ELSEVIER

The use of genomic tools for the molecular understanding of breast cancer and to guide personalized medicine

John A. Foekens[1], Yixin Wang[2], John W.M. Martens[1], Els M.J.J. Berns[1] and Jan G.M. Klijn[1]

[1] Department of Medical Oncology, Erasmus MC–Daniel den Hoed, Josephine Nefkens Institute and Cancer Genomics Centre, Dr. Molewaterplein 50, 3015 GE Rotterdam, The Netherlands
[2] Veridex LLC, a Johnson & Johnson Company, Technology Drive, Warren, NJ 07059, USA

The use of gene-expression microarray analysis to assess the expression levels of all the genes in the genome has tremendous potential. Important information has been obtained about many disease processes, particularly in classifying tumors in different subtypes and risk groups. Combining gene-expression data with other genomic information and the use of sophisticated bioinformatic tools enables the discovery of potential new targets for treatment, and is helpful for high-throughput drug screening and for designing new classes of drugs for targeted therapy. Here, we provide a short overview of the recent, promising developments in the field with emphasis on breast cancer.

Breast cancer is a pathological and clinical heterogeneous disease that progresses through accumulation of genomic and epigenetic aberrations. There is evidence that breast cancer originates from a small population of undifferentiated cells with self-renewal properties. These cancer stem cells, of which the origin is controversial, have not further been discussed in this report. Instead, we refer to excellent reviews regarding the role of stem cells in the pathogenesis of cancer development, their implications in breast cancer progression, treatment, and prognosis [1,2]. In current practice, treatment decisions are based on patient and tumor characteristics, and nowadays steroid hormone receptor and HER2 status are routinely assessed in tumor biopsies because they are potential targets for therapy. However, this information is insufficient to predict patient prognosis and the efficacy of the given systemic therapy accurately. Gene-expression microarray technology has become a popular tool to understand the biology of the disease and has made researchers realize that important information can be obtained to guide future treatments on a more individual basis. Since the landmark study of Perou *et al.* [3], in which the luminal, basal, Her2-like, and normal-like molecular breast cancer subtypes were described on the basis of approximately 500 differentially expressed genes (the intrinsic gene list), many studies addressing various important clinical questions have been performed. These studies, without trying to be complete, included the classification of breast tumors into histological subtypes [4–9], into subgroups with a different prognosis [4,5,7,10–17], subgroups with different preferences for relapse to different organs [18–21], and subgroups with different types of response to treatment [22–27].

Clinical application

Needless to say that many microarray studies suffer from the lack of sufficient sample size and proper validation in independent multicentric patient cohorts. There has been criticism and lack of understanding because the overlap of individual genes between various prognostic signatures was minimal or absent [28–32]. It is, therefore, important to note that the most widely studied prognostic signatures [16,17], in addition to in-house validation on independent patients [11,16], have successfully been validated in independent multicentric studies [33–35]. In these later studies, the observed sensitivities and/or specificities for 5-year TDM (time to distant metastasis) involving untreated lymph-node-negative breast cancer patients were 90%/42% for the 70-gene signature [33], and 90%/50% [35], and 97%/34% for the 76-gene signature [34], respectively. In Cox univariate analysis for TDM, the 70-gene signature yielded a hazard ratio of 2.3 [33], and the 76-gene signature hazard ratios were 7.4 [35] and 5.8 [34].

Corresponding author: Foekens, J.A. (j.foekens@erasmusmc.nl)

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2008.03.003

Moreover, both gene signatures outperformed the clinicopathologic risk as defined by Adjuvant! Online software (http://www.adjuvantonline.com) with adjusted hazard ratios for 10-year TDM of 3.5 for the 70-gene signature [33] and 5.1 for the 76-gene signature [34]. Interestingly, we observed a strong time dependency for both signatures, in the analysis for TDM peaking at four years with an adjusted hazard ratio of 9.2 for the 70-gene signature [33] and at five years for the 76-gene signature with an adjusted hazard ratio of 13.6 [34]. These independent multicenter validation studies demonstrate the reproducibility and robustness of these two gene expression signatures. Furthermore, the rigidity of the microarray technology is emphasized by the observation that the originally reported molecular portraits of breast tumors [3] are conserved across various microarray platforms [36]. Furthermore, the performance characteristics of different gene expression platforms carried out by the MicroArray Quality Control (MAQC) project showed a high level of intra-platform and inter-platform consistency [37]. This supports the view that clinical decision making based on gene expression profiles could be feasible in the future. All the above-described gene signatures were derived from global gene expression data using microarrays and were analyzed using frozen tissue biopsies. Time will tell whether gene expression assays, also those performed on routinely preserved formalin-fixed paraffin-embedded tissues, can provide reproducible independent prognostic information. In this respect we will have to await the results of the two recently started prospective randomized trials, MINDACT in Europe and TAILORx in USA [38], both including lymph-node-negative breast cancer patients and based on the 70-gene signature derived from global gene expression data using frozen samples [17] and the 21-gene signature derived from a set of 250 previously published candidate genes [39]. Early in 2007 the 70-gene signature [17], which is commercially available as the MammaPrint test, has been cleared for marketing by the FDA to determine breast cancer prognosis. Herewith it is the first cleared molecular test that profiles genetic activity (http://www.fda.gov/bbs/topics/NEWS/2007/NEW01555.html). The 21-gene signature [39], which is commercially available as the Oncotype DX test, was specifically designed to analyze RNA isolated from formalin-fixed paraffin-embedded tissues by quantitative real time (RT)-PCR. The possibility to use paraffin-embedded material has the advantage of a potentially more rapid implementation into the clinic because of relatively easy logistics with respect to tumor handling, shipping, and storage.

