the three key biological drivers in nine prognostic signatures are proliferation, ER signalling and ERBB2 amplification[7]. It seems that a lack of quality control, poor definitions and lack of reproducibility, which plagued breast histopathology in the past, are now major issues for those using molecular and protein profiling. Moreover, owing to a lack of standardized technology a mountain of misinformation has been generated in the rush to identify breast cancer subgroups. In this context, the recent conclusion that the gold standard for identifying breast cancer subtypes remains gene expression array analysis[8] seems somewhat premature.

In this Opinion article I argue that the identification of basal-like breast cancers, which form the majority of TNCs, on the basis of GEP data has been misleading in some respects and that a clear, basic understanding of breast cancer biology is needed to fully interpret GEP data in order to improve the treatment of patients with TNCs.

Normal breast and breast cancer

By 28 weeks *in utero* the human breast is composed of a branching ductal structure with two cell layers: inner luminal cells that line the ducts and produce milk at lactation and an outer layer that sits on the basement membrane and contracts in response to oxytocin[9]. This outer layer has many features of smooth muscle, giving it its classical name — the myoepithelium. The branching ductal structure ends in clusters of small ductules that form the terminal duct lobular units (TDLU) (FIG. 2). It is established that breast cancers, with rare exceptions, arise in the TDLU, not in the ducts[10], making the historical terminology of 'ductal' and 'lobular' breast cancer confusing and incorrect. How the luminal and myoepithelial lineages are maintained is largely unknown, but

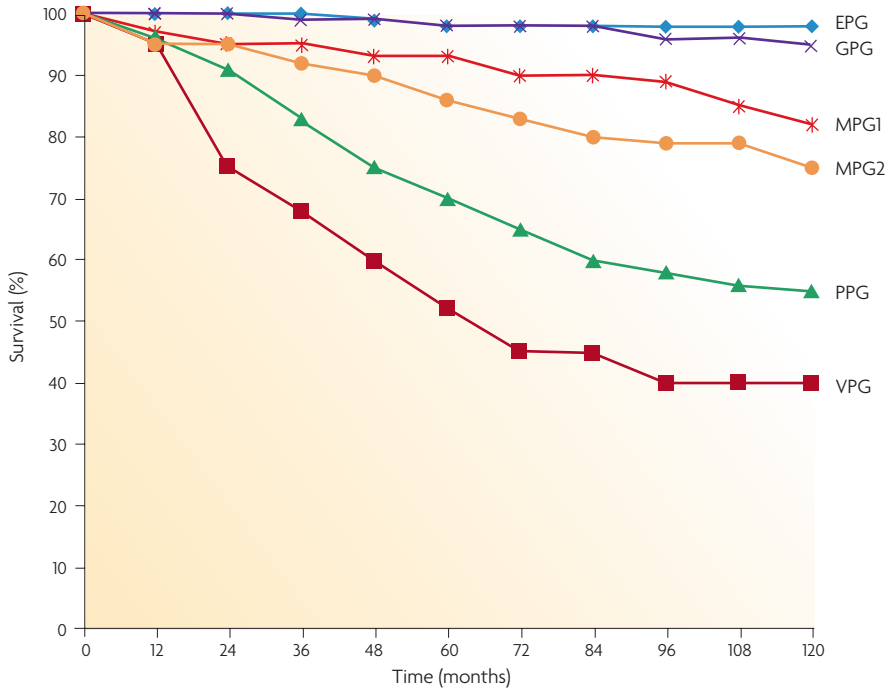

Figure 1 | **Breast cancer-specific survival (log–rank).** Data were generated using the Nottingham Prognostic Index on a series of patients assessed between 1990 and 1999. EPG, excellent prognostic group; GPG, good prognostic group; MPG1, moderate prognostic group 1; MPG2, moderate prognostic group 2; PPG, poor prognostic group; VPG, very poor prognostic group. Adapted from REF. 1 © Elsevier Ltd (2007).

recent work indicates that Notch signalling has a key role in luminal cell commitment[11] and that expression of GATA3 regulates luminal differentiation[12]. These results suggest the presence of epithelial hierarchies in the mammary gland.

Recent advances in mouse mammary and human breast stem cell biology have identified pluripotent cells with similar characteristics in both species. Such uncommitted stem cells can form an entire mammary gland[13], and elegant flow sorting combined with mammary gland reconstitution studies in the mouse[14,15] has demonstrated that mammary stem cells have markers of cells in the basal compartment (from here onwards in this article 'basal' is only used to define cells that are adjacent to the basement membrane in human breast tissue or the mouse mammary gland). These stem cells express cytokeratin 14, p63 and the epidermal growth factor receptor (EGFR), and lack expression of cytokeratin 18, ERBB2, ERα and PR[15,16]. In human mammosphere studies, progenitor cells with high proliferation potential also lacked expression of cytokeratin 8 or 18, α-smooth muscle actin (αSMA), the luminal cell marker mucin 1 (MUC1) and ER[17]. These cells were positive for cytokeratin 5 and were responsive to EGF. Earlier work identified a suprabasal population (sitting on the myoepithelial layer, but not reaching the lumen) that was positive for epithelial specific antigen and negative for MUC1 (REF. 18). These cells were αSMA-negative and pluripotent, giving rise to cells characterized as luminal and myoepithelial. The relationship of these cells to the stem cells that have been identified by others is unknown. The recent demonstration of a population of undifferentiated cells in humans that are positive for CD44 and negative for CD10 (a myoepithelial marker; also known as neprilysin) indicates that the results in mouse models are likely to be applicable to the human breast[19]. Thus, in both species a similar population of undifferentiated cells has been identified that has stem cell characteristics and that may reside in the basal layer. Although there is evidence that some tumours contain a minority of cells that have normal stem cell-like characteristics, there is no direct evidence for or against the origin of cancers from these cells.

