Cellular and Molecular Life Sciences

Molecular profiling of breast cancer: transcriptomic studies and beyond

A. C. Culhane[a,b,*] and J. Howlin[c]

[a] Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, 44 Binney St., Smith 822C, Boston, Massachusetts 02115 (USA)
[b] Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115 (USA),
Fax: +1 617 582 7760, e-mail: aedin@jimmy.harvard.edu
[c] Cell and Experimental Pathology, Lund University, Clinical Research Centre, Malmö University Hospital,
S-205 02 MALMÖ (Sweden), e-mail: jillian.howlin@med.lu.se

Abstract. Utilisation of 'omics' technologies, in particular gene expression profiling, has increased dramatically in recent years. In basic research, high-throughput profiling applications are increasingly used and may now even be considered standard research tools. In the clinic, there is a need for better and more accurate diagnosis, prognosis and treatment response indicators. As such, clinicians have looked to omics technologies for potential biomarkers. These prediction profiling studies have in turn attracted the attention of basic researchers eager to uncover biological mechanisms underlying clinically useful signatures. Here we highlight some of the seminal work establishing the arrival of the omics, in particular transcriptomics, in breast cancer research and discuss a sample of the most current applications. We also discuss the challenges of data analysis and integrated data analysis with emphasis on utilising the current publicly available gene expression datasets. (Part of a Multi-author Review)

Introduction

Increasingly, system-wide analyses are employed on a large scale, whether this be at the DNA – *genomics*, mRNA – *transcriptomics* or protein – *proteomics,* expression level. These approaches form part of a vast effort to create a more detailed view of biological systems (Fig. 1). The most familiar omic technology is without doubt transcriptomics and the study of gene expression profiles [1].

A gene expression microarray 'chip' consists of thousands of DNA molecules attached in fixed locations to a solid surface. These microarrays exploit preferential binding of mRNAs to their complimentary sequences. The abundance of mRNA molecules in a biological sample is assessed by chemically or fluorescently labelling its mRNA extract or cDNA, applying this to the microarray 'chip' and measuring the fluorescence intensities of bound DNA from each location on the array [2]. Although PCR products from cloned cDNA were used in the first microarrays, synthetic oligonucleotides are now widely used in the creation of gene expression microarrays [1, 3]. This technology has been adapted to measure genomic DNA copy number, exon expression and even tiling both coding and non-coding regions of whole genomes. Moreover this idea has expanded and an array may refer to a variety of biological matter immobilised on a solid support, including DNA, protein, antibody and tissue or cell lysates. Although the most widely used solid supports have been glass slides, emerging technologies now also use silicon beads (Illumina) or nanoparticles [4].

Gene expression microarrays require knowledge of the reference human genome to quantify RNA or

---

* Corresponding author.

Technology   Biological event

SNP arrays, aCGH   Copy number, LOH,
next generation seq   mutations, SNPs

ChIP-Chip, ChIP-seq   DNA-Protein
                          Interaction

Gene Expression   mRNA Expression
microarrays, SAGE
next generation seq

Antibody or   Protein
Lysate arrays   Expression
2D gels, Mass Spec

Yeast 2 Hybrid,   Protein-Protein
Lysate arrays   Interaction (PPI)

TMAs   Protein Expression
                         & Localisation

Metabolomics, NMR   Quantification of
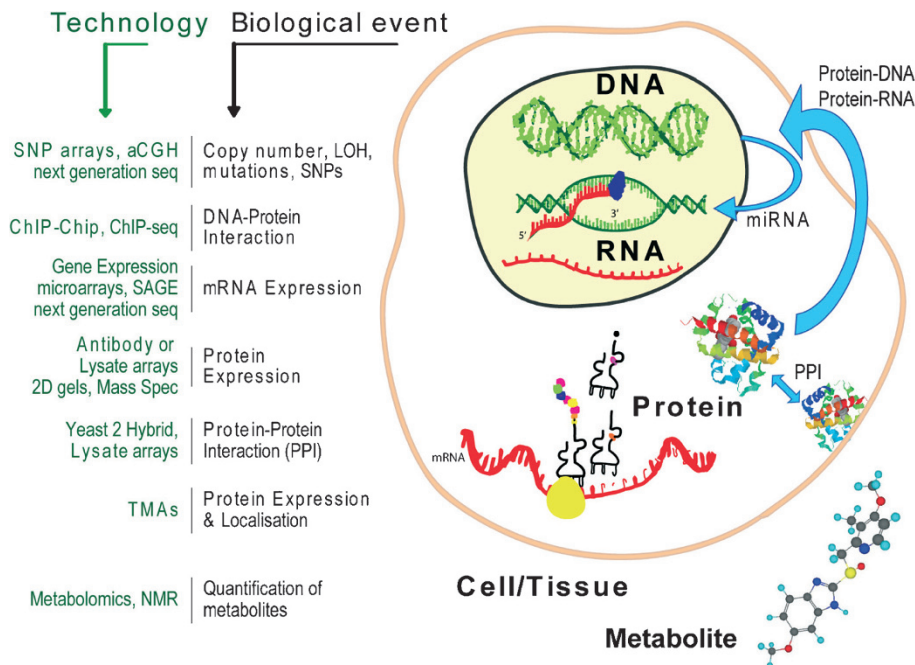   metabolites

Figure 1. High-throughput omics quantifies cellular components on a large or genome-wide scale. *Transcriptomics* is the measurement of mRNA within the cell. Gene expression microarrays (oligonucleotide or cDNA) or next-generation sequencing can be used to measure mRNA. *Genomics* is the quantification of DNA copy number, polymorphism (SNP) or methylation or protein-binding events. *Proteomics* is the analysis of the proteome or protein content of the cell. Each of these approaches forms part of a vast range of techniques now available to quantitatively understand a cell or biological system in a much more detailed manner.

DNA in a biological sample, and as such can be referred to as a 'closed' method. In contrast, another transcriptomics approach, SAGE (serial analysis of gene expression), does not require a pre-existing clone or knowledge of a gene [5], and is able to identify novel transcripts. SAGE is a less high throughput method but has some advantages over gene expression microarrays. For example, while microarrays quantify gene expression levels indirectly using fluorescence intensity, SAGE determines absolute mRNA levels. However, SAGE has been constrained by cost and lower data generation speed. The invention of paired-end (di-tag) sequencing and highly parallel multiplex new-generation sequencing technology from Solexa (Illumina) and 454 (Roche), among others, have reduced these limitations and provided greatly increased throughput. This new generation of sequencing technology has recently been applied to mapping of p53 binding sites in the genome of human colorectal carcinoma cells [6].

In addition to transcriptional aberrations, DNA copy number abnormalities, entire chromosomal or small interstitial DNA losses, gains and amplifications frequently characterise the development of cancer. Analyses of changes occurring at the level of DNA are referred to collectively as genomics. Changes in gene copy number can be measured by array-based comparative genomic hybridisation (array CGH or aCGH), which monitors genomic changes resulting in amplification or deletion and accurately quantifies changes in copy number on a genome-wide scale.

