Transcriptomic signatures in breast cancer

Jianjiang Fu and Stefanie S. Jeffrey*

DOI: 10.1039/b618163e

High throughput DNA microarray technology has been broadly applied to the study of breast cancer to classify molecular subtypes, to predict outcome, survival, response to treatment, and for the identification of novel therapeutic targets. Although results are promising, this technology will not have a full impact on routine clinical practice until there is further standardization of techniques and optimal clinical trial design. Due to substantial disease heterogeneity and the number of genes being analyzed, collaborative, multi-institutional studies are required to accrue enough patients for sufficient statistical power. Newer bioinformatic approaches are being developed to assist with the analysis of this important data.

Introduction

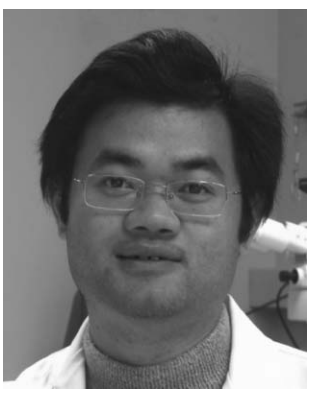
Breast cancer involves a wide range of pathological entities with diverse clinical courses. It is the most common malignancy and leading cause of cancer death among American women between the ages 20 to 59 and the second cause of cancer death in women aged 60 to 79.[1] In the UK and US, breast cancer mortality is declining,[2] attributed to the implementation of widespread screening mammography, earlier diagnosis, and advances in adjuvant treatment.[3] Treatment decisions are usually made according to general guidelines,[4] but not all patients benefit from their treatment. Efforts are now aimed at tailoring treatment for the individual patient, known as "personalized medicine".[5] This requires developing an accurate prognostic profile to define which patients should receive systemic therapy (hormone or chemotherapy) before and/or after surgery, and to decide which systemic treatments are most suitable for a given patient. The recent development of DNA microarray and related technologies provides an opportunity to perform more detailed and individualized tumor characterization.

Microarray technology, with its ability to simultaneously analyze tens of thousands genes, has transformed our understanding of human breast cancer. Previous classification systems historically relied on light microscopic findings and single marker-tumor features (*e.g.* estrogen receptor, ER). Expression profiling and other "-omic" technologies facilitate discovery of relevant signatures that may have an impact on prognosis (forecasting clinical outcome), prediction (forecasting tumor response to a specific therapy), and provide further insights into tumor biology,[6,7] informing both the clinician and the scientist.

Molecular classification by gene expression profiling

Using DNA microarrays, breast cancer heterogeneity has been confirmed at the gene expression level.[8] Multiple breast cancer subtypes with distinct gene expression patterns and different prognoses have been identified in primary breast cancers and their metastases. A Stanford–Norway collaborative research program revealed that breast cancers can be classified into five or six distinct

Stanford University Medical Center, MSLS Building, Room P214, 1201 Welch Road, M/C 5494, Stanford, CA, USA 94305-5494. E-mail: ssj@stanford.edu; Fax: +1 650-724-3229; Tel: +1 650-723-0799



Jianjiang Fu

Jianjiang Fu, PhD, graduated from China Pharmaceutical University with a BS degree. He earned his PhD from the Institute of Materia Medica at Peking Union Medical College as a pharmacologist specializing in tumor pharmacology. Dr Fu is a postdoctoral research fellow in the Jeffrey lab at Stanford University School of Medicine. His current interest is in the genomic characterization of circulating tumor cells and metastases in breast cancer.

Stefanie S. Jeffrey

Stefanie S. Jeffrey, MD, graduated from Harvard University with an AB in Chemistry and Physics and an MA in Chemistry. She did her medical and postgraduate training at the University of California, San Francisco. Dr Jeffrey is Associate Professor of Surgery at Stanford University School of Medicine and the John and Marva Warnock Faculty Scholar in Cancer Research. She was a key member of the Stanford–Norway team that pioneered the use of DNA microarrays to study breast cancer. Her lab is now developing and applying novel technologies for the study of circulating tumor cells and cancer metastases.

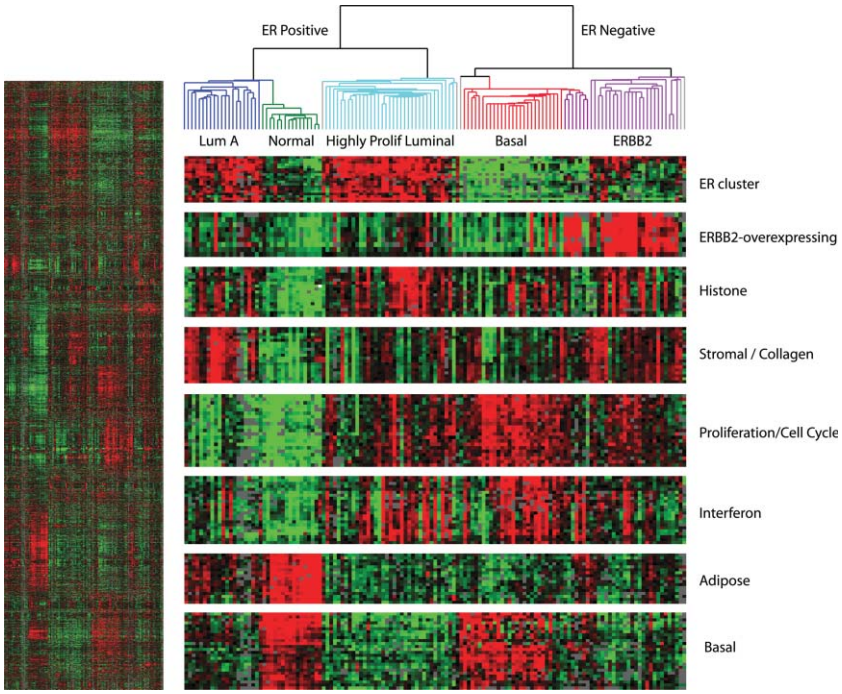


Fig. 1 Expression profiles of 119 breast cancers and 16 normal tissues. Tumors cluster into defined molecular subtypes.

molecular subtypes based their microarray expression data and this result has been established in other datasets worldwide[9–17] (Fig. 1).