In theory, the heterogeneity of human breast cancers could reflect differences in the target cell population and/or different combinations of oncogene activation and loss of tumour suppressor gene function in a common breast stem cell or committed progenitor. We know that the majority of human cancers express proteins that would suggest a luminal origin, but how the many clearly defined histopathological subtypes[20] and molecular subtypes of breast cancers arise remains the subject of speculation[21]. Some human breast cancers have been clearly shown to have myoepithelial characteristics, but such cancers are so rare that they are the subject of clinical case reports. However, it should be noted that there is a group of metaplastic breast cancers that have been described as frequently having myoepithelial markers and thus may take their origin from a committed myoepithelial cell or a pluripotent cell[22].

Cells that have similar properties to stem cells have been identified in human breast cancers. These cancer stem cells (CSCs) are $CD44^{+}$ and $CD24^{-/low}$, are highly tumorigenic at low cell inoculum in immunodeficient mice[23] and can self-renew. Gene signature studies of isolated $CD44^{+}$ cells from human breast cancers have shown that these cells express genes that are known stem cell markers[19]. Moreover, poorly differentiated adenocarcinomas of the breast overexpress *NANOG*, *OCT4*, *SOX2* and *MYC* (genes associated with embryonic stem cells); further evidence of stem cell-like features in some breast cancers[24].

Keeping in mind that we do not fully understand the cellular origins of breast cancer, there are many references in the literature to the molecular subclassification of breast cancer, with groupings into basal, normal, luminal A and luminal B, with the inference that this classification reflects histogenesis[25]. However, when histological data are compared with molecular data, some discrepancies are evident. For example, the original landmark publication in 2000 on the use of cDNA microarrays to produce a molecular portrait of breast cancers incorrectly stated that basal (and/or myoepithelial) cells and luminal epithelial cells are distinguished on the basis of the expression of cytokeratin 5/6 in basal cells and cytokeratin 8 or 18 in luminal cells[26]. However, the original paper by Nagle that was quoted to support this discrimination clearly states that the luminal cells in the ductules of the human breast express cytokeratin 5 (REF. 27), and numerous subsequent papers have shown the luminal expression of cytokeratins 5/6 (see REF. 21 for a review). Despite this, many subsequent publications have perpetuated the statement that cytokeratins 5 and 14 define basal and myoepithelial cells[28], and the term 'basal-like' has been used for tumours that express these cytokeratins.

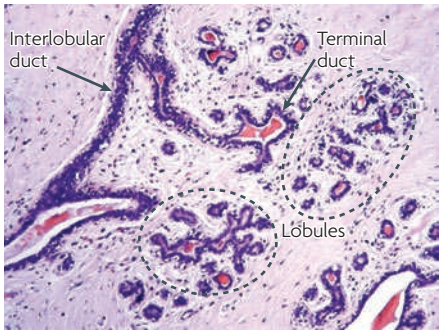

Figure 2 | **Normal breast tissue.** The image shows a haematoxylin and eosin-stained section of a normal human adult breast with a mature branching network. Note the interlobular duct branches to the terminal ducts. The terminal ducts together with the lobules form the terminal duct lobular unit. In three dimensions all parts of the lobule are linked in a structure analogous to a bunch of grapes.

Adding to this confusion is the use of the term basal-like, which has been used to characterize a heterogeneous class of breast cancers. To avoid further confusion when discussing these data, from here on in the term basal-like is only used in the context of quoting other authors' work. Moreover, the terms basal and myoepithelium tend to be used synonymously in publications, but it would appear that there is a rare subpopulation of cells in the mouse mammary gland[14,19] that have stem cell characteristics and that are likely to be basally positioned, but are not myoepithelial cells. For the remainder of this article the term myoepithelium is restricted to cells that are basal and have expression of αSMA and CD10. It should be noted that reference to expression of cytokeratins 5/6 is based on the use in these publications of an antibody (clone D5/16B4) that recognizes both proteins.