Biological pathways

Most, if not all, published gene expression signatures are based on a combination of individual genes. The implication of many of those genes for the studied phenotype is unclear, and there is only very limited overlap in genes between the different published gene signatures, even those addressing the same clinical question. It has been suggested that it might be more appropriate to interrogate the gene lists for biological themes, rather than just the individual genes [16,31,40–44]. Moreover, identification of the distinct biological processes between subtypes of cancer patients is more relevant to understand the disease and for targeted drug development [30,43]. In this respect, it is also important to realize that gene expression profiles of a primary tumor are maintained throughout the metastatic process, suggesting that the genomic composition of the primary tumor is also relevant for the metastasis to be treated [45]. Thus, identifying relevant pathways in the primary tumor might help to develop targeted drugs that may work in the adjuvant and metastatic setting. Taking into account that many genes carry prognostic information and have correlated expression on a gene expression array, especially genes involved in the same biological process, it is not surprising that different genes may be present in different signatures when separate training sets of patients and diverse statistical tools are used [16,28]. Hence, large overlaps in the genes present in the various gene signatures identified in different datasets are unlikely [31], but a common set of biological phenotypes may be represented [46,47].

Prognostic pathways

Because the construction of rational biological models will lead to a better understanding of the disease and, as a result, to the identification of potential drug targets, we have decided to study the biological processes associated with tumor metastatic capability in a large set of 344 lymph-node-negative breast cancer patients who had not received any adjuvant systemic therapy. Because gene expression patterns of estrogen receptor (ER) subgroups of breast tumors and also the genes related to prognosis are different, data analysis to derive gene signatures and subsequent pathways were conducted separately for ER-positive and ER-negative tumors. The datasets were re-sampled numerous times to construct a total of 1000 gene lists—where expression correlated with patients' distant metastasis-free survival [47]. On the basis of these gene lists, overrepresented pathways defined in Gene Ontology Biological Process (GOBP) were identified using Global Testing [40,41]. For ER-positive tumors, cell-division-related processes were frequently found in the overrepresented pathways, in addition to a couple of immune-related pathways. The two most significant pathways were 'Apoptosis' and 'Regulation of cell cycle'. For ER-negative tumors, many of the top significant pathways are related with RNA processing, transportation, and signal transduction, the two most significant being 'Regulation of cell growth' and 'Regulation of G-protein-coupled receptor signaling' [47]. Not surprisingly, the biological prognostic pathways between ER-positive and ER-negative tumors were entirely different. By comparing the pathways represented by the genes in four published prognostic gene signatures with 62 genes or more [7,16,17,39], each of the gene lists mapped to more than half of our core prognostic pathways for ER-positive or ER-negative tumors. Although similar pathways are represented in various signatures, it does not necessarily mean that the individual genes in a pathway are equally significant because in most pathways tens to hundreds of genes play a part (negative or positive), most of them with different individual contributions [47]. Gene-expression microarray data have also been used to identify the function of specific oncogenes [48,49] or to construct oncogenic pathway signatures [50]. Ultimately, the identification of distinct biological pathways and their activation mechanisms may guide the use of the most promising combination of drugs targeting multiple pathways leading to the actual implementation of personalized medicine.