Perou *et al.*[10] originally showed that breast tumor samples separate into two groups defined by ER status. Tumors in the ER-positive group have expression patterns reminiscent of the luminal epithelial cells of the breast, expressing luminal cytokeratins 8/18, ER (*ESR1*) and genes associated with ER overexpression such as *GATA3* and *NAT1*. There are at least two subtypes of ER-positive tumors,[11] luminal A and luminal B (we now term the luminal B group "highly proliferating luminals").[18] These two hormone receptor-overexpressing tumor classes are distinguished by gene expression patterns and markedly different clinical outcomes. Luminal A tumors have, in general, the highest expression of ER and ER-related genes. Luminal B tumors show expression of luminal/ER-related genes, but are further distinguished by relatively high expression of proliferation and cell cycle-related genes.

Conversely, hormone receptor negative breast cancers comprise two distinct subtypes, *ERBB2*-overexpressing and basal-like, that differ in biology and behavior; both show comparatively poor outcomes. Basal-like tumors highly express genes characteristic of breast basal epithelial cells, including strong expression of basal cytokeratins 5, 6 and 17. These tumors show high expression of proliferation/cell cycle-related genes, lack *ER* and ER-related genes, show low *ERBB2* (HER2/neu) expression; and show low expression of BRCA1 protein.[19,20] In addition, basal-like tumors usually have aggressive features such high tumor grade and *TP53* mutations.[11,18,21] The *ERBB2*-overexpressing subtype is characterized by overexpression of genes in a 17q amplicon that include *ERBB2* and *GRB7*. Like the basal-like subtype, *ERBB2*-overexpressing tumors have a high proportion of *TP53* mutations,[11,18,21] and are significantly more likely to be grade III.[21]

Finally, a normal breast-like subtype was identified that has some characteristics in common with normal breast tissue, including adipose and other non-epithelial tissue components of the tumor microenvironment. These tumors, like normal breast tissue, show relative overexpression of basal epithelial genes and relative underexpression of luminal epithelial genes. Invasive lobular breast cancers often show normal-like expression profiles.[11,12]

The distinct subtypes of breast cancer have also been characterized by immunohistochemical protein markers (*e.g.* ER, PR, HER2, HER1/EGFR, basal cytokeratins).[22] Recently, the luminal A and basal-like subtypes have been analyzed and validated on three different microarray platforms.[23] In addition to previously identified genes, signatures revealed distinct biological pathways: luminal A tumors expressed genes involved in fatty acid metabolism and steroid hormone-mediated signaling; basal-like tumors expressed cell proliferation and differentiation genes and pathway genes involved in p21-mediated and G1-S checkpoint signaling. A possible new subtype characterized by high expression of interferon (*IFN*)-regulated genes has been identified and linked to lymph node metastasis and poor prognosis.[16]

Using a novel approach and combined data from 599 microarrays, Kapp *et al.*[24] present evidence in support of the most consistently identifiable subtypes: ESR1+/ERBB2−, ESR−/ERBB2−, ERBB2+. Different sets of gene pairs were considered, statistically validated, and compared to clinical outcome. Tumors described by Sorlie centroids were generally grouped into expected categories (luminals were ESR+/ERBB2−; basals were ESR−/ERBB2−; and ERBB2-overexpressing tumors were mainly ERBB2+ subtype).

Outcome prediction by gene expression profiling

Metastases are the main cause of death in breast cancer patients, and improving the means of foretelling their development is a major goal of current clinical research.[25–30] Genomic-based tests predicting the likelihood of tumor recurrence provide information about the molecular biology of metastasis and provide a gauge of outcome prediction for new cases.[7–9,13,31–40]

A 70-gene prognosis signature was developed by van't Veer[31] *et al.* In this study, the investigators selected 78 lymph node-negative patients, with primary sporadic breast cancer, who were less than 55 years old; expression profiles were compared between 34 patients who developed distant metastasis within 5 years and 44 patients who remained disease-free for at least 5 years. ER status and other clinical variables were not

considered when the molecular predictor was developed. A 70-gene marker set was developed to classify tumors into good and poor prognosis groups. Not surprisingly, genes significantly up-regulated in the poor prognosis signature included those involved in cell cycle, invasion and metastasis, angiogenesis, and signal transduction. This 70-gene signature has been validated in larger series.[32] The largest series evaluated 302 multi-institutional tumor samples from 403 node-negative women, who were less than 61 years old and who did not receive systemic therapy, and found that the 70-gene signature added independent prognostic information to conventionally used prognostic criteria, although it did not perform as well in this larger trial with longer clinical follow-up.[41] Commercially available on the MammaPrint® array (Agendia BV, Amsterdam, The Netherlands), the 70-gene profile will be prospectively compared to a clinical-pathological prognostic tool (Adjuvant! Online) in selecting approximately 6000 node-negative patients for adjuvant chemotherapy in the *m*icroarray *i*n *n*ode negative *d*isease may *a*void *c*hemo*t*herapy (MINDACT, http://www.eortc.be/services/unit/mindact/) trial from the European Organization for Research and Treatment of Cancer.[42]

A different prognostic signature based on a different array platform was recently published by Wang *et al.*, specifying 76 genes (60 genes for ER-positive and 16 genes for ER-negative breast tumors) that distinguished lymph node-negative patients who developed distant metastases within five years.[33] This profile was found to be applicable to both pre- and postmenopausal patients and patients with 10–20 mm tumors, an especially common but not well-studied group. The genes in this prognostic signature belonged to many functional classes, including cell death, cell cycle and proliferation, transcriptional regulation, immune response, and growth, suggesting that different pathways can influence disease progression.[33,43] A multi-center trial testing this signature on a separate group of 180 breast tumors verified it as a strong predictor for remaining distant metastasis-free at 5 years. Moreover, comparing it to conventionally-used criteria, use of the signature would have potentially spared as many as 40% of node-negative patients for whom chemotherapy would have been recommended.[44] This signature also warrants prospective testing in larger clinical trials. However, as many clinicians know, ER-positive patients may relapse distantly 8 or more years after diagnosis, so 5 year follow-up may be too short a time interval for distinguishing good *vs.* poor prognostic signatures, particularly for patients of 60 years or less.[45–48]