Similar to the case with gene expression analysis, platforms vary and include bacterial artificial chromosome (BAC), cDNA or oligonucleotide formats [7, 8] and platforms initially developed for genotyping. Several studies have shown that variations in gene copy number do not necessarily result in a change in mRNA levels and highlight the importance of combining genomic and transcriptomic data sources in order to derive functional relevance. A recent study by Yao et al., for example, combined cDNA array CGH with SAGE and identified two over-expressed genes in an amplified region of the genome as putative breast cancer oncogenes [9].

Advances in array-printing technology together with the completion of the human genome project have led to generation of microarray chips with significantly more probe sequences per array, providing greater genome coverage. Arrays are available which detect all know exons in the genome and provide an ability to identify gene expression levels, alternative splicing events and mRNA-processing alterations. Single-nucleotide polymorphism (SNP) arrays (Affymetrix, Illumina) containing large numbers of known SNPs are being used to investigate genome-wide patterns of variation. Recently the Illumina Hap300 platform was utilized to study over 300 000 SNPs in 1600 Icelandic individuals with breast cancer and 11 563 controls. Two variants on chromosome 2q35 and 16q12 were reported to be associated with increased breast cancer risk [10]. Such genome-wide analysis can uncover genetic alteration occurring at the level of the DNA within exons of a transcript or surrounding regulatory

regions such as a promoter or, indeed, in non-coding transcribed regions of the genome.

Changes in the sequences of a given gene or genomic region may profoundly affect the level of expression or function of a protein. Interestingly, it has been seen that expression changes in a given gene do not necessarily link to SNPs in that gene transcript or adjacent regulatory regions (working in 'cis'). Rather, many have been found to work in 'trans', meaning that the genomic determinant is located at a distance from the gene it regulates, possibly even on a different chromosome [11, 12]. Transcription factor activity is one obvious way in which one gene may act in trans to regulate another. In the simplest scenario, genetic variation in a transcription factor may alter its ability to initiate transcription of another gene directly. Since the functioning of transcription factors is essential to understanding gene regulation, arrays which cover promoter and intergenic DNA regions of the genome, or even tile the whole genome, are used to map binding sites of transcription factors using 'ChIP on chip' (or ChIP-chip) assays, which are essentially large-scale formats of the original chromatin immunoprecipitation (ChIP) technique [13]. Regions of DNA that interact with a protein of interest are immunoprecipitated using a protein-specific antibody, and the DNA fragments are identified by hybridisation to a microarray. ChIP-Seq is a sequencing-based alternative to ChIP-chip assays. In ChIP-Seq, the ends of the DNA fragments are sequenced, enabling discovery of novel binding sites across the whole genome. Variations on this technique which are not dependent on antibodies also exist, such as the DamID approach, in which a transcription factor of interest is overexpressed as a fusion protein with a methylating enzyme. The methylated DNA is then extracted, amplified, labelled and hybridised to an array. The difference between fusion protein-induced methylation and background methylation with enzyme alone is then analysed [14]. By mapping the actual interactions of protein transcription factor with the genome, one can build a profile of a specific transcription factor's DNA-protein interactome at a given time under certain conditions.

Of course cell processes governing transcription and translation in a cell are complex and may be influenced by epigenetic changes. These DNA modifications, such as methylation, can also be analysed on a genome-wide level by incorporating epigenetic analysis with microarray technology. Gene expression microarrays can be used to identify genes with upregulated expression following demethylation treatment, as these are most likely to have been silenced by promoter hypermethylation [15]. However, CpG island microarrays and methylation-specific oligonucleotide microarrays have also been developed [16, 17].

The resolution and genome coverage of arrays has increased as array technologies have developed. Tiling microarrays, which consist of overlapping probes spanning large genomic regions, have also been used in a variety of contexts, including identification of previously unknown coding and non-coding transcripts, and high-resolution DNA-protein interactions using ChIP-chip protocols and DNA methylation changes. Even higher resolution can be achieved using resequencing arrays which probe (nearly) every base within a region and next-generation sequencing technologies, both of which promise to open new areas of research. For example, a recent study resequenced the exons of 13 023 genes in 11 breast and 11 colorectal cancers and revealed that each tumour accumulated an average of 90 mutant genes, though only a subset of these are likely to be directly implicated in cancer progression. Intriguingly, the majority of these genes were not known to be genetically altered in tumours [18]. As these and other high-throughput techniques become commonplace in genomics laboratories, a systematic characterisation of genetic events that give rise to a cancer cell is rapidly becoming possible.

However, nucleic acid changes, either at the DNA or RNA level, do not provide a complete picture. Binding of a transcription factor to the promoter of a gene may not necessarily result in transcription of that gene; the expression of a gene may not result in translation of a protein or indeed in protein activity. Therefore, we would also like to know the relative protein expression, subcellular location and interaction with other proteins in order to predict or explain a particular biological outcome. To this end, a number of proteomic technologies have been developed to analyse proteins in a high-throughput fashion [19]. Proteomic techniques can be broadly divided into those which are either antibody-dependent or -independent.

Antibody arrays consist of characterised antibodies to known proteins spotted on solid supports, whereas reverse-phase arrays consist of cell lysates or purified proteins and are in turn probed with antibodies [20, 21]. These array-based proteomic approaches are heavily reliant upon specific antibody generation. Discovery of novel proteins or those for which antibodies are unavailable often relies on 2D gel or mass spectrometry proteomics. 2D gel electrophoresis separates proteins in a sample based on their isoelectric properties and mass. The mass, charge and amino acid sequence of variants or differentially expressed proteins can be discovered using mass spectrometry and proteins identified by comparison with peptide signature databases. Part of this process can also be

reduced to chip format using technologies such as Protein Chips (Ciphergen) [22]. Many other proteomic technologies have been developed and have been used to screen samples for autoantibodies, to detect protein:protein/peptide interaction, or to measure protein activity.

Although probably not considered a proteomics approach *per se*, tissue and cell line microarrays have also greatly increased the speed at which protein immunoreactivity can be assessed in tissue specimens or cell lines [19, 23]. A cell line or tissue microarray consists of thousands of cell-line pellets or microcores from tissue sections robotically spotted onto glass slides. Such arrays bring us a step closer to high-throughput validation of candidate genes or proteins by allowing direct measurements *in situ*. However, their success may also be limited by both the availability of antibodies and the reliability of immunohistochemical techniques.

Because so many proteomic approaches depend on antibodies, a recently initiated Swedish project, the Human Protein Atlas project [http://www.proteinatlas.org/] [24], aims to provide antibodies to all human proteins using high-throughput antibody generation [25]. As part of this project, each new antibody is assessed on a number of tissue microarrays containing normal human tissues and tissue from different human cancers. Immunohistological images of these analyses are made available in an online database [24]. The second release of the Human Protein Atlas in October 2006 provided over 1 million images from more than 1500 antibodies. This data will soon be available within the public repository ArrayExpress [personal communication, Dr Alvis Bramza] and will provide a significant resource for the study of the expression and localization of the human proteome in cancer.