Protease systems and common pathways

The metastatic behavior of tumor cells is facilitated by cell-associated protease systems that are able to degrade components of the

extracellular matrix (ECM). The urokinase-type plasminogen activator (uPA) and its main inhibitor PAI-1 have been shown to be associated with the aggressive behavior of many tumor types, including breast cancer [51,52]. uPA activates plasminogen to plasmin, which in turn cleaves a wide range of ECM and basement membrane components (either directly or indirectly) by activation of other ECM degrading proteases [53]. Many factors are involved in the invasive and metastatic capacity of cancer cells and the concerted action of the urokinase system, matrix metalloproteinases (MMPs), and their inhibitors (TIMPs) play crucial parts. For this matter, the development of drugs that interfere in the proteolytic cascade of tumor invasion and metastasis has attracted accumulating attention. In this review, we present an example of an approach to identify potential drug targets based on the identification of the key biological pathways associated with two prognostic protease systems in breast cancer. Using our previously published Affymetrix U133A oligonucleotide array data (http://www.ncbi.nlm.nih.gov/geo; accession numbers GSE2034 and GSE5327) of 344 primary tumors of lymph-node-negative breast cancer patients [16,47], we analyzed the biological pathways associated with the expression levels of uPA, PAI-1, MMP-2, and TIMP-1. Samples were ranked according to the level of the different 'protease' genes and the top 20% were compared with the bottom 20% of samples ($n = 69$ each). The Global Test program [40,41] was used to associate biological pathways [54] to samples expressing high or low levels of a particular gene and to assess the contribution of individual genes in a pathway. The top five most significant biological processes associated with differential uPA expression were Focal adhesion, ECM receptor interaction, Hedgehog signaling pathway, Complement and coagulation cascade, and TGFβ signaling pathway. Using a p-value of <0.01 for all four factors studied (uPA, PAI-1, MMP2, TIMP-1), only Focal adhesion and ECM receptor interaction remained as significant pathways. Because there is a large overlap in genes between these two biological processes, and the ECM receptor interaction pathway does not contain well-known drugable intracellular signaling molecules, further analyses were focused on the Focal adhesion pathway. The top 10 significant genes that showed an association with uPA, PAI-1, MMP2, and TIMP-1 (for all four, $p < 0.001$), included CAV1, CAV2, FLNA, FLNC, ILK, PARVA, VEGFC, and the platelet-derived growth factor receptor system with the genes PDGFC and receptors PDGFRA and PDGFRB. In total these analyses identified some potential interesting therapeutic targets, such as the PDGF receptor system, the Hedgehog and TGFβ signaling pathways, and the ILK–PARVA system of which it has been shown that the ILK-α-parvin complex protects cells from apoptosis and that β-parvin promotes apoptosis via inhibition of the ILK-α-parvin complex formation [55]. Thus, although tumors may utilize different protease systems, they do share upstream and downstream biological pathways. Identification of the key components playing a part in overlapping signaling pathways may identify potential effective drug targets.

Copy number alterations

Gene amplification is one of the mechanisms underlying the activation of oncogenes and may be associated with poor clinical outcome. Therefore, the identification of amplified oncogenes may have diagnostic and therapeutic potential. ERBB2, the best characterized breast cancer oncogene, is located on chromosome 17q21 and is amplified in 20–30% of breast cancers [56]. Array comparative genomic hybridization (aCGH) and fluorescence *in situ* hybridization (FISH) have been widely used to study gene copy number changes. In earlier studies in breast cancer, these technologies have revealed a number of gene copy number alterations (CNAs), including regions with high-level amplification that were associated with poor clinical outcome [57–59].

The combination of gene-expression data with those obtained with aCGH to assess CNAs allowed the identification of new amplicons and their candidate targets [60]. Using such a combined approach for analyzing breast cancers, 66 genes deregulated by high-level amplification were identified. These genes or their downstream effectors could serve as potential targets for treatment [10]. Many of these genes (of which their expression correlated with amplification) are known to be involved in various metabolic signaling pathways [10]. As a higher resolution alternative for CGH arrays, single nucleotide polymorphism (SNP) arrays in combination with gene-expression microarrays can also be used to assess clinically relevant CNAs. In this respect, by genome-wide SNP analysis in conjunction with gene-expression profiling we recently identified DNA amplification loci that predict breast cancer aggressiveness [61]. The above studies suggest that CNAs, together with gene-expression profiles, provide a strategy to combine molecular markers and build mathematical models of risk assessment.