Hypoxia (low oxygen) is clinically recognized as an important determinant of metastasis and poor patient outcome of breast cancer.[49,50] Chi *et al.* analyzed differential expression in global transcript levels in response to hypoxia in primary epithelial cells, including renal proximal tubule epithelial cells, normal breast epithelial cells, and endothelial cells. This gene expression signature was proposed to serve as a measure of hypoxia response activation in different human cancers, and to provide clinical outcome prediction in breast cancer.[51] Furthermore, when lysyl oxidase (*LOX*), an extracellular matrix enzyme up-regulated by hypoxia, is overexpressed in human tumors, it is associated with poorer distant metastasis-free and overall survival in ER-negative breast cancer patients, and *LOX* inhibition eliminates metastasis formation in a model system.[50,52] Given that cancer invasion and metastasis possess many histological similarities and may parallel some cellular behaviors of normal wound healing,[53,54] Chang *et al.* identified a 677-gene signature based on gene expression profiles of fibroblasts from ten anatomic sites in response to serum exposure, which seems to represent the role of fibroblasts in wound healing.[55] More recently, the reproducibility of the association between this serum-response gene expression signature and breast cancer progress was examined on a database of 295 breast cancer patients, which also had been used to identify and validate a 70-gene profile.[31,32] The results revealed that not only distant metastasis-free survival but also overall survival was strikingly reduced in patients whose tumors expressed this serum-response signature compared to tumors that did not express the signature.[34,35]

Histological grading of breast cancer provides clinically important prognostic information.[56–59] Gene expression profiling of tumors having different histological grades is under study. Ma *et al.* created distinct low grade (grade I) and high grade (grade III) signatures based on laser microdissection and DNA microarrays.[60] Sotiriou *et al.*[61] identified a 97-gene signature that they designate the gene expression grade index (GGI). This was comprised mainly of genes involved in cell cycle regulation and proliferation. Based on this signature, intermediate grade (grade II) tumors were subdivided into two groups with high and low risks of recurrence, demonstrating that the GGI may be used to classify intermediate grade tumors more accurately.[62] More recently, a group from the Genome Institute of Singapore reported six genes highlighted from 264 grade-associated genes identified by the expression profiles of 347 primary invasive breast cancers. These six genes were able to subdivide high and low grade tumors and also classify intermediate grade tumors into two distinct subtypes, termed G2a and G2b, which possessed similar, but not identical, clinical outcome to grade I and grade III tumors, respectively.[63]

Metastases to sites distant from the breast may contain tissue-specific expression profiles. Breast cancer most commonly spreads to bone marrow, lung, liver, brain, and adrenal gland.[64,65] Discovering genes that are functionally important for tissue-specific metastasis could help explain underlying tissue-tropism.[66] Using the MDA-MB-231 human breast cancer cell line as a model system, Kang and Minn *et al.*[67,68] identified a gene set that acted cooperatively to cause osteolytic metastasis. Analyzing this cell line-derived bone profile in 25 primary breast tumors, there was 67% sensitivity and 80% specificity for developing bone-only metastases in patients. An 86-gene bone metastatic signature was identified by Woelfle and colleagues between tumors from BM-positive and BM-negative patients. This gene set was mainly characterized by transcriptional repression genes and requires validation in a different data set.[69] Smid *et al.* reported that 69 genes may be involved in bone metastasis based on 107 primary breast tumors in patients

who all were lymph node-negative at the time of diagnosis and who experienced relapse in the bone and other organs.[69,70] They found that bone-only metastases involved the fibroblast growth factor receptor-MAPK pathway. A study analogous to Kang's report described a lung metastasis profile for breast cancer.[72] A 95-gene set was identified and reduced to 54 candidate genes by requiring that these genes also be differentially expressed across multiple independent lung-metastatic clones of breast cancer cell lines. The 54-gene signature was validated in a cohort of 82 breast cancer patients with a 10 year follow-up, and appeared to be a strong clinical predictor for lung selective metastasis.[72] The ability to predict site-specific metastasis based on gene expression profiling of a primary tumor may permit targeted therapeutic intervention and could help increase patient survival.

Gene expression profiling for prediction of response to treatments

Breast cancer is among the most sensitive to chemotherapy compared to other solid tumors. Systemic therapy for breast cancer includes hormonal therapy, chemotherapy, and novel agents.[5] Several single agent and combination chemotherapy regimens are effective treatments for breast cancer, such as cyclophosphamide, doxorubicin, 5-fluorouracil (5-FU) and taxanes.[5] In clinical practice, chemotherapy is applied empirically despite the fact that not all patients benefit from those agents. Hence, there is a need to identify predictive biomarkers for its efficacy. Currently, with the recent technological advances, it is anticipated that gene expression profiling in predicting efficacy and safety of breast cancer treatment may allow the definition of a pattern of clinically useful discriminatory gene expression.

Taxanes are a class of antimicrotubule agents that are proven to be effective and are routinely used in multidrug therapy of primary and advanced breast cancers.[73–77] Several groups have analyzed the gene expression profiles of patients that may benefit from taxane therapy. A Japanese group studied 44 primary or locally recurrent breast cancers treated with docetaxel using a 2453-gene high-throughput reverse transcriptase polymerase chain reaction technique.[78] They developed an 85-gene classifier for partial or complete response, which was validated in an additional 26 patients. Two groups, one from Baylor College of Medicine and the other from Millennium Pharmaceuticals and the M. D. Anderson Cancer Center, have used microarrays to study the response of locally advanced breast cancers to primary chemotherapy in 30 and 42 patients, respectively.[79–82] Multi-gene predictors of response were generated. These included a 92-gene list that initially predicted at least 75% tumor regression in response to four cycles of docetaxol in 24 patients, which was validated in an independent set of six patients. However, further analysis of residual tumors after three months of docetaxel treatment, revealed the residual cells had similar expression profiles to those that were initially resistant. Differentially expressed genes between initially sensitive and ultimately docetaxel-resistant cells included those involved in cell cycle arrest at G2/M, fatty acid/phospholipid metabolism or the mammalian target of rapamycin (mTOR) survival pathway that involves vesicular trafficking, oxidative bursts, and protein/organelle metabolism.[83] A separate 74-gene list predicting pathologic complete response to sequential weekly paclitaxel and fluorouracil + doxorubicin + cyclophosphamide (T/FAC) neoadjuvant chemotherapy in 24 patients was independently validated in an additional 18 patients. Another group from Germany reported a 512-gene signature, enriched for genes involved in transforming growth factor-beta and RAS-mediated signaling pathways, to predict pathologic complete response to primary systemic therapy with gemcitabine, epirubicin, and docetaxel.[84] Although all groups reported an association between gene expression profile and treatment outcome, the predictive power was too low for current clinical use and larger validation studies are underway or planned.