Finally, metabolic profiling (especially of urine or blood plasma samples) can be used to detect physiological changes in a cell or system. Metabolomics or metabonomics measures small-molecule intermediate products of metabolism using chromatography, nuclear magnetic resonance (NMR) spectroscopy or mass spectrometry-based technique.

Each of these 'omics' approaches (Fig. 1) described has been applied in one way or another to the study of breast cancer. Several recent reviews have discussed the application of proteomics [19, 26, 27] and metabolomics [28] to the study of breast cancer. Here, we will principally focus on transcriptomic analysis of breast cancer and the integrated or meta-analysis of these data.

Breast cancer model systems

Basic breast cancer research seeks to understand the molecular mechanisms of cancer origin, progression and invasion leading to metastatic disease. Although many studies search for diagnostic or prognostic biomarkers, most studies ultimately seek to develop therapeutic interventions that disturb the workings of the genes, proteins or pathways that are altered in the disease. In both of these broad research classes, 'omic' technologies have produced a dramatic change in how research is conducted. Rather than focusing on a single gene or protein model, we can now look for 'signatures' consisting of multiple genes or proteins that are altered in some way and that together define a molecular phenotype or a particular cancer type or subtype.

As model systems have played such a significant role in our current understanding of cancer, one of the first applications of omic approaches was the molecular characterisation of such models to a level previously unattainable. The NCI-60, a set of 60 human cancer cell lines, including breast cancer cells, selected by the Developmental Therapeutics program of the National Cancer Institute for the purposes of screening anti-cancer agents [29], was among the first to be analysed using omic technologies. They have been extensively profiled for mRNA expression, DNA copy number, DNA methylation profile, mutations and protein expression [30–33]. Attempts to match these cancer cells lines with their putative tissue of origin by examination of the expression profiles successfully demonstrated that cell lines from the same tissue group together. This profiling was also useful in identifying changes that are of functional importance for cancer and not simply indicative of tissue-type variation [34]. Recently a panel of 51 breast cancer cell lines have been defined, and initial characterization indicates that these may provide a useful *in vitro* model system for investigation of breast cancer phenotypes [35].

Because cancer does not develop in isolation, but involves the interaction of the tumour with the surrounding tissue, more advanced models are being developed, including 3D tissue-culture systems that more closely mimic the real *in vivo* situation. When breast epithelial cells are grown on a laminin-rich extracellular matrix, they form polarised acini that are organised and growth-arrested, and are reported to be more representative of normal breast cell behaviour. Fournier and colleagues identified 19 genes that were expressed during organisation and growth arrest of cultured mammary acini, and demonstrated that expression of these genes are associated with good prognosis in breast cancer [36]. These 19 good

prognosis genes were predominately associated with cell cycle and cell division, but a subset also mapped to cytosketal regulation, cell survival and cell-cell interaction processes. These experiments highlight the applicability of the 3D system as a breast cancer model [36], and the prognostic success of these gene signatures suggests that the cellular microenvironment and tumour cell both need to be considered as molecular targets (reviewed by [37]).

The tumour microenvironment contains many cells, including stromal fibroblasts, endothelial cells and infiltrating leukocytes such as macrophages, T lymphocytes, and dendritic cells. These produce and respond to an array of chemokines, cytokines and growth factors, resulting in a complex network of cell-cell interactions which in turn control differentiation, activation, function and survival of multiple cell types in the tumour microenvironment. In an interesting study, Chang and colleagues investigated an hypothesis which likened cellular behaviour during cancer progression and wound healing [38]. To identify a molecular signature of wound healing, they used a transcriptomic approach to identify genes that were expressed by fibroblasts in response to serum exposure. This wound-healing gene expression signature consisted of genes involved in matrix remodelling, cell motility and angiogenesis; the study found these genes were expressed either by the tumour cells themselves, tumour-associated fibroblasts or both. Critically, tumours which expressed this wound-healing gene expression signature were significantly more likely to progress to metastasis and death [38]. The observation that the expression signature of activated cells in the tumour microenvironment is predictive of poor prognosis is significant and has opened new avenues for basic tumour biology research.

Xenograft models of breast cancer are being used to advance our understanding of the propensity of a tumour to metastasize. Weinberg and colleagues [39] profiled mRNA expression levels in a series of isogenic murine breast cancer cell lines (67NR, 168FARN, 4TO7 and 4T1) of differing metastatic ability, revealing a significant increase in the expression of the transcription factor Twist in metastatic variants. The role of Twist was functionally validated using RNA interference, highlighting the ability of transcriptomic studies to identify key targets for functional analysis. Tumours formed by 4T1 cells that expressed *Twist*-siRNA (small interfering RNA) resulted in significantly fewer metastases compared to controls [39].

Transcriptomic approaches have been used to understand the basic biology of breast development and breast cancer cells *in vivo* and the links between them (see O'Connor et al., in this review series and [40]).

Mostly, such work has been carried out in mouse or rat models with a variety of induced genetic alterations. Omic profiling of breast cancer mouse models have also been used to elucidate the mechanism by which a particular genetic alteration or environmental insult leads to an observed phenotype, whether that phenotype produces a developmental defect in the breast or a susceptibility for cancer development [41, 42].

Gene expression profiling has also been applied to define the intrinsic changes which occur during cancer progression in clinical tumour samples. Using a combination of laser capture microdissection and cDNA microarrays, normal epithelium, premalignant, preinvasive and invasive tumours were profiled [43]. Perhaps surprisingly, the results from this study indicated that discrete pathological stages, e.g. ductal carcinoma in situ versus infiltrating ductal carcinoma, were not that dissimilar in terms of their expression profiles. This supports the idea that genes controlling infiltrating or invasive behaviour are active even in preinvasive stages, and is consistent with reports of significant correlation between germ-line mutations or SNPs, and tumour mRNA levels [44]. These observations support the hypothesis that progression of disease may not be the primary factor resulting in genetic rearrangements or expression change, but that the propensity to tumorigenesis is determined much earlier [45].

Molecular Breast Cancer Subtypes

Breast cancer is phenotypically diverse in prognosis and responsiveness to treatment. One of the main observations arising from gene expression studies of breast cancer is that this diversity is reflected in the intrinsic heterogeneity of breast cancer gene expression profiles. Gene expression profiles of breast tumours group into distinct classes. In a seminal study, Perou et al. [46] distinguished four subtypes of breast cancer based on their gene expression profiles. These subtypes were identified using unsupervised hierarchical clustering analysis of cDNA microarray profiles of 8102 human genes in 65 cancer and normal breast samples; these 65 samples were derived from 42 different individuals and included profiles from individuals (n = 10) who provided specimens before and after a 16-week course of doxorubicin chemotherapy. Perou et al. [46] observed that gene expression patterns in pairs of tumour samples from the same individual were almost always more similar to each other than either was to any other sample. Therefore, they defined a set of genes whose variation was significantly greater between samples from different tumours than between samples from the same tumour

before and after treatment. This set of 496 *intrinsic genes* clustered the breast tumour profiles into four groups, which they designed as *basal-like, Erb-B2+, normal breast-like,* and *luminal epithelial/ER+* [46]. The *luminal* and *basal* clusters were defined based on their expression of keratin genes, and were not specifically associated with epithelial cell lineages. The *luminal* group formed the largest cluster, and the expression profiles of these tumours were characterized by the expression of estrogen receptor (ESR1), keratins 8/18, GATA3 and several transcription factors. Tumours which were negative for ESR1 expression segregated into two distinct clusters, *basal-like* and *ERBB2+*. The *basal* gene expression cluster expressed genes which are considered characteristic of basal epithelial cells including keratin 5/6, keratin 17, integrin b4 and laminin. The *ERBB2+* cluster expressed genes associated with expression of the Erb-B2 oncogene and GRB7. Finally, a few tumours, a fibroadenoma and the normal breast samples formed a cluster designated *normal-like*.