Expression profiling and terminology

Perou[26] and Sørlie[25] demonstrated using cDNA microarray studies that an 'intrinsic gene set' could segregate breast cancers into five main subgroups, one of which they equated to the gene expression of 'basal' cells in the normal breast as discussed above. This basal-like group of breast cancers comprise 80–90% of the TNC group[5]. Some have now equated the two groups[29], although others argue that it makes no sense to group together such a heterogeneous collection of cancers[30,31]. The Perou paper[26], based on 38 invasive cancers of which 6 had basal-like features, is at the base of an exponentially increasing number of papers on this 'newly defined cancer type' (FIG. 3). However, the known facts about the cellular origins of breast cancer question whether there is a definable subtype of basal-like breast cancer that has any clinico-pathological significance. To support a new classification of breast cancers into basal, luminal A, luminal B, luminal C and normal, as proposed by Sørlie, there needs to be good evidence that the tumours do in fact arise from cells in these compartments. Unfortunately this is not the case. A recent review by Moinfar[32] provides a critical and excellent analysis of the papers by Perou[26] and Sørlie[25]. Having analysed in detail all of the data presented, including the supplementary information, his findings support the conclusion that there is little evidence to support the luminal-like or a basal-like nomenclature using hierarchical clustering. In this analysis, Moinfar points out that the basal-like subtype was characterized by cytokeratin 5 and cytokeratin 17, whereas many immunohistochemical studies use cytokeratin 14 to define basal-like cancers. Also cytokeratin 7 and cytokeratin 13 were used to define a basal epithelial cell-enriched gene cluster, although neither of these proteins are basal cytokeratins. Both of these studies were of insufficient size for reliable statistical analyses.

Origins and use of the term basal-like

The term basal originated in the now classic papers of Moll, reviewed in REF. 21. In many stratified epithelia, the cells sitting on the basement membrane in a basal position express cytokeratin 14, cytokeratin 17 and cytokeratin 5. What has apparently not been realized by many groups is that, in contrast to the mouse, these basal cytokeratins are also expressed in luminal cells in the TDLU in the human breast (FIG. 4; reviewed in REF. 21). Clearly this is important because, as stated above, a body of evidence points to the vast majority of breast cancer arising in the luminal cells of the TDLU and not from the myoepithelial cells. If the expression of cytokeratin 14, cytokeratin 5 or cytokeratin 17 is thus used as a criterion for any classification, one would have to conclude that these tumours have a profile that suggests they are of TDLU luminal origin not of myoepithelial origin. This is further supported by the fact that the cytokeratin 8, cytokeratin 18 and cytokeratin 19 are expressed in luminal cells, but have never been described in myoepithelial cells. Indeed, in 1998, Moll showed that 100% of breast cancers of all grades express cytokeratin 18 and up to 20% express cytokeratin 14 and this increases markedly in aggressively growing tumours[33]. Immunohistochemical studies have also shown that 10% or fewer of cells positive for cytokeratins 5/6 and 17 (even single-positive cells in a few cases) have been used to classify a tumour as basal-like[32]. In their 2006 paper[34] the Perou group stated that the myoepithelial markers CD10, p63 and αSMA were infrequently expressed in basal-like breast cancer and concluded that their results do not support a myoepithelial derivation for basal-like breast cancer.

Previous publications from Diarkee[35] and Malzahn[33] described that a subgroup of breast cancers expressed cytokeratin 4, cytokeratin 14 and/or cytokeratin 17 and that this was associated with a poor prognosis. The Malzahn paper also raised the possibility that the expression of stratified epithelial cytokeratins (such as cytokeratin 4, cytokeratin 14 and cytokeratin 17) may represent rudimentary squamous metaplasia[33]. Using immunohistochemistry on 21 cases of basal-like breast cancer as defined by gene set analyses, Neilsen and colleagues showed that 76% of the tumours could be classified as basal-like on the basis of ER and ERBB2 negativity in conjunction with expression of cytokeratin 5/6 and/or EGFR[36]. Importantly, any level of staining for cytokeratin 5/6 and EGFR was considered as a positive result. In a more recent publication, five biomarkers (ER, PR, ERBB2, EGFR and cytokeratin 5/6) were assessed using a breast cancer tissue array[37]. Using 3,744 interpretable samples, a basal-like classification was made for tumours with a negative result for cytokeratin 5/6. Finally, Carey and colleagues have equated ER-negative and ERBB2-negative staining as defining the basal-like group, eliminating EGFR and cytokeratin 5/6 as necessary criteria to define basal-like disease[29].

These contradictory findings raise considerable issues. It seems that different antibodies and staining techniques can convert a positive result to a negative result and vice versa, and this of course would present serious validation problems if such biomarkers were to be introduced into the clinic. Moreover, reliance on a negative result for any classification is a difficult quality control issue. There is also cause for concern in terms of whether or not data based on tissue arrays are appropriate given the degree of heterogeneity of these proteins in the tumour samples. This is a particular problem for ER, PR and ERBB2, for which the agreed cut-off points for staining intensity or percentage of positive cells in order to stratify patients for treatment have not been validated on tissue arrays or received international approval.