Huang *et al.* established several gene expression phenotypic models controlled by oncogenes, such as *HRAS*, *MYC* and *E2Fs*, and applied microarrays to analyze regulatory pathways that predicted oncogenic phenotypes.[85] More recently, Bild *et al.* from the same group used microarrays to analyze the gene expression profiling of oncogene and tumor-suppressor gene regulated pathways.[86–88] In each case, they identified a signature that represented the activated status of each oncogenic pathway. In addition to successfully predicting the activation status of each of the pathways in a range of human and mouse tumors, oncogenic pathway activation predicted the *in vitro* sensitivity of a broad range of human tumor cell lines to drugs targeting specific pathways. The analysis of oncogenic pathway signatures may offer guidance in the appropriate selection of tumor-specific combination therapies—multiple drugs that target multiple pathways—based on information specifying the activation state of these pathways.

Oncotype DX™ is a commercial clinically-validated multi-gene assay that provides a quantitative assessment of the likelihood of distant breast cancer recurrence in lymph node-negative, ER-positive breast cancer and also assesses the benefit from adjuvant chemotherapy in these patients. This 21-gene signature was established based on the paraffin-embedded tumor tissue from tamoxifen-treated patients[89–94] and was shown to predict risk for distant recurrence or death in several independent studies. Studying similar tumors, another signature, the HOXB13 : IL17BR expression ratio was identified and confirmed to predict survival and recurrence in tamoxifen-treated patients with ER-positive node-negative breast cancer.[95–97] In a different study, Jansen and colleagues identified genes that predicted response to tamoxifen in recurrent ER-positive breast carcinomas.[98] Golub's group recently described a "connectivity map" linking bioactive small-molecule perturbagens, gene signatures, and disease states based on DNA microarray assessment.[99]

Limitation and prospects

DNA microarray analysis has been shown to be a powerful tool in uncovering the mechanistic insights into tumor biology. It is being used to study almost all the aspects of cancer biology, from diagnosis to prognosis to drug responses, and for the development of new anticancer agents. It provides opportunities for more detailed characterization of

cancer biology and impacts on our understanding of malignant diseases. While reports of DNA microarray studies in oncology are exciting, the use of this technology has not yet been implemented clinically. Different signatures are reported by different studies of the same disease,[100–104] resulting in classifiers with little overlap between predictive gene lists for the same cancer,[103–106] even when the same microarray platform is used. Among the many limitations to deriving a stable molecular predictor for new cases of breast cancer is variability in technique platform,[105] sample size,[107,108] patient selection criteria,[40–42,109] statistical methods of data analysis,[110] noise and bias analysis,[111] and prediction rules. In most studies, the sample size (number of tumors assayed) is an order of magnitude smaller than the number of genes analyzed, leading to a lack of statistical power. Moreover, when large numbers of genes are analyzed, traditional approaches to multiple hypothesis correction are too conservative, finding few if any significant genes; alternative approaches use permutation methods to estimate false discovery rates.[112] Other approaches involve the use of gene sets to scale down the numbers of genes being tested and using the biological strength across a gene set to increase statistical power,[113,114] or using mathematical modeling to determine only the pathological component of high dimensional data.[115]

Because of the molecular heterogeneity of breast cancer, future large-scale clinical validation trials will, by necessity, be collaborative and multi-institutional. They must be guided by sound trial design and use analytic methods that incorporate developed rules of evidence[116] prior to their acceptance into routine clinical practice.

References

1 A. Jemal, R. Siegel, E. Ward, T. Murray, J. Xu, C. Smigal and M. J. Thun, *Ca Cancer J. Clin.*, 2006, **56**, 106–130.

2 R. Peto, J. Boreham, M. Clarke, C. Davies and V. Beral, *Lancet*, 2000, **355**, 1822.

3 Early Breast Cancer Trialists' Collaborative Group (EBCTCG), *Lancet*, 2005, **365**, 1687–1717.

4 R. W. Carlson and B. McCormick, *J. Natl. Compr. Cancer Network*, 2005, **3**, S7–S11.

5 V. Stearns, N. E. Davidson and D. A. Flockhart, *Pharmacogenomics J.*, 2004, **4**, 143–153.

6 S. S. Jeffrey, M. J. Fero, A. L. Borresen-Dale and D. Botstein, *Mol. Interventions*, 2002, **2**, 101–109.

7 V. A. Funari and S. S. Jeffrey, in *Advanced therapy of breast disease*, ed. S. E. Singletary, G. L. Robb and G. N. Hortobagyi, B. C. Decker, Hamilton, Ontario, Canada, 2nd edn, 2004, ch. 8, pp 845–858.

8 S. S. Jeffrey, P. E. Lonning and B. E. Hillner, *J. Natl. Compr. Cancer Network*, 2005, **3**, 291–300.

9 C. M. Perou, S. S. Jeffrey, M. van de Rijn, C. A. Rees, M. B. Eisen, D. T. Ross, A. Pergamenschikov, C. F. Williams, S. X. Zhu, J. C. Lee, D. Lashkari, D. Shalon, P. O. Brown and D. Botstein, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 9212–9217.

10 C. M. Perou, T. Sorlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. Lonning, A. L. Borresen-Dale, P. O. Brown and D. Botstein, *Nature*, 2000, **406**, 747–752.

11 T. Sorlie, C. M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, T. Thorsen, H. Quist, J. C. Matese, P. O. Brown, D. Botstein, P. E. Lonning and A. L. Borresen-Dale, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 10869–10874.