In a follow-up study using the same cDNA microarray platform, hierarchical clustering of these same 496 genes in 85 breast cancer, fibroadenoma and normal breast tissues also produced three ESR1-negative groups (*basal epithelial-like*, *ERBB2-overexpressing, normal-like*). However the luminal/ESR1-positive cluster could be further subdivided into three groups, *luminal A*, *B* and *C* [47]. Two years later these subtypes were further refined in a study that combined 84 of 85 of these cDNA arrays with a new set of 38 cDNA array gene expression profiles [48]. This investigation identified a set of *534 intrinsic* genes and defined five subtypes: *basal-like*, *normal-like*, *ERBB2+*, and two *luminal* subtypes, *A* and *B* [48]. Luminal A tumours had a higher expression of ESR1-related genes (such as GATA3) and lower expression of proliferative genes. By contrast, luminal B tumours expressed more proliferative genes. These subgroupings or clusters were independent of stage and grade [46, 47]. Significantly, the clinical importance and distinction between these clusters was supported by their associated prognosis [49–51]. Survival analyses showed different outcomes for each subtype. The luminal tumours were associated with more favourable outcome, whilst the basal-like subtype was associated with poor prognosis [46–48]. Women with BRCA1 mutations are more likely to have triple negative hormone receptor status, *i.e.* negative for ESR1, progesterone receptor (PGR) and ERBB2, and a gene expression profile correlated to the basal subtype of breast cancer [48]. Furthermore, these gene expression subtypes may be predictive of chemotherapeutic response. It is reported that the basal-like and ERBB2+ subtypes are more sensitive to paclitaxel- and doxorubicin-containing preoperative chemotherapy than the luminal and normal-like cancers [52]. These intrinsic subtypes are discussed in more detail in the accompanying review in this series by Mullan and Milikan.

Numerous studies have now confirmed the existence of breast cancer molecular subtypes; however, the number of intrinsic subtypes remains to be resolved (Table 1). A recent meta-analysis of 599 microarrays from five separate cDNA microarray studies of breast cancer found support for only three subtypes of breast cancer [53]. These three subtypes were defined by their ESR1/ERBB2 status. These subtypes were ESR1+/ERBB2 (group 1), ERBB2+ (group 2) and ESR1–/ERBB2– (group 3), and although similar, did not exactly match the intrinsic subtypes defined by the Sørlie et al. (2003) classification [48]. The *normal-like*, *luminal A* and most of *luminal B* samples were classified to group 1 (ESR1+/ERBB2– subtype). Most of intrinsic ERBB2+ samples [48] were classified in the group 2 subtype [53]. Group 3 captured only a subset of intrinsic basal subtype [48]. The remainder of intrinsic basal samples were distributed among the three ESR1/ERBB2 subtypes. Whether these three subtypes can be further subdivided into four, five, or six subtypes remains to be confirmed.

Genomic and proteomic studies reinforce the robustness of the molecular taxonomy that emerged from these gene expression microarray studies. Genome-wide aCGH of 89 breast tumours has demonstrated that different DNA copy number alterations are associated with each molecular subtype [54]. Not surprisingly, ERBB2 subtype tumours exhibited more frequent amplification at 17q12-q21, a region which harbours the ERBB2 gene. Higher numbers of gains and losses were associated with basal-like tumours. For example, loss on 5q was associated with high-grade, ESR1-negative, TP53-mutant and basal-like breast tumours, whereas high levels of DNA amplification, including more frequent gains on 8q, were observed in luminal B tumours [54], which have high proliferation. Interestingly, the oncogene MYC resides on 8q and may be implicated in higher levels of cell proliferation in these tumours. The observation that different DNA copy number alterations are associated with different subtypes may suggest that different genomic instability mechanisms are implicated in the pathogenesis of different subtypes. The existence of three ESR1/ERBB2 subtypes is supported in immunohistological studies [19, 49] (please refer to the accompanying review in this series by Mullan and Milikan).

In order for this molecular subtyping to be translated into a clinically useful assay, a definitive gene signature with predictive accuracy needs to be

Table 1. Gene Expression Studies of Intrinsic Breast Cancer Subtypes.

Breast cancer subtypes								Number of features (probes)	Number of arrays	Microarray platform	References
Basal	Basal II	Normal-like	ERBB2+	Luminal A	Luminal B	Luminal C	IFN				
√		√	√	√				496	84	8K cDNA microarrays	[45]
√		√	√	√	√	√		456	85	8K cDNA microarrays	[46]
√		√	√	√	√			552	122	42K, 23K, 8K cDNA microarrays	[47]
√	√		√	√	√	√		706	99	7.6K cDNA microarrays	[50]
√		√	√					367	126	Affymetrix U133A	[118]
√		√	√	√				37	123	Agilent A1, A2, custom qRT-PCR	[54]
√				√				54	20	ABI, Stanford cDNA, Agilent	[49]
√		√	√	√	√		√	1410	146	Agilent 1Av1, 1Av2 and custum	[100]

defined, and tested. To date such a set of intrinsic genes has yet to be defined and intrinsic gene expression lists of between 133 and 1300 genes are reported (Table 1). A qRT-PCR assay with a set of 59 genes that stratifies breast tumours into molecular subtypes [55] is promising, but the robustness of this signature must be confirmed. Further studies on larger cohorts of patients such as the EORTC-MINDACT (Microarray In Node negative Disease may Avoid ChemoTherapy) Trial are under way. Future analyses of these data together with meta-analyses of public data will lead to refinements in our understanding of breast cancer subtypes and provide more precise gene signatures that are predictive of intrinsic breast cancer molecular subtypes.

Diagnosis and prognosis

There has been an increased focus on the implementation of rational molecular targeted therapy for cancer [56]. The failure of current markers and prognostic indicators to accurately predict the patients that would eventually succumb to fatal meta-static disease without intervention, and those whose disease course could not be altered, prompted clinicians to look for something new to try. Avoidance of aggressive adjuvant therapy for patients who ultimately would not benefit is highly desirable, but current pathological markers do not have the ability to classify such patients based on initial diagnosis. Currently, therapies based on molecular diagnosis are limited to hormonal based chemotherapy such as tamoxifen for estrogen-responsive tumours or treat-

ment of ERBB2/Her2-positive tumours with Herceptin. The hope for omics in the clinic is to ultimately tie a molecular classification and signature to a specific treatment regime.