MicroRNA

MicroRNAs (miRNAs) are recently discovered, highly conserved, noncoding small 19-nucleotide to 24-nucleotide RNAs (ncRNAs) that are cleaved from hairpin precursors that are produced from large primary transcripts [62,63]. These genes lack a significant open reading frame and use the endogenous RNA interference pathway to modulate gene expression [64]. In mammals, targeted messengers are mostly posttranscriptionally regulated through a blockage of protein translation [64]. It has become evident that miRNAs show tremendous potential to contribute to various normal cellular and developmental processes and to malignant transformation and progression [62,63,65]. Currently, over 530 human miRNA genes have been identified (http://microrna.sanger.ac.uk/sequences). Biochemical and *in silico* analyses have already suggested a plethora of predicted mRNA targets and evidence is accumulating that miRNAs can act as tumor suppressor genes and as oncogenes [62,63,65]. In breast cancer, miRNA levels are different between lymph-node-positive and lymph-node-negative disease, and between steroid hormone receptor-positive and receptor-negative tissue [66]. In a variety of cancer types, miRNA-expression profiling has identified various signatures that are associated with staging, prognosis, or response to treatment [63]. Furthermore, the intrinsic breast cancer subtypes luminal A, luminal B, basal-like, HER2-like and normal-like show a different miRNA expression [67], and recently miRNA10b has been causally linked to breast cancer invasion [68]. miRNAs also display a high frequency of genomic alterations in various tumor types and are located within fragile sites of certain chromosomes [62]. In a study in which miRNA expression was compared with CNAs determined by array CGH, 41 miRNAs of the total of 283 studied were identified with gene copy number changes (25 gains, 15

losses) in ovarian and breast cancers, and melanoma samples [69]. In addition, Zhang *et al.* showed that many other miRNA genes with copy number changes were unique for each of the three tumor types and that miRNA copy number changes correlate with miRNA expression [69]. Alongside miRNAs, a particular class of ncRNAs encoded by transcribed ultra conserved regions (T-UCRs) has recently shown to be consistently altered at the genomic level in a fair number of human leukemias and carcinomas, and that miRNAs may interact with these T-UCRs [70]. These UCRs are located in genomic regions putatively involved in human tumorigenesis. miRNAs may interact with these T-UCRs, and Calin *et al.* proposed a model in which alteration in both coding and non-coding RNAs cooperate in the initiation and progression of malignancy [70]. miRNA profiling is a very promising new field to explore since miRNAs have the potential to act as therapeutic molecules. Furthermore, the identification of miRNA genes that target genes and pathways, which are candidates for therapeutic intervention, might guide the biotechnological development of new drugs [62,63,70–72].

Mutations, SNPs

Currently, only a limited number of genes are known to be causally linked to the development of a specific type of cancer. It was hypothesized that a comprehensive analysis of human cancers could lead to the discovery of a set of genes that are linked together through a shared phenotype [73]. Therefore, Sjöblom *et al.* performed a comprehensive sequencing analysis of consensus coding sequences in 11 breast and 11 colon cancers. Of the 13 023 genes analyzed, 1 146 were found to be mutated. By applying stringent criteria, 236 of the mutated genes remained after validation [73]. Of these, 189 were defined as the so-called candidate cancer (CAN) genes that were mutated at a significant frequency. Of these genes, 122 were identified in breast and 69 in colon cancer. Two of these genes (OBSCN, TP53) were identified in both cancer types. Individual breast cancers on average had 12 CAN genes mutated and colon cancers had nine, while none of the cancers had more than six mutated CAN genes in common with any of the other 21 studied cancers [73]. Most of the 189 CAN genes were not known previously to be mutated in human tumors and they predict many cellular functions. Subsequently, Wood *et al.* analyzed 20 857 exons representing transcripts of 18 191 genes of the same set of breast and colon cancers and concluded that the genomic landscape of breast and colorectal cancers are composed of a handful of frequently mutated individual CAN genes and a much larger number or CAN genes that are mutated at a low frequency [74]. Together, these interesting results provide new clues to the pathogenesis and to future research opportunities. These studies also potentially provide new leads for therapeutic intervention strategies particularly for those genes that code for cell surface proteins and proteins with drugable enzymatic activities [73]. Apart from sequencing analysis, which is potentially available for the entire human genome, genome-wide SNP analysis is also currently feasible. Recently, genome-wide association studies of breast, colorectal, and prostate cancer were conducted by genotyping over 2 00 000 to over 5 00 000 SNPs. Some of the highlights of the obtained results are the identification of alleles in FGFR2 as risk factors in postmenopausal breast cancer [75], common variants on chromosomes 2q35 and 16q12 as risk factors for ER-positive breast cancer [76], new loci on chromosome 8q24 that confer prostate cancer [77] and colorectal cancer [78], SMAD7 association with colorectal cancer risk [79], and five novel breast cancer susceptibility loci [80]. In the latter study, plausible causative genes in the two most significant defined loci were again FGFR2 [75] and TNRCP, the latter was reported earlier by our group at the mRNA level to be associated with metastasis to the bone [19]. These SNP studies may have important consequences for the counseling, follow-up, and treatment of the person carrying the respective SNP. Furthermore, they provide leads for the development of the new targeted therapies.