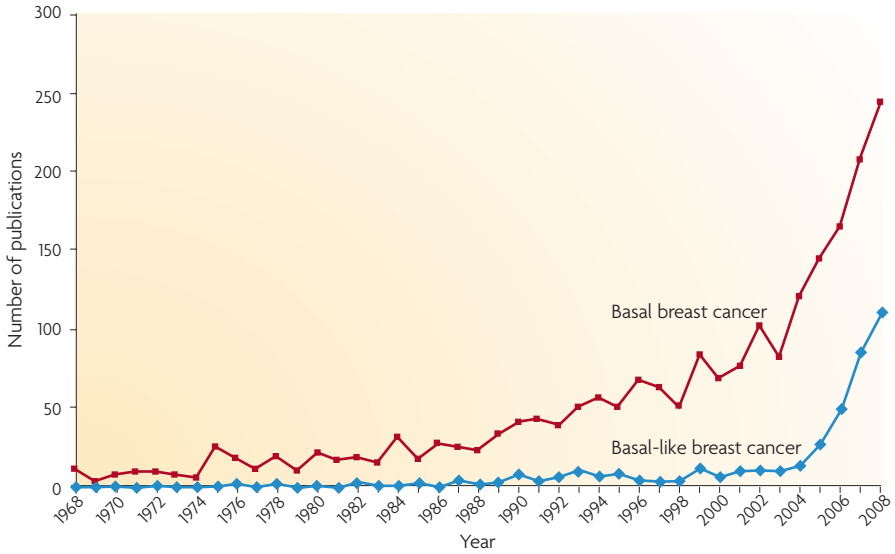

Figure 3 | **The rise of the term basal-like.** These data are taken from PubMed and demonstrate the number of publications per annum that are identified using the terms 'basal', 'breast' and 'cancer' or 'basal-like', 'breast' and 'cancer'.

Overall, at a morphological level, the term basal-like breast cancers seems to cover tumours that appear to arise in the luminal epithelium, can produce glandular structures, express mucin and express either cytokeratin 8 or cytokeratin 18 in 100% of cases. These cytokeratins have only been described in luminal cells to date. There is currently no reason to believe that these tumours arise from myoepithelial cells or from the basal layer. As discussed earlier, there is a population of cells in the basal layer that are distinguishable from myoepithelial cells and that may be the target for carcinogenesis, but if markers are an indication of lineage origin these cells would not be good candidates for basal-like breast cancers or TNCs.

Basal-like, BRCA1 and metaplasia?

Even if a tumour has a protein expression profile that is similar to that of some normal cells, is this evidence that the tumour arose from cells of this differentiation profile or is this the malignant equivalent of convergent evolution? The histogenesis implications of the Perou papers have resulted in lineage hypothesis models that try to explain the *in vivo* and *in vitro* data generated[38–40]. However, as stated by Stingl and Caldos[41], these novel molecular signatures may not have a normal cellular equivalent, but are patterns that have evolved during tumour progression. Therefore, the expressions of cytokeratin 5, cytokeratin 14, cytokeratin 17 and EGFR may be not drivers but passengers in this progression[39]. The decision by some groups to drop the original criterion of increased expression of cytokeratin 5/6, cytokeratin 14 and cytokeratin 17 in the latest classification realignment of basal-like breast cancer may be overlooking an important piece of relevant biology.

It is well known that the breast epithelium arises from the overlying skin and that it is a common finding for the breast epithelium to undergo metaplasia (a change in state of differentiation from one specialist type of epithelium to another). This is seen most commonly as a change to a skin-like structure (squamous metaplasia) or apocrine metaplasia. Metaplastic carcinomas are a recognized histopathological subgroup of breast cancers and fall within the definition of 'basal-like' breast cancer[42]. In squamous metaplasia there is an associated increase in cytokeratin 5, cytokeratin 14 and cytokeratin 17. In addition, normal squamous epithelium has a higher level of EGFR expression than normal breast epithelium, and this is reflected in metaplastic epithelium undergoing squamous change. Expression of cytokeratin 5/6 or cytokeratin 14 is used to confirm squamous differentiation in breast cancer, where the areas with cells positive for these proteins are associated with increased expression levels of EGFR[43]. Single-cell expression of cytokeratin 5 or cytokeratin 14 might therefore represent a 'forme fruste' of squamous change. This view is consistent with the earlier work of Moll, who suggested that breast cancers that express cytokeratin 5 or cytokeratin 14 were undergoing squamous metaplasia[33].

The initiating events of squamous metaplasia are not clear. In a recent study, Ince and colleagues demonstrated that human mammary epithelial cells that had an expression profile similar to myoepithelial cells produced tumours with squamous metaplasia, whereas human breast epithelial cells that had luminal characteristics produced adenocarcinomas[44]. As many of the original myoepithelial markers were retained in these squamous tumours, this would argue that myoepithelial cells have an increased propensity for squamous metaplasia. Squamous metaplasia, however, is a common finding in rodent tumours and is a feature of activation of the Wnt–β-catenin pathway. Stabilization of β-catenin in mammary cells causes loss of differentiation followed by transdifferentiation to squamous epithelium[45]. In transgenic mice in which the WNT1 pathways have been activated, mammary tumours show squamous metaplasia and pilar differentiation[46]. Thus, the squamous changes seen in some basal-like carcinomas may be due to activation of these pathways.