While unsupervised clustering of gene expression profiles has been used to identify intrinsic molecular subtypes, supervised analysis of breast cancer gene expression profiles has been used to find genes associated with clinically important prognostic information such as grade [57] and histology [58]. A number of gene expression signature assays developed from microarray studies of breast cancer are already commercially available. These include the 70-gene prognosis signature MammaPrint [59] and the 21-gene Oncotype DX [60]. Both of these are designed to be applied in selection of adjuvant therapy for individual patients. Genomic Health released the Oncotype DX Recurrence Score gene signature of breast cancer prognosis test in 2004. That test tracks a 21 gene expression signature [60, 61] using qRT-PCR and can be performed on formalin-fixed, paraffin-embedded tumour tissue. The US Food and Drug Administration (FDA) approved the 70-gene signature, MammaPrint, developed by Agendia (Amsterdam, Netherlands) as the first *in vitro* diagnostic multivariate index assay device. It will be marketed as a test that distinguishes lymph node-negative breast cancer patients who would benefit from additional therapy and those who would not. Though the prognostic ability of the 70-gene signature has been supported by subsequent studies from this research team [62, 63], concerns regarding the design and statistical analysis used to derive the original 70-gene signature have been raised [64–66]. A diagnostic test

is also being developed by Veridex, LLC, using a set of 76 genes that were identified in a study of gene expression profiles of 286 lymph node-negative patients who had not received adjuvant systemic treatment [67, 68].

Large prospective studies have been initiated to determine the accuracy of test predictions from these assays. The ongoing European MINDACT clinical trial will examine the MammaPrint test and may also prospectively evaluate the 76-gene Veridex signature. The TAILORx study, which is sponsored by the US National Cancer Institute and led by the Eastern Cooperative Oncology Group, will test the Oncotype DX Breast Cancer Assay. The TAILORx study plans to enrol more than 10000 women with hormone-positive (ESR1+ and/or PgR+), ERBB2− breast cancer that has not spread to the lymph nodes and will use the Oncotype DX array to determine which women will receive adjuvant chemotherapy in addition to hormone therapy. Initial studies of the 21-gene Oncotype DX signature are inconclusive. RT-PCR analysis of the Oncotype DX 21 genes in archived material from 4964 lymph node-negative breast cancers that were not treated with adjuvant chemotherapy reported that it was associated with risk of breast cancer-specific mortality among ER-positive, tamoxifen-treated patients [69]. However, another study at MD Anderson Cancer Center reported no clear association between the Recurrence Score and risk for distant recurrence in 149 patients who were not treated with adjuvant therapy [70]. Further studies are required to establish the clinical usefulness of this assay.

Whilst these assays may help in decision making regarding whether patients should undergo adjuvant therapy, molecular tests are also required for prediction of risk of local or distant recurrence. To address this issue Kreike et al., used cDNA microarrays, to study gene expression profiles of pairs of primary tumours and their recurrences [71]. They report that primary tumours and their recurrences have highly correlated gene expression profiles. Neither unsupervised or supervised analysis of gene expression profiles could reveal significant differences in gene expression profiles between recurring and non-recurring primary breast tumours [71].

Similarly, Nuyten et al., were unable to recover gene predictors of recurrence in a recent study in which they applied a supervised analysis approach to identify predictors of recurrence in a set of 161 gene expression profiles of patients [62] with stage I or II breast cancer who were treated with complete surgical excision of the tumour followed by whole breast irradiation (breast-conserving therapy) at the Netherlands Cancer Institute between 1984 and 1995 [72]. They report

also that other published prognostic signatures, including MammaPrint [59], did not predict recurrence in these data. The most successful predictor of local recurrence identified in this study was the Chang et al. 2004 wound-healing signature [38], which correctly predicted local recurrence in 7 out of 8 cases of the test dataset (88 % sensitivity) with a specificity of 74 % (53 out of 72) [72]. However, the number of local recurrences in the dataset was low (17/161), and these results need to be confirmed in a larger study [62].

Breast tumours normally metastasize to bone, lung or other visceral sites. To investigate lung-specific metastasis, Minn et al. [73] generated a xenograft model from a derivative of the MDAMB-231 cell line which was selected to be highly metastatic to lung, and using transcriptomic analysis identified several genes whose altered expression constituted a 54-gene Lung Metastatic Signature (LMS). When cross-validated on a cohort of 82 patients, the LMS predicted patients at high risk for selective distant relapse to lung, but not to bone. Recently, this group has supported this LMS in a study on a larger cohort of 738 primary tumours [74] in which they show that tumours which expressed the LMS, and consequently were more likely to relapse in lung than in bone, liver or pleura, were mostly ESR1-negative (73 %), grade 3 (69 %), and the majority belonged to basal (65 %) or ERBB2 (19 %) subtype [74]. Interestingly, tumours which were LMS-positive were also likely to be predicted as poor prognosis using the MammaPrint signature (92 %) [59], Chang et al. wound response signature [38, 75] or the Oncotype DX Recurrence Score [60, 61].

Serious methodological issues, including small or biased cohorts of patients, inappropriate statistical analysis or validation are evident in landmark papers on breast cancer expression profiling and need to be addressed before these signatures are adopted as clinical tools [27, 76–78]. The transition from an observation that gene expression profiles can predict clinical outcome to actual routine use of such profiles is a long process, and success in well-designed clinical trials must be demonstrated. A disappointing finding in this regard was the fact that gene expression measurements may not necessarily perform better than conventional markers as reported by Eden et al. 2004 [79, 80]. These issues together with problems relating to design and analysis of results of the MINDACT clinical trial are described in detail in several recent papers [76, 77, 81].

Omics for drug discovery and therapeutics

Ultimately, even if the newly defined breast cancer signatures and subtypes are not quite ready for the clinic, they still have an enormous potential to contribute to the drug discovery process. This is by virtue of the fact that they can be used to identify potential new rational molecular targets specific for individual cancer subtypes. One example of this may be the observation that the triple receptor negative (ESR1, PGR, ERBB2) basal cancer subtype (see above) may in fact be amenable to anti-EGFR targeted therapies, as most of this subgroup express EGFR, a fact previously unappreciated [82]. This provides at least a possible molecular target in a subgroup that has few satisfactory treatment options at present.

The promise of connecting a cancer subtype to one gene or protein drug target with efficacious compounds is oversimplified as an ultimate objective. However, the goal of individualised therapy tailored to specific patients' phenotypes is enticing. Current treatment of breast cancer is limited to a small range of options. Following biopsy and/or surgery, patients may be treated with antihormonal or growth factor drugs either alone or in combination with conventional cytotoxic chemotherapy and possibly radiation. These treatment options fall short of ideal and can lead to overtreatment of those who either do not need or will not benefit from adjuvant therapy.