DNA methylation

A common and early event in cancer is aberrant DNA methylation of cytosine phosphoguanine dinucleotides (CpG) within gene regulatory regions that, in general, adversely affects gene expression [81]. Excellent up-to-date reviews from leading scientists in the field, also describing other epigenetic events such as histone deacetylation, are available [82,83]. DNA-methylation patterns are tumor-specific and may be used in clinical practice for early detection of disease, for subclassification of tumors, and have potential as prognostic or predictive markers in a variety of cancers [82,83]. DNA-methylation tags are chemically stable, can be readily amplified, are very important for routine diagnostic application and can easily be obtained from formalin-fixed paraffin-embedded tissues and from body fluids. Furthermore, several DNA methylating agents and drugs that inhibit histone deacetylases are in clinical development. The finding that miRNA expression, also of those with a tumor suppression function, can be regulated by epigenetic mechanisms [84–86] opens the possibility that future anticancer therapy may also be based on the epigenetic regulation of miRNAs. Thus, ''our expectations are high'', as rightly concluded by Esteller [87].

Finding patients sensitive to individualized treatment

Exploiting *in vitro* cell line models for drug sensitivity profiling appears to be a very promising and feasible approach [27,88–90], and it is our belief that there is a strong clinical potential for this pharmacogenomic approach. *Potti et al.* [88,89,91] have proposed that future individualized treatment strategies, also for combination regimens, might be guided by a tumor's drug sensitivity profile based on cell lines that serve as surrogate phenotypes. As an extension of this, Lamb *et al.* have described 'The Connectivity Map' as a systematic approach to discover functional connections among diseases, genetic perturbation, and drug action [92,93]. Using this resource the investigators showed that gene-expression signatures could be used to find functional connections between small molecule compounds and the response of cells cultured *in vitro*. By *in silico* assessment of factors in the database and their ability to reverse signatures associated with drug resistance and sensitivity profiles, this approach allowed for the identification of drugs with a common mechanism of action and the discovery of new mechanisms of action [92,93], and may even be used to identify potential new therapeutics [93]. For example, in a search for androgen receptor (AR) signaling inhibitors, connections to heat-shock protein 90 (HSP90) inhibition were found by mapping the gene-expression signatures for AR inhibition with those obtained by celastrol and gedunin (and their derivatives) to the

Connectivity Map library [94]. This implies that inhibition of AR signaling might be achieved via inhibition of HSP90 activity by the structurally similar drugs celastrol and gedunin [94]. Another example involves the identification of a connection between mTOR inhibition and glucocorticoid sensitivity in acute lymphoblastic leukemia via mapping the gene-expression signatures to the Connectivity Map [95]. These examples show that the Connectivity Map approach, with its database on the efficacy of thousands of small molecule drugs, and other comparable (but possibly more powerful) approaches linking also upstream and downstream signaling networks [96,97], may contribute significantly to the efficiency of drug discovery and development.

Conclusion

The continuous progress in technological advances and the large-scale high-throughput analyses in global genomics studies allow for a comprehensive collection of genomics and transcriptomics information from all tumor types. It is important to realize that biomarkers, often as a combination, are crucial links between pathophysiology, diagnostics, and personalized medicine. Many different clinical questions have already been addressed by high-throughput modern technologies, and many more will follow in the forthcoming years. Important refined information can also be collected from combining data obtained by various techniques, such as from high-density DNA microarrays, miRNA arrays, CGH arrays, SNP arrays, and DNA-methylation arrays. In addition, very important data (including pathways and interacting networks) can be obtained by combining genomics data with those obtained with advanced proteomics, including mass spectrometry and reverse-phase protein (micro)arrays [98–100] (not discussed in this report). It is imperative that better tools become available to allow a full integration of genomic and transcriptomic data, also in the context of a systems biology approach including metabolic networks as well. In this respect, mathematicians, statisticians, and bioinformaticians will continue to play key parts in the forthcoming decade. A lot of important genomics and transcriptomics information has already been gathered, particularly during the past few years, and this flow of information is rapidly increasing. Routine implementation and clinical use of the current predictive or prognostic profiles will be many years away if we wait for the results of ongoing, and yet to be planned, prospective randomized clinical trials. A pitfall here is that, long before a trial may have given conclusive results, another and more promising and possibly easier applicable assay will be available, and the procedures of validation have to start all over again. Nonetheless, at this very moment, promising approaches to guide future drug development, and to classify patients into different risk groups with respect to prognosis and response to therapy prediction, have been described and the current outlook has the potential to make cancer patient care in the near future truly personal.