12 T. Sorlie, R. Tibshirani, J. Parker, T. Hastie, J. S. Marron, A. Nobel, S. Deng, H. Johnsen, R. Pesich, S. Geisler, J. Demeter, C. M. Perou, P. E. Lonning, P. O. Brown, A. L. Borresen-Dale and D. Botstein, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 8418–8423.

13 C. Sotiriou, S. Y. Neo, L. M. McShane, E. L. Korn, P. M. Long, A. Jazaeri, P. Martiat, S. B. Fox, A. L. Harris and E. T. Liu, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 10393–10398.

14 K. Yu, C. H. Lee, P. H. Tan, G. S. Hong, S. B. Wee, C. Y. Wong and P. Tan, *Cancer Res.*, 2004, **64**, 2962–2968.

15 K. Yu, C. H. Lee, P. H. Tan and P. Tan, *Clin. Cancer Res.*, 2004, **10**, 5508–5517.

16 Z. Hu, C. Fan, D. S. Oh, J. S. Marron, X. He, B. F. Qaqish, C. Livasy, L. A. Carey, E. Reynolds, L. Dressler, A. Nobel, J. Parker, M. G. Ewend, L. R. Sawyer, J. Wu, Y. Liu, R. Nanda, M. Tretiakova, A. R. Orrico, D. Dreher, J. P. Palazzo, L. Perreard, E. Nelson, M. Mone, H. Hansen, M. Mullins, J. F. Quackenbush, M. J. Ellis, O. I. Olopade, P. S. Bernard and C. M. Perou, *BMC Genomics*, 2006, **7**, 96.

17 S. H. Cheng, C. F. Horng, M. West, E. Huang, J. Pittman, M. H. Tsou, H. Dressman, C. M. Chen, S. Y. Tsai, J. J. Jian, M. C. Liu, J. R. Nevins and A. T. Huang, *J. Clin. Oncol.*, 2006, **24**, 4594–4602.

18 A. Langerød, H. Zhao, Ø. Borgen, J. M. Nesland, I. R. K. Bukholm, T. Ikdahl, R. Kåresen, A.-L. Børresen-Dale and S. S. Jeffrey, *Breast Cancer Res.*, 2007, **9**(R30), DOI: 101186/bcr1675.

19 D. M. Abd El-Rehim, G. Ball, S. E. Pinder, E. Rakha, C. Paish, J. F. Robertson, D. Macmillan, R. W. Blamey and I. O. Ellis, *Int. J. Cancer*, 2005, **116**, 340–350.

20 N. Turner, A. Tutt and A. Ashworth, *Nat. Rev. Cancer*, 2004, **4**, 814–819.

21 L. A. Carey, C. M. Perou, C. A. Livasy, L. G. Dressler, D. Cowan, K. Conway, G. Karaca, M. A. Troester, C. K. Tse, S. Edmiston, S. L. Deming, J. Geradts, M. C. Cheang, T. O. Nielsen, P. G. Moorman, H. S. Earp and R. C. Millikan, *JAMA, J. Am. Med. Assoc.*, 2006, **295**, 2492–2502.

22 T. O. Nielsen, F. D. Hsu, K. Jensen, M. Cheang, G. Karaca, Z. Hu, T. Hernandez-Boussard, C. Livasy, D. Cowan, L. Dressler, L. A. Akslen, J. Ragaz, A. M. Gown, C. B. Gilks, M. van de Rijn and C. M. Perou, *Clin. Cancer Res.*, 2004, **10**, 5367–5374.

23 T. Sorlie, Y. Wang, C. Xiao, H. Johnsen, B. Naume, R. R. Samaha and A. L. Borresen-Dale, *BMC Genomics*, 2006, **7**, 127.

24 A. V. Kapp, S. S. Jeffrey, A. Langerød, A. Børresen-Dale, W. Han, D. Noh, I. Bukholm, M. Nicolau, P. O. Brown and R. Tibshirani, *BMC Genomics*, 2006, **7**, 231; (Erratum in: *BMC Genomics*, 2007, **8**(1), 101).

25 I. J. Fidler, *Nat. Rev. Cancer*, 2003, **3**, 453–458.

26 A. F. Chambers, A. C. Groom and I. C. MacDonald, *Nat. Rev. Cancer*, 2002, **2**, 563–572.

27 D. Tarin, J. E. Price, M. G. Kettlewell, R. G. Souter, A. C. Vass and B. Crossley, *Br. Med. J. (Clin. Res. Ed.)*, 1984, **288**, 749–751.

28 D. Tarin, A. C. Vass, M. G. Kettlewell and J. E. Price, *Invasion Metastasis*, 1984, **4**, 1–12.

29 I. J. Fidler, *J. Natl. Cancer Inst.*, 1970, **45**, 773–782.

30 M. D. Cameron, E. E. Schmidt, N. Kerkvliet, K. V. Nadkarni, V. L. Morris, A. C. Groom, A. F. Chambers and I. C. MacDonald, *Cancer Res.*, 2000, **60**, 2541–2546.

31 L. J. van't Veer, H. Dai, M. J. van de Vijver, Y. D. He, A. A. Hart, M. Mao, H. L. Peterse, K. van der Kooy, M. J. Marton, A. T. Witteveen, G. J. Schreiber, R. M. Kerkhoven, C. Roberts, P. S. Linsley, R. Bernards and S. H. Friend, *Nature*, 2002, **415**, 530–536.

32 M. J. van de Vijver, Y. D. He, L. J. van't Veer, H. Dai, A. A. Hart, D. W. Voskuil, G. J. Schreiber, J. L. Peterse, C. Roberts, M. J. Marton, M. Parrish, D. Atsma, A. Witteveen, A. Glas, L. Delahaye, T. van der Velde, H. Bartelink, S. Rodenhuis, E. T. Rutgers, S. H. Friend and R. Bernards, *New Engl. J. Med.*, 2002, **347**, 1999–2009.

33 Y. Wang, J. G. Klijn, Y. Zhang, A. M. Sieuwerts, M. P. Look, F. Yang, D. Talantov, M. Timmermans, M. E. Meijer-van Gelder, J. Yu, T. Jatkoe,

E. M. Berns, D. Atkins and J. A. Foekens, *Lancet*, 2005, **365**, 671–679.

34 H. Y. Chang, D. S. Nuyten, J. B. Sneddon, T. Hastie, R. Tibshirani, T. Sorlie, H. Dai, Y. D. He, L. J. van't Veer, H. Bartelink, M. van de Rijn, P. O. Brown and M. J. van de Vijver, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 3738–3743.