Data from models of BRCA1 mutation-associated tumours demonstrate a link between mutation of this gene and expression of EGFR and cytokeratin 14. Two groups have independently generated conditional BRCA1 knockouts. In one case the activation is in β-lactoglobulin (luminal cells)[47] and in the other it is in cells positive for cytokeratin 14 (REF. 48). It should be noted that, in the mouse, cytokeratin 14 expression at the protein level is specific for the basal cell population, in contrast to the dual luminal and basal expression in the human breast. In both models, the majority of the tumours lack hormone receptors and ERBB2 expression, and express cytokeratin 14 and EGFR. This would indicate that, in this particular instance, the cellular target is less important than the transforming event. This raises the question of how BRCA1 can produce this squamous-like change. Also, owing to the similarity of this expression pattern to that of mammary stem cells, it raises the question of whether BRCA1 is a stem cell regulator.

It is known that BRCA1 is crucial to normal mammary gland development[49] and that it can directly modulate expression of ERα[50]. Recent work also indicates that BRCA1 is required for the conversion of ER-negative cells to ER-positive cells, with the proposal that BRCA1 regulates progenitor cell fate[51]. This also fits with the hypothesis that BRCA1 is involved in stem cell regulation[52]. It is thus possible that loss of BRCA1 can affect the normal differentiation in the mammary gland with a reversion to a potentially default squamous pattern. It would be of interest to know the BRCA1 status of the

cells in metaplastic breast cancers and whether expressing BRCA1 in keratinocytes under appropriate hormone supplements induces glandular differentiation.

Clinical significance. In 2001, Hedenfalk demonstrated that expression profiling could identify a gene set that would specifically recognize breast cancers that had a mutation in *BRCA1* (REF. 53). It was predicted in this paper that cDNA microarrays would identify sporadic breast cancers with a phenotype that resembles tumours arising in patients with germline *BRCA1* mutations. The majority of *BRCA1* mutation-associated breast cancers are TNCs and are associated with increased expression of cytokeratins 5/6 (REF. 54) and EGFR[36,55]. BRCA1-mutant tumours also have a high incidence of *TP53* gene mutations[56]. Using the Breast Cancer Linkage Consortium data set, Lakhani has clearly demonstrated the similarities between BRCA1-associated breast cancers and basal-like breast cancers that have a high incidence of cytokeratin 5/6 and cytokeratin 14 expression and hormone receptor negativity[55].

Owing to some of the similarities between the morphology and the proteins expressed in familial BRCA1 breast cancers and sporadic basal-like breast cancers, many have asked whether these tumours share characteristics that make them amenable to similar treatment protocols. Recently, using Nielsen's criteria, Ellis's laboratory has demonstrated a strong association between loss of BRCA1 nuclear expression and basal-like carcinomas[57]. The similarity between *BRCA1* mutation-associated tumours and basal-like tumours (or TNCs) suggests that there may be a dysfunction of BRCA1 in a proportion of these tumours that could lead to new therapeutic approaches[58].

The role of BRCA1 as a regulator of the DNA damage response indicates that *BRCA1*-mutated tumours should have greater sensitivity to a DNA crosslinking agent such as platinum than to a spindle poison such as docetaxel. In fact, there is evidence that loss of BRCA1 may lead to resistance to spindle poisons[59]. In this context, the current Triple Negative Trial (TNT)[60] is aimed at dissecting the response of patients with TNC to therapies with a view to placing them in clinically meaningful subgroups and then to develop robust markers that define these populations to assist in the design of future trials with biomarkers and predictive indicators of response.

As recently reviewed, there are a number of possible novel targets emerging for the TNC[42]. Based on the increased levels of EGFR and KIT in these tumours[36], it has been suggested that these proteins may also be therapeutic targets. However, as EGFR is part of the differentiation phenotype and in other tumours there is no correlation of response to inhibitors and levels of expression, it is unlikely that increased expression of EGFR will be useful as a therapeutic target. Similarly, the absence of activating KIT mutations in TNCs indicates that this is an equally poor prospect.