In response to this need for increased individualised therapy, increased understanding of drug mechanisms of action is required. As a result there have been many developments to optimize genomic-scale readouts applicable to drug development. High-throughput omics technologies have been applied to screening of new breast cancer drugs, elucidating the molecular mechanism of drug action, predicting response and potential harmful effects. Traditional drug discovery processes have been revolutionised by high-throughput approaches; now genome-wide high-throughput approaches are being used to facilitate pharmacological studies *in silico*. Toxicological screens have been performed using traditional gene arrays and also so called tox-chips, which contain only a subset of genes most likely to be involved in various toxic responses [83, 84]. These can be used both as rapid screens and as a means of generating patterns of expression that may help identify mechanisms of toxicological action. This approach has been successful in identifying a mechanism of drug action which was previously unknown; the response to 5-fluorouracil (5-FU), used to treat breast cancer and colorectal cancer, was found to directly correlate to the levels of DPYD (dihydropyrimidine dehydrogenase) mRNA. Tumours with high levels of this enzyme showed increased resistance to 5-FU owing to the fact that this enzyme catalyses the breakdown of 5-FU [85].

Successful chemotherapy can in many cases lead to a significantly better outcome after breast cancer surgery; however, intrinsic or acquired drug resistance can reduce potential responders by up to 50% [86]. Identification of molecular markers of drug resistance is one approach to uncovering common mechanisms of drug resistance and finding ways to circumvent them. Distinct gene expression and DNA methylation profiles associated with acquired resistance to fulvestrant or tamoxifen were described by Fan et al. (2006). These authors demonstrated that different molecular changes result from resistance to different antiestrogens. Most notably, there was dramatically reduced expression of ESR1 in cells refractory to estrogen treatment in fulvestrant-resistant lines and maintenance of functional ESR1 but altered signalling patterns in tamoxifen-resistant cells [87].

Omic approaches have also been applied to increase our understanding of response to radiation therapy. Radiation-induced gene or protein expression changes have the potential to identify radiation-specific genes that can be used as biomarkers to assess radiation exposure and to probe the mechanism of cellular responses to radiation. In 1999, Amundsen et al. discovered various transcripts not previously known to be induced by radiation in MCF-7 breast cancer cells. Additionally, they demonstrated that FRA1 was involved in p53-mediated radiation response by virtue of its upregulation following radiation treatment only in breast cancer cells with wild-type p53 [88]. Park et al. (2002) developed a radiation-specific cDNA array or Rad Chip of 384 genes, based on gene expression responses to gamma irradiation. From this defined set it was then possible to distinguish different types of genotoxic stress and detect radiation exposure *ex vivo* [89]. As with drug compounds, radio resistance can become a problem for certain tumours, and omic approaches have also been used to define gene expression responses indicative of resistance in order to better understand the mechanism and to develop strategies to overcome it [90, 91]. Identifying gene targets for radiosensitisation and/or chemosensitisation is an important strategy in improving anticancer treatments. Ideally, we will get to a stage where our understanding of drug action allows us to tailor therapeutic regimens so that each patient receives the maximum benefit based on the omic characteristics of their particular disease.

Genes which contribute to gene expression signatures of breast cancer

As discussed previously, there have been a number of studies that have produced gene expression signatures that characterise different breast cancer phenotypes. If one looks at these in detail, the overlap in the gene lists comprising each signature is generally small, causing some to question their validity. For example, there are only 3 genes in common between the 70-gene MammaPrint and the 76-gene Rotterdam signature [59,68]. Even the genes in the various intrinsic breast cancer molecular subtype signatures vary between studies (Table 1). However, a recent study that examined different gene expression signatures applied to a single dataset of 295 samples [92] found similar patient outcome was predicted using the intrinsic gene signature or many different prognostic gene classifiers [38, 59, 60, 75]. Most tumours with a basal-like, ERRB2+ or luminal B subtypes were classified as having a poor 70-gene MammaPrint profile signature [59], activated wound response gene signature [38] and high Recurrence Score [60, 92] while all luminal A subtype tumours were predicted to have a good outcome using the 70-gene profile [59, 92]. Overall, there was a high level of concordance in prognosis prediction even though the overlap in gene signatures was minimal. In many ways, this is not unexpected and can be partly explained by the highly correlated nature of gene expression. Because many genes interact in overlapping pathways and networks in producing a phenotype, correlations reflecting their co-regulation allow one to select several combinations, each of which has similar predictive power. The observation that many genes (Table 1) contribute to the intrinsic subtypes and breast cancer signatures may reflect different transcriptional or epigenetic processes.

The predominant phenotypic distinction in breast cancer is between ESR1+ and ESR1− and this is evident in the intrinsic breast cancer molecular subtypes and in prognostic gene predictors. Each intrinsic gene list includes a significant number of genes that are responsive to estrogen, and overlaps with a set of 822 estrogen-responsive genes that were identified by exposing ESR1+ MCF-7 breast cancer cell line to 17-beta-estradiol [93]. Perhaps unsurprisingly, hormone receptor-positive (ESR+) breast cancer patients who are treated with tamoxifen have a better prognosis if they express estrogen-responsive genes [93].

Other genes within the intrinsic list are genes associated with proliferation (CENF, BIRC5, BUB1, GTPBP4, TTK) and have a high correlation with tumour grade and poor patient prognosis [57,93]. In a study of gene expression profiles of 189 invasive breast carcinomas, a set of 97 genes were identified that strongly associated with the distinction between grade 1 and 3. Most of these 97 were involved in cell cycle regulation and proliferation [57]. This set of genes was validated on an independent set of 597 expression profiles from previously published studies, and it was demonstrated that high expression levels for these genes were associated with a higher risk of recurrence. Proliferation-associated genes are a significant component of both the Oncotype Dx [60] and MammaPrint 70-gene prognosis signature [59]. Furthermore, proliferation-associated genes also form a major part (14/53) of the recently described qRT-PCR intrinsic gene list [55].

Other studies have examined which genes are expressed in tumour stroma and in tumour cells exposed to different environmental conditions. Bacac et al. examined gene expression profiles of laser capture microdissected murine stromal cells derived from an animal model of non-invasive and invasive prostate cancer. Investigation of human orthologues of these murine stromal genes showed they were also expressed in invasive human cancer and were associated with poor prognosis in prostate and breast cancer [94]. Two recent studies have derived significant predictors of prognosis of breast tumour by examining the genes that are induced when cells are grown *in vitro* under different conditions [36, 95]. Park et al. identified a set of genes that were differentially expressed between transplanted mouse plasma cell tumor tissue and cell lines derived from them [95]. Genes induced when mouse plasma cell tumours are first grown *in vivo* then *in vitro* were associated with ESR expression, grade, the 70-gene prognosis signature, the wound-healing signature and ERBB2+ genes of human breast cancer [95]. Network analysis of pathways in which these genes are expressed implicated genes activated by EGF (epidermal growth factor), IFNγ, IL-4, cyclin A2, TNF and AKT1 in the poor prognosis group. In contrast, genes that were activated by PTGS2 (COX2) had lower expression in poor prognosis tumours [95].