35 E. T. Liu, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 3531–3532.

36 B. E. Hillner, *ACP J. Club*, 2003, **138**, 82.

37 G. Sauter and R. Simon, *New Engl. J. Med.*, 2002, **347**, 1995–1996.

38 A. Kallioniemi, *New Engl. J. Med.*, 2002, **347**, 2067–2068.

39 B. Weigelt, A. M. Glas, L. F. Wessels, A. T. Witteveen, J. L. Peterse and L. J. van't Veer, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 15901–15905.

40 K. Mikhitarian, W. E. Gillanders, J. S. Almeida, R. H. Martin, J. C. Varela, J. S. Metcalf, D. J. Cole and M. Mitas, *Clin. Cancer Res.*, 2005, **11**, 3697–3704.

41 M. Buyse, S. Loi, L. van't Veer, G. Viale, M. Delorenzi, A. M. Glas, M. S. d'Assignies, J. Bergh, R. Lidereau, P. Ellis, A. Harris, J. Bogaerts, P. Therasse, A. Floore, M. Amakrane, F. Piette, E. Rutgers, C. Sotiriou, F. Cardoso and M. J. Piccart, *J. Natl. Cancer Inst.*, 2006, **98**, 1183–1192.

42 R. Simon, *J. Natl. Cancer Inst.*, 2006, **98**, 1169–1171.

43 T. K. Jenssen and E. Hovig, *Lancet*, 2005, **365**, 634–635.

44 J. A. Foekens, D. Atkins, Y. Zhang, F. C. Sweep, N. Harbeck, A. Paradiso, T. Cufer, A. M. Sieuwerts, D. Talantov, P. N. Span, V. C. Tjan-Heijnen, A. F. Zito, K. Specht, H. Hoefler, R. Golouh, F. Schittulli, M. Schmitt, L. V. Beex, J. G. Klijn and Y. Wang, *J. Clin. Oncol.*, 2006, **24**, 1665–1671.

45 V. R. Grann, A. B. Troxel, N. J. Zojwalla, J. S. Jacobson, D. Hershman and A. I. Neugut, *Cancer (N. Y., NY, U. S.)*, 2005, **103**, 2241–2251.

46 C. Schairer, P. J. Mink, L. Carroll and S. S. Devesa, *J. Natl. Cancer Inst.*, 2004, **96**, 1311–1321.

47 M. Kaufmann and A. Rody, *J. Cancer Res. Clin. Oncol.*, 2005, **131**, 487–494.

48 H. F. Kennecke, I. A. Olivotto, C. Speers, B. Norris, S. K. Chia, C. Bryce and K. A. Gelmon, *Ann. Oncol.*, 2007, **18**, 45–51.

49 G. L. Semenza, *Trends Mol. Med.*, 2002, **8**, S62–S67.

50 J. T. Erler, K. L. Bennewith, M. Nicolau, N. Dornhofer, C. Kong, Q. T. Le, J. T. Chi, S. S. Jeffrey and A. J. Giaccia, *Nature*, 2006, **440**, 1222–1226.

51 J. T. Chi, Z. Wang, D. S. Nuyten, E. H. Rodriguez, M. E. Schaner, A. Salim, Y. Wang, G. B. Kristensen, A. Helland, A. L. Borresen-Dale, A. Giaccia, M. T. Longaker, T. Hastie, G. P. Yang, M. J. Vijver and P. O. Brown, *PLoS Med.*, 2006, **3**, e47.

52 N. C. Denko, L. A. Fontana, K. M. Hudson, P. D. Sutphin, S. Raychaudhuri, R. Altman and A. J. Giaccia, *Oncogene*, 2003, **22**, 5907–5914.

53 H. F. Dvorak, *New Engl. J. Med.*, 1986, **315**, 1650–1659.

54 M. J. Bissell and D. Radisky, *Nat. Rev. Cancer*, 2001, **1**, 46–54.

55 H. Y. Chang, J. B. Sneddon, A. A. Alizadeh, R. Sood, R. B. West, K. Montgomery, J. T. Chi, M. van de Rijn, D. Botstein and P. O. Brown, *PLoS Biol.*, 2004, **2**, E7.

56 C. W. Elston and I. O. Ellis, *Histopathology*, 1991, **19**, 403–410.

57 C. W. Elston and I. O. Ellis, *Histopathology*, 2002, **41**, 151.

58 M. H. Galea, R. W. Blamey, C. E. Elston and I. O. Ellis, *Breast Cancer Res. Treat.*, 1992, **22**, 207–219.

59 J. Brown, M. Jones and E. A. Benson, *Breast Cancer Res. Treat.*, 1993, **25**, 283.

60 X. J. Ma, R. Salunga, J. T. Tuggle, J. Gaudet, E. Enright, P. McQuary, T. Payette, M. Pistone, K. Stecker, B. M. Zhang, Y. X. Zhou, H. Varnholt, B. Smith, M. Gadd, E. Chatfield, J. Kessler, T. M. Baer, M. G. Erlander and D. C. Sgroi, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 5974–5979.

61 C. Sotiriou, P. Wirapati, S. Loi, A. Harris, S. Fox, J. Smeds, H. Nordgren, P. Farmer, V. Praz, B. Haibe-Kains, C. Desmedt, D. Larsimont, F. Cardoso, H. Peterse, D. Nuyten, M. Buyse, M. J. Van de Vijver, J. Bergh, M. Piccart and M. Delorenzi, *J. Natl. Cancer Inst.*, 2006, **98**, 262–272.

62 C. Desmedt and C. Sotiriou, *Cell Cycle*, 2006, **5**, 2198–2202.

63 A. V. Ivshina, J. George, O. Senko, B. Mow, T. C. Putti, J. Smeds, T. Lindahl, Y. Pawitan, P. Hall, H. Nordgren, J. E. Wong, E. T. Liu, J. Bergh, V. A. Kuznetsov and L. D. Miller, *Cancer Res.*, 2006, **66**, 10292–10301.