Acknowledgements

We thank Marcel Smid, Yi Zhang, and Jack Yu for performing bioinformatic analyses. The work is supported, partly, by the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO).

References

1 Cariati, M. and Purushotham, A.D. (2008) Stem cells and breast cancer. *Histopathology* 52, 99–107
2 Stingl, J. and Caldas, C. (2007) Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis. *Nat. Rev. Cancer* 7, 791–799
3 Perou, C.M. *et al.* (2000) Molecular portraits of human breast tumours. *Nature* 406, 747–752
4 Sorlie, T. *et al.* (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U. S. A.* 98, 10869–10874
5 West, R.B. *et al.* (2005) Determination of stromal signatures in breast carcinoma. *PLoS Biol.* 3, e187
6 Gruvberger, S. *et al.* (2001) Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Res.* 61, 5979–5984
7 Sotiriou, C. *et al.* (2006) Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J. Natl. Cancer Inst.* 98, 262–272
8 Bertucci, F. *et al.* (2006) Gene expression profiling shows medullary breast cancer is a subgroup of basal breast cancers. *Cancer Res.* 66, 4636–4644
9 Hedenfalk, I. *et al.* (2001) Gene-expression profiles in hereditary breast cancer. *N. Engl. J. Med.* 344, 539–548
10 Chin, K. *et al.* (2006) Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* 10, 529–541
11 van de Vijver, M.J. *et al.* (2002) A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* 347, 1999–2009
12 Chang, H.Y. *et al.* (2005) Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc. Natl. Acad. Sci. U. S. A.* 102, 3738–3743
13 Winter, S.C. *et al.* (2007) Relation of a hypoxia metagene derived from head and neck cancer to prognosis of multiple cancers. *Cancer Res.* 67, 3441–3449
14 Liu, R. *et al.* (2007) The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N. Engl. J. Med.* 356, 217–226
15 Yu, K. *et al.* (2004) A molecular signature of the Nottingham prognostic index in breast cancer. *Cancer Res.* 64, 2962–2968
16 Wang, Y. *et al.* (2005) Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365, 671–679
17 Van 't Veer, L.J. *et al.* (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530–536
18 Kang, Y. *et al.* (2005) Breast cancer bone metastasis mediated by the Smad tumor suppressor pathway. *Proc. Natl. Acad. Sci. U. S. A.* 102, 13909–13914
19 Smid, M. *et al.* (2006) Genes associated with breast cancer metastatic to bone. *J. Clin. Oncol.* 24, 2261–2267
20 Minn, A.J. *et al.* (2007) Lung metastasis genes couple breast tumor size and metastatic spread. *Proc. Natl. Acad. Sci. U. S. A.* 104, 6740–6745
21 Minn, A.J. *et al.* (2005) Genes that mediate breast cancer metastasis to lung. *Nature* 436, 518–524
22 Jansen, M.P. *et al.* (2005) Molecular classification of tamoxifen-resistant breast carcinomas by gene expression profiling. *J. Clin. Oncol.* 23, 732–740
23 Hannemann, J. *et al.* (2005) Changes in gene expression associated with response to neoadjuvant chemotherapy in breast cancer. *J. Clin. Oncol.* 23, 3331–3342
24 Rouzier, R. *et al.* (2005) Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clin. Cancer Res.* 11, 5678–5685
25 Chang, J.C. *et al.* (2003) Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet* 362, 362–369
26 Hess, K.R. *et al.* (2006) Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *J. Clin. Oncol.* 24, 4236–4244
27 Gyorffy, B. *et al.* (2005) Prediction of doxorubicin sensitivity in breast tumors based on gene expression profiles of drug-resistant cell lines correlates with patient survival. *Oncogene* 24, 7542–7551
28 Dupuy, A. and Simon, R.M. (2007) Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J. Natl. Cancer Inst.* 99, 147–157
29 Michiels, S. *et al.* (2005) Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 365, 488–492
30 Massague, J. (2007) Sorting out breast-cancer gene signatures. *N. Engl. J. Med.* 356, 294–297