Experimental challenges in breast cancer transcriptomics

Although microarrays and other omic technologies allow the generation of vast quantities of data on each individual sample, the wealth of data they provide does not mean that one can ignore the basics of good experimental design. Many early studies did not include sufficient replication to estimate and account for experimental and biological variance and such replication is necessary to assure that the signatures

that are found can be translated from the laboratory to the clinic. Replicates of each covariate are essential. A covariate is a variable that is potentially predictive of the outcome under study. Whilst it is hoped that these predictive covariates are biologically or clinically relevant, sometimes covariates may explain experimental variance (e.g. sample batch date). Analyses should be designed so as to minimize confounding covariates. For example, if all ESR1+ cases were hybridized on a specific date, then it would be impossible to determine whether a dependent effect was due to ESR1 status or a batch effect. This hidden covariate is called confounding. Sampling, mRNA extractions, microarray hybridisation batches and other experimental procedures should be randomised to minimise the chance of this occurring.

But other elements of sound experimental design cannot be ignored either. For example, high-quality, reproducible arrays and optimised, standardised protocols should be used. Similarly, high-quality RNA extracts should be used and the quality confirmed before labelling and hybridisation. Poor-quality RNA or hybridisation assays failing to meet quality standards should be excluded from analysis. One of the key steps in data analysis is normalisation, which adjusts the measurements across arrays to allow effective comparisons. Many normalisation algorithms assume that either the sum or the mean of gene expression should be equal across all arrays, but outlying or poor-quality data can skew the results.

Another often ignored element of good omic analysis is the use of a well-designed data management system to track information regarding clinical samples and laboratory processes. This is essential not only for day-to-day laboratory management and quality control, but also to effectively link clinical information necessary for analysis of the data that are ultimately produced. For both legislative (Health Insurance Portability and Accountability compliance) and ethical reasons it is important that such a sample annotation database or data management system provide a confidential and secure process for de-identifying or anonymising subject records. Many laboratories have found that commercial Laboratory Information Management Systems (LIMS) software is designed for large-scale or institutional use and cannot afford the high investment of time and cost that these require. However, groups that neglect the data management aspect of an omic study and implement poorly managed in-house databases with inadequate access to subject and sample information considerably limit the analysis that can be performed on experimental data. Indeed, due to the high dimensional nature of omic data and the fact that sample parameters may be confounded, insufficient metadata (such

as incomplete clinical information) may significantly bias interpretation. This may result in a study that produces a large amount of data that was expensive to generate but reveals no significant biological insights. An open source, lightweight, simple Web-based tool, PASSIM, that fulfils many of the above criteria and may meet the needs of research projects was recently developed by the microarray group at the European Bioinformatics Institute [96].

A complete description of methods for the analysis of microarrays is beyond the scope of this article, but it has been discussed in depth in several other reviews [97, 98]. Several early microarray gene expression studies suffered from poor design, confounding covariates and insufficient or overfitting during classifier cross-validation [65]. Different feature selection approaches, even when applied to the same dataset, can produce different gene lists [99], and similarly different classification algorithms may result in different classifier success. Consequently, the findings from high-throughput omic studies need to be validated. Gene expression studies are now frequently supported with qRT-PCR data, and increasingly tissue microarrays are used. Data mining of published microarray studies represents another approach to validate a selected gene signature [48, 100, 101], although this requires that public databases include appropriate clinical information on each sample so that valid conclusions can be drawn.

Integrated data analysis – gene expression microarrays

Several public repositories have been established to collect the publicly available microarray data – Gene Expression Omnibus (GEO) at the NCBI [102], Center for Information Biology gene EXpression (CIBEX) in DDBJ, Japan [103], and ArrayExpress at the European Bioinformatics Institute [104]. A researcher can query these databases for all experiments of a given type and retrieve the respective data, which then can be combined with the researcher's own data or used for designing new experiments. These repositories represent a considerable data resource. The public microarray repositories ArrayExpress and the GEO now contain over 100 000 microarray gene expression profiles (Table 2). With growing amounts of microarray data, improving quality and new analysis methods being developed, these form a rich resource for both data mining and validation of omic studies of breast cancer (reviewed by [105]). It is likely that these data will generate findings and produce new information long after the experiments have been completed. Several new gene lists have been gener-

Table 2. Omics data in public repositories*.

Repository	Experiments	Number of arrays	URL
ArrayExpress	2077	67 863	http://www.ebi.ac.uk/arrayexpress/
GEO	5614	141 830	http://www.ncbi.nlm.nih.gov/geo/
CIBEX	10	483	http://cibex.nig.ac.jp/
SMD		12 385	http://smd.stanford.edu/

GEO, Gene Expression Omnibus; SMD, Stanford Microarray Database; CIBEX, An array is a single-microarray hybridisation; an experiment is a collection of arrays. Counts compiled May 2007. ArrayExpress statistics based on monthly report April 2007. SMD is a repository for researchers at Stanford and their collaborators. SMD holds > 66 000 arrays; however, much of this data is private. Only public data in SMD listed.

ated from reanalysis of published data. For example, CENPF was recently identified as a marker of poor prognosis and chromosomal instability in a reanalysis of previously published data [59], and the significance of this gene was confirmed using tissue microarray analysis [64].

Datasets can also be combined in meta analysis. The average number of arrays per study is only 30–40 (Table 2). Given that the number of features (genes) on microarrays now regularly exceeds 50 000, this presents a considerable dimensionality problem. Since the cost and availability of biomaterial, such biopsy tissue, limits the number of samples analysed, this low case-to-feature ratio is likely to remain an issue. As a result, meta-analysis, or merging data from multiple studies, is an attractive option. Meta-analysis of many gene expression datasets has been applied to obtain more robust breast cancer gene signatures [101] and to define the number of well-supported intrinsic molecular subtypes of breast cancer [53]. Simple methods, such as *co-inertia analysis*, can be used to compare the global correlation between gene expression profiles of the same tissues or cell lines obtained in different studies, even if these studies have used arrays with different catalogues and numbers of genes [106].