64 S. P. Leong, B. Cady, D. M. Jablons, J. Garcia-Aguilar, D. Reintgen, J. Jakub, S. Pendas, L. Duhaime, R. Cassell, M. Gardner, R. Giuliano, V. Archie, D. Calvin, L. Mensha, S. Shivers, C. Cox, J. A. Werner, Y. Kitagawa and M. Kitajima, *Cancer Metastasis Rev.*, 2006, **25**, 221–232.

65 A. Stevanovic, P. Lee and N. Wilcken, *Aust. Fam. Physician*, 2006, **35**, 309–312.

66 T. A. DiMeo and C. Kuperwasser, *Breast Cancer Res.*, 2006, **8**, 301.

67 A. J. Minn, Y. Kang, I. Serganova, G. P. Gupta, D. D. Giri, M. Doubrovin, V. Ponomarev, W. L. Gerald, R. Blasberg and J. Massague, *J. Clin. Invest.*, 2005, **115**, 44–55.

68 Y. Kang, P. M. Siegel, W. Shu, M. Drobnjak, S. M. Kakonen, C. Cordon-Cardo, T. A. Guise and J. Massague, *Cancer Cell*, 2003, **3**, 537–549.

69 U. Woelfle, J. Cloos, G. Sauter, L. Riethdorf, F. Janicke, P. van Diest, R. Brakenhoff and K. Pantel, *Cancer Res.*, 2003, **63**, 5679–5684.

70 M. Smid, Y. Wang, J. G. Klijn, A. M. Sieuwerts, Y. Zhang, D. Atkins, J. W. Martens and J. A. Foekens, *J. Clin. Oncol.*, 2006, **24**, 2261–2267.

71 S. L. Kominsky and N. E. Davidson, *J. Clin. Oncol.*, 2006, **24**, 2227–2229.

72 A. J. Minn, G. P. Gupta, P. M. Siegel, P. D. Bos, W. Shu, D. D. Giri, A. Viale, A. B. Olshen, W. L. Gerald and J. Massague, *Nature*, 2005, **436**, 518–524.

73 B. Fisher, J. Bryant, N. Wolmark, E. Mamounas, A. Brown, E. R. Fisher, D. L. Wickerham, M. Begovic, A. DeCillis, A. Robidoux, R. G. Margolese, A. B. Cruz, J. L. Hoehn, A. W. Lees, N. V. Dimitrov and H. D. Bear, *J. Clin. Oncol.*, 1998, **16**, 2672–2685.

74 E. Rivera, F. A. Holmes, D. Frye, V. Valero, R. L. Theriault, D. Booser, R. Walters, A. U. Buzdar, K. Dhingra, G. Fraschini and G. N. Hortobagyi, *Cancer (N. Y., NY, U. S.)*, 2000, **89**, 2195–2201.

75 N. Wolmark, J. Wang, E. Mamounas, J. Bryant and B. Fisher, *J. Natl. Cancer Inst. Monogr.*, 2001, **30**, 96–102.

76 A. K. Nowak, N. R. Wilcken, M. R. Stockler, A. Hamilton and D. Ghersi, *Lancet Oncol.*, 2004, **5**, 372–380.

77 R. W. Carlson, E. Brown, H. J. Burstein, W. J. Gradishar, C. A. Hudis, C. Loprinzi, E. P. Mamounas, E. A. Perez, K. Pritchard, P. Ravdin, A. Recht, G. Somlo, R. L. Theriault, E. P. Winer and A. C. Wolff, (National Comprehensive Cancer Network), *J. Natl. Compr. Cancer Network*, 2006, **4**, S1–S26.

78 K. Iwao-Koizumi, R. Matoba, N. Ueno, S. J. Kim, A. Ando, Y. Miyoshi, E. Maeda, S. Noguchi and K. Kato, *J. Clin. Oncol.*, 2005, **23**, 422–431.

79 J. C. Chang, E. C. Wooten, A. Tsimelzon, S. G. Hilsenbeck, M. C. Gutierrez, R. Elledge, S. Mohsin, C. K. Osborne, G. C. Chamness, D. C. Allred and P. O'Connell, *Lancet*, 2003, **362**, 362–369.

80 J. D. Brenton and C. Caldas, *Lancet*, 2003, **362**, 340–341.

81 M. Ayers, W. F. Symmans, J. Stec, A. I. Damokosh, E. Clark, K. Hess, M. Lecocke, J. Metivier, D. Booser, N. Ibrahim, V. Valero, M. Royce, B. Arun, G. Whitman, J. Ross, N. Sneige, G. N. Hortobagyi and L. Pusztai, *J. Clin. Oncol.*, 2004, **22**, 2284–2293.

82 M. Ellis and K. Ballman, *J. Clin. Oncol.*, 2004, **22**, 2267–2269.

83 J. C. Chang, E. C. Wooten, A. Tsimelzon, S. G. Hilsenbeck, M. C. Gutierrez, Y. L. Tham, M. Kalidas, R. Elledge, S. Mohsin, C. K. Osborne, G. C. Chamness, D. C. Allred, M. T. Lewis, H. Wong and P. O'Connell, *J. Clin. Oncol.*, 2005, **23**, 1169–1177.

84 O. Thuerigen, A. Schneeweiss, G. Toedt, P. Warnat, M. Hahn, H. Kramer, B. Brors, C. Rudlowski, A. Benner, F. Schuetz, B. Tews, R. Eils, H. P. Sinn, C. Sohn and P. Lichter, *J. Clin. Oncol.*, 2006, **24**, 1839–1845.

85 E. Huang, S. Ishida, J. Pittman, H. Dressman, A. Bild, M. Kloos, M. D'Amico, R. G. Pestell, M. West and J. R. Nevins, *Nat. Genet.*, 2003, **34**(2), 226–30.