31 Simon, R. (2006) Development and evaluation of therapeutically relevant predictive classifiers using gene expression profiling. *J. Natl. Cancer Inst.* 98, 1169–1171
32 Ransohoff, D.F. (2004) Rules of evidence for cancer molecular-marker discovery and validation. *Nat. Rev. Cancer* 4, 309–314
33 Buyse, M. *et al.* (2006) Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J. Natl. Cancer Inst.* 98, 1183–1192
34 Desmedt, C. *et al.* (2007) Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin. Cancer Res.* 13, 3207–3214
35 Foekens, J.A. *et al.* (2006) Multicenter validation of a gene expression-based prognostic signature in lymph node-negative primary breast cancer. *J. Clin. Oncol.* 24, 1665–1671
36 Hu, Z. *et al.* (2006) The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* 7, 96
37 Consortium, M. *et al.* (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat. Biotechnol.* 24, 1151–1161
38 Sotiriou, C. and Piccart, M.J. (2007) Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? *Nat. Rev. Cancer* 7, 545–553
39 Paik, S. *et al.* (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* 351, 2817–2826
40 Goeman, J.J. *et al.* (2005) Testing association of a pathway with survival using gene expression data. *Bioinformatics* 21, 1950–1957
41 Goeman, J.J. *et al.* (2004) A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics* 20, 93–99
42 Vogelstein, B. and Kinzler, K.W. (2004) Cancer genes and the pathways they control. *Nat. Med.* 10, 789–799
43 Tinker, A.V. *et al.* (2006) The challenges of gene expression microarrays for the study of human cancer. *Cancer Cell* 9, 333–339
44 O'Shaughnessy, J.A. (2006) Molecular signatures predict outcomes of breast cancer. *N. Engl. J. Med.* 355, 615–617
45 Weigelt, B. *et al.* (2005) Breast cancer metastasis: markers and models. *Nat. Rev. Cancer* 5, 591–602
46 Fan, C. *et al.* (2006) Concordance among gene-expression-based predictors for breast cancer. *N. Engl. J. Med.* 355, 560–569
47 Yu, J.X. *et al.* (2007) Pathway analysis of gene signatures predicting metastasis of node-negative primary breast cancer. *BMC Cancer* 7, 182
48 Lamb, J. *et al.* (2003) A mechanism of cyclin D1 action encoded in the patterns of gene expression in human cancer. *Cell* 114, 323–334
49 Huang, E. *et al.* (2003) Gene expression phenotypic models that predict the activity of oncogenic pathways. *Nat. Genet.* 34, 226–230
50 Bild, A.H. *et al.* (2006) Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 439, 353–357
51 Jänicke, F. *et al.* (2001) Randomized adjuvant chemotherapy trial in high-risk, lymph node-negative breast cancer patients identified by urokinase-type plasminogen activator and plasminogen activator inhibitor type 1. *J. Natl. Cancer Inst.* 93, 913–920
52 Look, M.P. *et al.* (2002) Pooled analysis of prognostic impact of urokinase-type plasminogen activator and its inhibitor PAI-1 in 8377 breast cancer patients. *J. Natl. Cancer Inst.* 94, 116–128
53 Andreasen, P.A. *et al.* (1997) The urokinase-type plasminogen activator system in cancer metastasis: a review. *Int. J. Cancer* 72, 1–22
54 Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30
55 Zhang, Y. *et al.* (2004) Distinct roles of two structurally closely related focal adhesion proteins, alpha-parvins and beta-parvins, in regulation of cell morphology and survival. *J. Biol. Chem.* 279, 41695–41705
56 Slamon, D.J. *et al.* (1989) Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* 244, 707–712
57 Kallioniemi, O.P. *et al.* (1992) ERBB2 amplification in breast cancer analyzed by fluorescence *in situ* hybridisation. *Proc. Natl. Acad. Sci. U. S. A.* 89, 5321–5325
58 Kallioniemi, A. *et al.* (1994) Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridisation. *Proc. Natl. Acad. Sci. U. S. A.* 91, 2156–2160
59 Isola, J.J. *et al.* (1995) Genetic aberrations detected by comparative genomic hybridization predict outcome in node-negative breast cancer. *Am. J. Pathol.* 147, 905–911
60 Pollack, J.R. *et al.* (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* 23, 41–46
61 Zhang, Y. *et al.* (2006) Genome-wide assessment of DNA copy number alterations in conjunction with gene expression profiling identifies DNA amplification loci that predict distant recurrence in lymph node-negative (LNN) primary breast cancer patients. *San Anontio Breast Cancer Conference* Abstract 32
62 Calin, G.A. and Croce, C.M. (2007) Chromosomal rearrangements and microRNAs: a new cancer link with clinical implications. *J. Clin. Invest.* 117, 2059–2066
63 Calin, G.A. and Croce, C.M. (2006) MicroRNA signatures in human cancers. *Nat. Rev. Cancer* 6, 857–866