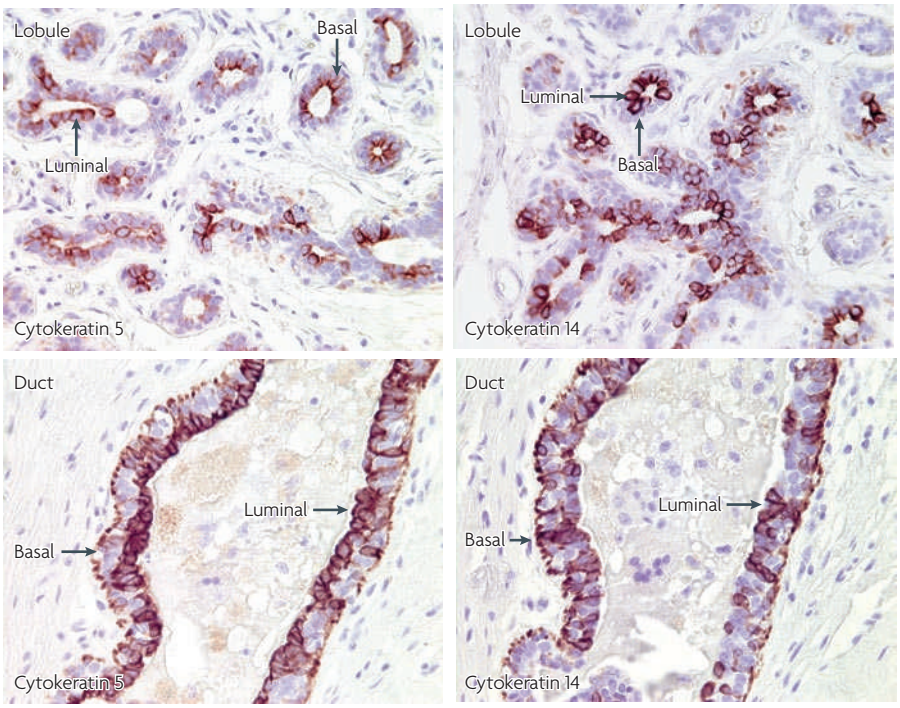

Figure 4 | **Cytokeratin expression in the human breast.** Cytokeratin 5 (CK5) and CK14 show a similar pattern of expression in both ducts and lobules and can stain both luminal cells and myoepithelial cells. It should be noted that in some areas of the same breast that there were no cells that are positive for these proteins in the basal or luminal cells in lobules or ducts. The pattern is heterogeneous, both in the same breast and between breasts.

Conclusions and future developments

The term basal-like breast cancer has been shown to be misleading and the use of molecular profiling is clearly not currently a 'gold standard' for treatment decision making or for useful classification. As recently stated there are considerable problems in the adaptation of gene expression analysis to the routine setting[8]. Only an international committee of unbiased experts can designate gold standards and these must be based on rigorous criteria. There is no justification for perpetuating incorrect terminology and, until we have identifiers of clinically relevant breast cancer subgroups using markers or signatures that have rigorous quality control, we should resist the temptation to re-classify breast cancer. The application of signatures that start with a biological question, such as the analysis of cancer stem cells[19,23,28] and endocrine resistance[50], are more likely to have a clinical impact than prognostic signature approaches.

Many prognostic gene signatures have been reported that, although sharing few common genes, segregate breast cancers into broadly similar subgroups, which indicates that they have shared biological phenotypes[61]. It is of interest that recent papers have demonstrated that prognostic gene signatures have as their primary discriminators the expression of genes related to proliferation[62]. The other key drivers are ER and ERBB2 expression[7,63], so, in effect, we have come full circle.

In some countries the reason for developing signatures to replace current assays for ER, ERBB2 and pathology grading is because of poor quality control in some pathology laboratories. There is no evidence that the quality control with molecular techniques is any better than that provided by an excellent and dedicated breast pathologist. Perhaps some of the funding going to molecular profiling should be invested in development of international standards and external quality assurance systems in routine pathology.

Barry Gusterson is at the Division of Cancer Sciences and Molecular Pathology, University of Glasgow, Western Infirmary (Pathology), Dumbarton Road, Glasgow G11 6NT, Scotland, UK.
e-mail: B.A.Gusterson@clinmed.gla.ac.uk