86 A. H. Bild, G. Yao, J. T. Chang, Q. Wang, A. Potti, D. Chasse, M. B. Joshi, D. Harpole, J. M. Lancaster,

A. Berchuck, J. A. Olson, J. R. Marks, H. K. Dressman, M. West and J. R. Nevins, *Nature*, 2006, **439**, 353–357.
87 J. Downward, *Nature*, 2006, **439**, 274–275.
88 A. H. Bild, A. Potti and J. R. Nevins, *Nat. Rev. Cancer*, 2006, **6**, 735–741.
89 M. Cronin, M. Pho, D. Dutta, J. C. Stephans, S. Shak, M. C. Kiefer, J. M. Esteban and J. B. Baker, *Am. J. Pathol.*, 2004, **164**, 35–42.
90 L. A. Habel, S. Shak, M. K. Jacobs, A. Capra, C. Alexander, M. Pho, J. Baker, M. Walker, D. Watson, J. Hackett, N. T. Blick, D. Greenberg, L. Fehrenbacher, B. Langholz and C. P. Quesenberry, *Breast Cancer Res.*, 2006, **8**, R25.
91 S. Paik, S. Shak, G. Tang, C. Kim, J. Baker, M. Cronin, F. L. Baehner, M. G. Walker, D. Watson, T. Park, W. Hiller, E. R. Fisher, D. L. Wickerham, J. Bryant and N. Wolmark, *New Engl. J. Med.*, 2004, **351**, 2817–2826.
92 S. Paik, G. Tang, S. Shak, C. Kim, J. Baker, W. Kim, M. Cronin, F. L. Baehner, D. Watson, J. Bryant, J. P. Costantino, C. E. Geyer, D. L. Wickerham and N. Wolmark, *J. Clin. Oncol.*, 2006, **24**, 3726–3734.
93 F. J. Esteva, A. A. Sahin, M. Cristofanilli, K. Coombes, S. J. Lee, J. Baker, M. Cronin, M. Walker, D. Watson, S. Shak and G. N. Hortobagyi, *Clin. Cancer Res.*, 2005, **11**, 3315–3319.
94 S. Paik, G. Tang, S. Shak, C. Kim, J. Baker, W. Kim, M. Cronin, F. L. Baehner, D. Watson, J. Bryant, J. P. Costantino, C. E. Geyer, D. L. Wickerham and N. Wolmark, *J. Clin. Oncol.*, 2006, **24**, 3726–3734.
95 X. J. Ma, S. G. Hilsenbeck, W. Wang, L. Ding, D. C. Sgroi, R. A. Bender, C. K. Osborne, D. C. Allred and M. G. Erlander, *J. Clin. Oncol.*, 2006, **24**, 4611–4619.
96 X. J. Ma, Z. Wang, P. D. Ryan, S. J. Isakoff, A. Barmettler, A. Fuller, B. Muir, G. Mohapatra, R. Salunga, J. T. Tuggle, Y. Tran, D. Tran, A. Tassin, P. Amon, W. Wang, W. Wang, E. Enright, K. Stecker, E. Estepa-Sabal, B. Smith, J. Younger, U. Balis, J. Michaelson, A. Bhan, K. Habin, T. M. Baer, J. Brugge, D. A. Haber, M. G. Erlander and D. C. Sgroi, *Cancer Cell*, 2004, **5**, 607–616.
97 M. P. Goetz, V. J. Suman, J. N. Ingle, A. M. Nibbe, D. W. Visscher, C. A. Reynolds, W. L. Lingle, M. Erlander, X. J. Ma, D. C. Sgroi, E. A. Perez and F. J. Couch, *Clin. Cancer Res.*, 2006, **12**, 2080–2087.
98 M. P. Jansen, J. A. Foekens, I. L. van Staveren, M. M. Dirkzwager-Kiel, K. Ritstier, M. P. Look, M. E. Meijer-van Gelder, A. M. Sieuwerts, H. Portengen, L. C. Dorssers, J. G. Klijn and E. M. Berns, *J. Clin. Oncol.*, 2005, **23**, 732–740.
99 J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J. P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander and T. R. Golub, *Science*, 2006, **313**, 1929–1935.
100 A. A. Ahmed and J. D. Brenton, *Breast Cancer Res.*, 2005, **7**, 96–99.
101 C. Fan, D. S. Oh, L. Wessels, B. Weigelt, D. S. Nuyten, A. B. Nobel, L. J. van't Veer and C. M. Perou, *New Engl. J. Med.*, 2006, **355**, 560–569.
102 J. A. O'Shaughnessy, *New Engl. J. Med.*, 2006, **355**, 615–617.
103 L. Ein-Dor, I. Kela, G. Getz, D. Givol and E. Domany, *Bioinformatics*, 2005, **21**, 171–178.
104 R. Simon, M. D. Radmacher, K. Dobbin and L. M. McShane, *J. Natl. Cancer Inst.*, 2003, **95**, 14–18.
105 L. Pusztai, C. Mazouni, K. Anderson, Y. Wu and W. F. Symmans, *Oncologist*, 2006, **11**, 868–877.
106 L. Shi, W. Tong, F. Goodsaid, F. W. Frueh, H. Fang, T. Han, J. C. Fuscoe and D. A. Casciano, *Expert Rev. Mol. Diagn.*, 2004, **4**, 761–777.
107 J. F. Reid, L. Lusa, L. De Cecco, D. Coradini, S. Veneroni, M. G. Daidone, M. Gariboldi and M. A. Pierotti, *J. Natl. Cancer Inst.*, 2005, **97**, 927–930.
108 J. P. Ioannidis, *Lancet*, 2005, **365**, 454–455.
109 S. Michiels, S. Koscielny and C. Hill, *Lancet*, 2005, **365**, 488–492.
110 R. Shen, D. Ghosh and A. M. Chinnaiyan, *BMC Genomics*, 2004, **5**, 94.
111 P. G. Febbo and P. W. Kantoff, *J. Clin. Oncol.*, 2006, **24**, 3719–3721.
112 F. Dudbridge, A. Gusnanto and B. P. Koeleman, *Hum. Genomics*, 2006, **2**, 310–317.
113 A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander and J. P. Mesirov, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 15545–15550.
114 B. Efron and R. Tibshirani, On testing the significance of sets of genes, 2006, http://stat.stanford.edu/~brad/papers/genesetpaper.pdf.
115 M. Nicolau, R. Tibshirani, A. L. Børresen-Dale and S. S. Jeffrey, *Bioinformatics*, 2007, **23**, 957–265.
116 D. F. Ransohoff, *Nat. Rev. Cancer*, 2004, **4**, 309–314.