64 Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297
65 Caldas, C. and Brenton, J.D. (2005) Sizing up miRNAs as cancer genes. *Nat. Med.* 11, 712–714
66 Iorio, M.V. *et al.* (2005) MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.* 65, 7065–7070
67 Blenkiron, C. *et al.* (2007) MicroRNA expression profiling of human breast cancer identifies new markers of tumour subtype. *Genome Biol.* 8, R214
68 Ma, L. *et al.* (2007) Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature* 449, 682–688
69 Zhang, L. *et al.* (2006) microRNAs exhibit high frequency genomic alterations in human cancer. *Proc. Natl. Acad. Sci. U. S. A.* 103, 9136–9141
70 Calin, G.A. *et al.* (2007) Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell* 12, 215–229
71 Volinia, S. *et al.* (2006) A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl. Acad. Sci. U. S. A.* 103, 2257–2261
72 Negrini, M. *et al.* (2007) MicroRNAs in human cancer: from research to therapy. *J. Cell Sci.* 120 (Pt 11), 1833–1840
73 Sjöblom, T. *et al.* (2006) The consensus coding sequences of human breast and colorectal cancers. *Science* 314, 268–274
74 Wood, L.D. *et al.* (2007) The genomic landscapes of human breast and colorectal cancers. *Science*
75 Hunter, D.J. *et al.* (2007) A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.* 39, 870–874
76 Stacey, S.N. *et al.* (2007) Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat. Genet.* 39, 865–869
77 Yeager, M. *et al.* (2007) Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.* 39, 645–649
78 Zanke, B.W. *et al.* (2007) Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat. Genet.* 39, 989–994
79 Broderick, P. *et al.* (2007) A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat. Genet.* 39, 1315–1317
80 Easton, D.F. *et al.* (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447, 1087–1093
81 Herman, J.G. and Baylin, S.B. (2003) Gene silencing in cancer in association with promoter hypermethylation. *N. Engl. J. Med.* 349, 2042–2054
82 Jones, P.A. and Baylin, S.B. (2007) The epigenomics of cancer. *Cell* 128, 683–692
83 Esteller, M. (2007) Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat. Rev. Genet.* 8, 286–298
84 Lujambio, A. and Esteller, M. (2007) CpG island hypermethylation of tumor suppressor microRNAs in human cancer. *Cell Cycle* 6, 1455–1459
85 Saito, Y. *et al.* (2006) Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. *Cancer Cell* 9, 435–443
86 Lujambio, A. *et al.* (2007) Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. *Cancer Res.* 67, 1424–1429
87 Esteller, M. (2007) Epigenetic gene silencing in cancer: the DNA hypermethylome. *Hum. Mol. Genet.* 16 (Spec No 1), R50–R59
88 Potti, A. *et al.* (2006) Genomic signatures to guide the use of chemotherapeutics. *Nat. Med.* 12, 1294–1300
89 Hsu, D.S. *et al.* (2007) Pharmacogenomic strategies provide a rational approach to the treatment of cisplatin-resistant patients with advanced cancer. *J. Clin. Oncol.* 25, 4350–4357
90 Huang, F. *et al.* (2007) Identification of candidate molecular markers predicting sensitivity in solid tumors to dasatinib: rationale for patient selection. *Cancer Res.* 67, 2226–2238
91 Nevins, J.R. and Potti, A. (2007) Mining gene expression profiles: expression signatures as cancer phenotypes. *Nat. Rev. Genet.* 8, 601–609
92 Lamb, J. *et al.* (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929–1935
93 Lamb, J. (2007) The Connectivity Map: a new tool for biomedical research. *Nat. Rev. Cancer* 7, 54–60

94 Hieronymus, H. *et al.* (2006) Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell* 10, 321–330
95 Wei, G. *et al.* (2006) Gene expression-based chemical genomics identifies rapamycin as a modulator of MCL1 and glucocorticoid resistance. *Cancer Cell* 10, 331–342
96 Tomlins, S.A. *et al.* (2007) Integrative molecular concept modeling of prostate cancer progression. *Nat. Genet.* 39, 41–51
97 Pujana, M.A. *et al.* (2007) Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nat. Genet.* 39, 1338–1349
98 Araujo, R.P. *et al.* (2007) Proteins, drug targets and the mechanisms they control: the simple truth about complex networks. *Nat. Rev. Drug Discov.* 6, 871–880
99 Petricoin, E.F. *et al.* (2006) The blood peptidome: a higher dimension of information content for cancer biomarker discovery. *Nat. Rev. Cancer* 6, 961–967
100 Petricoin, E.F., 3rd *et al.* (2007) Phosphoprotein pathway mapping: Akt/mammalian target of rapamycin activation is negatively associated with childhood rhabdomyosarcoma survival. *Cancer Res.* 67, 3431–3440