doi:10.1038/nrc2571
Published online 29 December 2008

1. Blamey, R. W. *et al.* Survival of invasive breast cancer according to the Nottingham Prognostic Index in cases diagnosed in 1990–1999 *Eur. J. Cancer* **43**, 1548–1555 (2007).
2. National Institutes of Health Consensus Development Conference statement: adjuvant therapy for breast cancer, November 1–3, 2000. *J. Natl Cancer Inst. Monogr.* **93**, 5–15 (2001).
3. Goldhirsh, A. *et al.* Meeting highlights: updated international expert consensus on the primary therapy of early breast cancer. *J. Clin. Oncol.* **21**, 3357–3365 (2003).
4. Reis-Filho, J. S. & Tutt, A. N. Triple negative tumours: a critical review. *Histopathology* **52**, 108–118 (2008).
5. Reis-Filho, J. S. & Lakhani, S. R. Breast cancer special types: why bother? *J. Pathol.* **216**, 394–398 (2008).
6. Sotiriou, C. & Piccart, M. J. Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? *Nature Rev. Cancer* **7**, 545–553 (2007).
7. Wirapati, P. *et al.* Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res.* **10**, R65 (2008).
8. Peppercorn, J., Perou, C. M. & Carey, L. A. Molecular subtypes in breast cancer evaluation and management: divide and conquer. *Cancer Invest.* **26**, 1–10 (2008).
9. Anbazhagan, R. *et al.* The development of epithelial phenotypes in the human fetal and infant breast. *J. Pathol.* **184**, 197–206 (1998).
10. Wellings, S. R. & Jensen, H. M. On the origin and progression of ductal carcinoma in the human breast. *J. Natl Cancer Inst.* **50**, 1111–1118 (1973).
11. Bouras, T. *et al.* Notch signaling regulates mammary stem cell function and luminal cell-fate commitment. *Cell Stem Cell* **3**, 429–441 (2008).
12. Asselin-Labat, M. L. *et al.* Gata-3 is an essential regulator of mammary-gland morphogenesis and liminal-cell differetiation. *Nature Cell Biol.* **9**, 201–209 (2007).
13. Kordon, E. C., Smith, G. H. An entire functional mammary gland may comprise the progeny from a single cell. *Development* **125**, 1921–1930 (1998).
14. Stingl, J. *et al.* Purification and unique properties of mammary epithelial stem calls. *Nature* **439**, 993–997 (2006).
15. Shackleton, M. *et al.* Generation of a functional mammary gland from a single stem cell. *Nature* **439**, 84–88 (2006).
16. Asselin-Labat, M.-L. *et al.* Steroid hormone receptor status of mouse mammary stem cells. *J. Natl Cancer Inst.* **98**, 1011–1014 (2006).
17. Dontu, G. *et al. In vitro* propagation and transcriptional prfiling of human mammary stem/progenitor cells. *Genes Dev.* **17**, 1253–1270 (2003).
18. Gudjonsson, T. *et al.* Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties. *Genes Dev.* **16**, 693–706 (2002).
19. Shipitsin, M. *et al.* Molecular definition of breast tumor heterogeneity. *Cancer Cell* **11**, 259–273 (2007).
20. Rosen, P. P. *Rosen's Breast Pathology*. 3rd edn (ed. Pine, J. W.) (Lippincott Williams and Wilkins, Philadelphia, 2008).
21. Gusterson, B. A., Ross, D. T., Heath, V. J. & Stein, T. Basal cytokeratins and their relationship to the cellular origin and functional classification of breast cancer. *Breast Cancer Res.* **7**, 143–148 (2005).
22. Leibl, S., Gogg-Kammerer, M., Sommersacher, A., Denk, H. & Moinfar, F. Metaplastic breast carcinomas: are they of myoepithelial differentiation?: immunohistochemical profile of the sartcomatoid sybtype using novel myoepithelial markers. *Am. J. Surg. Pathol.* **29**, 347–353 (2005).
23. Al-Hajj, M. *et al.* Prospective identification of tumorigenic breast cancer cells. *Proc. Natl Acad. Sci. USA* **100**, 3983–3988 (2003).
24. Ben-Porath, I. *et al.* An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumours. *Nature Genet.* **40**, 499–507 (2008).
25. Sorlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl Acad. Sci. USA.* **98**, 10869–10874 (2001).
26. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
27. Nagle RB. *et al.* Characterization of breast carcinmoas by two monoclonal antibodies distinguishing myoepithelial from luminal epthelial cells. *J. Histochem. Cytochem.* **34**, 869–881 (1986).
28. Livasy, C. A., *et al.* Identification of a basal-like subtype of breast ductal carcinoma *in situ*. *Hum. Pathol.* **38**, 197–204, (2007).
29. Carey, L. A. *et al.* The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. *Clin. Cancer Res.* **13**, 2329–2334 (2007).
30. Rakha, E. A. *et al.* Are triple-negative tumours and basal-like breast cancer synonymous? *Breast Cancer Res.* **9**, 404 (2007).
31. Rakha, E. A., El-Sayed, M. E., Reis-Filho, J. & Ellis, I. O. Patho-biological aspects of basal-like breast cancer. *Breast Cancer Res. Treat.* 9 Mar 2008 (doi 10.1007/s10549-008-0007–0004).
32. Moinfar, F. Is 'basal-like' carcinoma of the breast a distinct clinicopathological entity? A critical review with cautionary notes. *Pathobiology* **75**, 119–131 (2008).
33. Malzahn, K., Mitze, M., Thoenes, M. & Moll, R. Biological and prognostic significance of stratified epithelial cytokeratins in infiltrating ductal breast carcinomas. *Virchows Arch.* **433**, 119–129 (1998).
34. Livasy, C. A. *et al.* Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. *Mod. Pathol.* **19**, 264–271 (2006).
35. Dairkee, S. H. Monoclonal marker that predicts early recurrence of breast cancer. *Lancet* **1**, 514 (1987).
36. Nielsen, T. O. *et al.* Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin. Cancer Res.* **10**, 5367–5374 (2004).
37. Cheang, M. C. *et al.* Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. *Clin. Cancer Res.* **14**, 1368–1376 (2008).
38. Behbod, F. & Rosen, J. M. Will cancer stem cells provide new therapeutic targets? *Carcinogenesis* **26**, 703–711 (2005).
39. Sims, A. H., Howell, A., Howell, S. J. & Clarke, R. B. Origins of breast cancer subtypes and therapeutic implications. *Nature Clin. Pract. Oncol.* **4**, 516–525 (2007).
40. Korsching, E. *et al.* Basal carcinoma of the breast revisited: an old entity with new interpretations. *J. Clin. Pathol.* **61**, 553–560 (2008).
41. Stingl, J. & Caldas, C. Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis. *Nature Rev. Cancer* **7**, 791–799 (2007).
42. Rakha, E. A., Reis-Filho, J. S. & Ellis, I. O. Basal-like breast cancer: a critical review. *J. Clin. Oncol.* **26**, 2568–2581 (2008).
43. Bossuyt, B. *et al.* Remarkably high frequency of EGFR expression in breast carcinomas with squamous differentiation. *Int. J. Surg. Pathol.* **13**, 319–327 (2005).
44. Ince, T. A. *et al.* Transformation of different human breast epithelial cell types leads to distinct tumor phenotypes. *Cancer Cell* **12**, 160–170 (2007).
45. Renou, J.-P. *et al.* Identification of genes differentially expressed in mouse mammary epithelium transformed by an activated β-catenin. *Oncegene* **22**, 4594–4610 (2003).
46. Rosner, A., *et al.* Pathway pathology: histological differences between ErbB/Ras and Wnt pathway transgenic mammary tumours. *Am. J. Pathol.* **161**, 1087–1097 (2002).
47. McCarthy, A. *et al.* A mouse model of basal-like breast carcinoma with metaplastic elements. *J. Pathol.* **211**, 389–398 (2007).
48. Liu, X. *et al.* Somatic loss of BRCA1 and p53 in mice induces mammary tumors with features of human BRCA1-mutated basal-like breast cancer. *Proc. Natl Acad. Sci. USA* **104**, 12111–12116 (2007).
49. Xu, X. *et al.* Conditional mutation of Brca1 in mammary epithelial cells results in blunted ductal morphogenesis and tumour formation. *Nature Genet.* **22**, 37–43 (1999).
50. Hosey, A. M. *et al.* Molecular basis for estrogen receptor α deficiency in BRCA1-linked breast cancer. *J. Natl Cancer Inst.* **99**, 1683–1694 (2007).
51. Liu, S. *et al.* BRCA1 regulates human mammary stem/progenitor cell fate. *Proc. Natl Acad. Sci. USA* **105**, 1680–1685 (2008).
52. Foulkes, W. D. *BRCA1* functions as a breast stem cell regulator. *J. Med. Genet.* **41**, 1–5 (2004).
53. Hedenfalk, I. *et al.* Gene-expression profiles in hereditary breast cancer. *N. Engl. J. Med.* **344**, 539–548 (2001).
54. Foulkes, W. D. *et al.* Germline *BRCA1* mutations and a basal epithelial phenotype in breast cancer. *J. Natl Cancer Inst.* **95**, 1482–1485 (2003).
55. Lakhani, S. R. *et al.* Prediction of BRCA1 status in patients with breast cancer using estrogen receptor and basal phenotype. *Clin. Cancer Res.* **11**, 5175–5180 (2005).
56. Crook, T., Crossland, S., Crompton, M. R., Osin, P. & Gusterson, B. A. p53 mutations in BRCA1-associated familial breast cancer. *Lancet* **350**, 638–639 (1997).
57. Rakha, E. A. *et al.* Expression of BRCA1 protein in breast cancer and its prognostic significance. *Hum. Pathol.* **39**, 857–865 (2008).
58. Turner, N. C. *et al.* BRCA1 dysfunction in sporadic basal-like breast cancer. *Oncogene* **26**, 2126–2132 (2007).
59. Mullan, P. B., Gorski, J. J. & Harkin, D. P. BRCA1 — a good predictive marker of drug sensitivity in breast cancer treatment? *Biochim. Biophys. Acta* **1766**, 205–216 (2006).
60. Kilburn, L. S. 'Triple negative' breast cancer: a new area for phase III breast cancer clinical trials. *Clin. Oncol.* **20**, 35–39 (2008).
61. Fan, C. *et al.* Concordance among gene-expression-based predictors for breast cancer. *N. Engl. J. Med.* **355**, 560–569 (2006).
62. Haibe-Kains, B. *et al.* Comparison of prognostic gene expression signatures for breast cancer. *BMC Genomics* **9**, 394 (2008).
63. Marchio, C. *et al.* The genomic profile of *HER2*-amplified breast cancers: the influence of ER status. *J. Pathol.* **216**, 399–407 (2008).

Acknowledgements

The laboratory of B.G. is supported by Breakthrough Breast Cancer.

DATABASES

National Cancer Institute Drug Dictionary: http://www.cancer.gov/drugdictionary/docetaxel

UniProtKB: http://www.uniprot.org

αSMA | BRCA1 | CD10 | cytokeratin 13 | cytokeratin 14 | cytokeratin 17 | cytokeratin 18 | cytokeratin 19 | cytokeratin 4 | cytokeratin 5 | cytokeratin 7 | cytokeratin 8 | EGFR | ERα | ERBB2 | GATA3 | KIT | MUC1 | p63 | PR | WNT1

FURTHER INFORMATION

B. Gusterson's homepage: http://www.gla.ac.uk/departments/cancersciences/pathologyandgeneregulation/

ALL LINKS ARE ACTIVE IN THE ONLINE PDF