Matching of genes or DNA probes across microarray platforms remains challenging. Older or custom microarrays may contain different subsets of the genome. Even when whole-genome arrays are used, microarray platforms may not be directly comparable. Probes may be designed to different gene regions or splice variants. Moreover, the quality of probe design may vary. Whilst older cross-platform analysis of microarray data showed poor concordance, more rigorous matching of probes greatly improved correlation across platforms [107]. Several reports have demonstrated that cross-platform correlation is increased when DNA probes are sequence-matched [108, 109]. Matched sets of genes across different microarray studies can be combined in a meta-analysis using simple non-parametric rank statistics [110] or a Bayesian framework [111]. Bayesian methods are

particularly attractive as they use prior knowledge (or assumptions) about relative accuracies of data sources, and their utility has been demonstrated in an integrated analysis of four breast cancer datasets [112]. Shen et al. applied a two-stage Bayesian mixture modelling approach to analyse four independent microarray studies obtained on different technological platforms (n = 305 samples) and identified a 90-gene meta-interstudy signature that was predictive of survival in breast cancer patients [112]. Much information can be obtained by comparing gene profiles across studies or even between species. For instance, comparison of the expression profiles of homologues across a range of organisms can help identify orthologous genes [113], or even orthologous processes such as the cell cycle in different organisms. Furthermore, gene signatures observed using animal models have been validated using human clinical expression profiles [94,95]. Increasingly, studies integrate microarray data with data from other omic technologies [9, 114], and software algorithms and tools are being developed to support such a systems approach. New integrated meta-analysis methods are being developed and applied to many types of omic data from ongoing large clinical trials and other studies, including transcriptomic, proteomics [115] and metabolomic data, as well as others. These integrated datasets and their analyses promise to yield a more complete understanding and insight than can be obtained from any single approach.

Conclusions and perspectives

In the past decade, high-throughput biology and gene expression profiling have matured considerably. The technology has developed rapidly, and these have been complemented by better experimental design and laboratory protocols, statistical tools, data standards, public data repositories and analysis software. Whilst considerable developments in all of these areas are still required, the development of commercial gene signature assays such as MammaPrint and

Oncotype DX demonstrates the potential of omic studies. These gene signatures were derived when microarrays were a relatively new and immature technology. Whilst the ultimate success of these assays remains to be determined, through several ongoing large-scale trials, the next generation of omic signatures is already being developed using more advanced technology and with computational methods that utilise integrated approaches.

In the context of breast cancer, new omic studies are likely to provide unprecedented capabilities to derive biomarkers of prognosis and drug response. These may be combined with, new less-invasive approaches for detection and monitoring of breast tissue. Currently, most breast tumours are detected in breast examination and mammography, and if abnormalities are found, these are followed up with invasive needle or surgical biopsy. Less-invasive approaches would be a significant advance, and the use of fine-needle aspirate, nipple aspirate and ductal lavage fluid is being assessed; recent studies have used microarrays and proteomics to identify ESR1-positive and -negative molecular differences and biomarkers in nipple aspirate samples [116–118]. For patients, the feasibility of developing omic applications that rely on such less-invasive approaches would represent a particularly important advance, as they may encourage more women to be screened more often and to allow earlier detection, when survival and response to therapy are optimal.

1   Lipshutz, R. J., Fodor, S. P., Gingeras, T. R. and Lockhart, D. J. (1999) High density synthetic oligonucleotide arrays. Nat. Genet. 21, 20–24.

2   Duggan, D. J., Bittner, M., Chen, Y., Meltzer, P. and Trent, J. M. (1999) Expression profiling using cDNA microarrays. Nat. Genet. 21, 10–14.

3   Cheung, V. G., Morley, M., Aguilar, F., Massimi, A., Kucherlapati, R. and Childs, G. (1999) Making and reading microarrays. Nat. Genet. 21, 15–19.

4   Yezhelyev, M. V., Gao, X., Xing, Y., Al-Hajj, A., Nie, S. and O'Regan, R. M. (2006) Emerging use of nanoparticles in diagnosis and treatment of breast cancer. Lancet Oncol. 7, 657–667.

5   Porter, D., Yao, J. and Polyak, K. (2006) SAGE and related approaches for cancer target identification. Drug Discov. Today. 11, 110–118.

6   Ng, P., Tan, J. J., Ooi, H. S., Lee, Y. L., Chiu, K. P., Fullwood, M. J., Srinivasan, K. G., Perbost, C., Du, L., Sung, W. K., Wei, C. L. et al. (2006) Multiplex sequencing of paired-end ditags (MS-PET): a strategy for the ultra-high-throughput analysis of transcriptomes and genomes. Nucleic Acids Res. 34, e84.

7   Mantripragada, K. K., Buckley, P. G., de Stahl, T. D. and Dumanski, J. P. (2004) Genomic microarrays in the spotlight. Trends Genet. 20, 87–94.

8   Albertson, D. G. and Pinkel, D. (2003) Genomic microarrays in human genetic disease and cancer. Hum. Mol. Genet. 12 Spec. No. 2, R145–152.

9   Yao, J., Weremowicz, S., Feng, B., Gentleman, R. C., Marks, J. R., Gelman, R., Brennan, C. and Polyak, K. (2006) Combined cDNA array comparative genomic hybridization and serial analysis of gene expression analysis of breast tumor progression. Cancer Res. 66, 4065–4078.

10 Stacey, S. N., Manolescu, A., Sulem, P., Rafnar, T., Gudmundsson, J., Gudjonsson, S. A., Masson, G., Jakobsdottir, M., Thorlacius, S., Helgason A. et al. (2007) Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat. Genet. 39, 865–869.

11 Morley, M., Molony, C. M., Weber, T. M., Devlin, J. L., Ewens, K. G., Spielman, R. S. and Cheung, V. G. (2004) Genetic analysis of genome-wide variation in human gene expression. Nature 430, 743–747.

12 ENCODE Project Consortium (2007) Identification and analysis of functional elements in 1 % of the human genome by the ENCODE pilot project. Nature 447, 799–816.

13 Das, P. M., Ramachandran, K., vanWert, J. and Singal, R. (2004) Chromatin immunoprecipitation assay. Biotechniques 37, 961–969.

14 Bulyk, M. L. (2006) DNA microarray technologies for measuring protein-DNA interactions. Curr. Opin. Biotechnol. 17, 422–430.

15 Suzuki, H., Gabrielson, E., Chen, W., Anbazhagan, R., van Engeland, M., Weijenberg, M. P., Herman, J. G. and Baylin, S. B. (2002) A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer. Nat. Genet. 31, 141–149.

16 Yan, P. S., Chen, C. M., Shi, H., Rahmatpanah, F., Wei, S. H., Caldwell, C. W. and Huang, T. H. (2001) Dissecting complex epigenetic alterations in breast cancer using CpG island microarrays. Cancer Res. 61, 8375–8380.

17 Gitan, R. S., Shi, H., Chen, C. M., Yan, P. S. and Huang, T. H. (2002) Methylation-specific oligonucleotide microarray: a new potential for high-throughput methylation analysis. Genome Res. 12, 158–164.

18 Sjoblom, T., Jones, S., Wood, L. D., Parsons, D. W., Lin, J., Barber, T. D., Mandelker, D., Leary, R. J., Ptak, J., Silliman, N. et al. (2006) The consensus coding sequences of human breast and colorectal cancers. Science 314, 268–274.

19 Bertucci, F., Birnbaum, D. and Goncalves, A. (2006) Proteomics of breast cancer: principles and potential clinical applications. Mol. Cell. Proteomics 5, 1772–1786.

20 MacBeath, G. and Schreiber, S. L. (2000) Printing proteins as microarrays for high-throughput function determination. Science 289, 1760–1763.

21 Ivanov, S. S., Chung, A. S., Yuan, Z. L., Guan, Y. J., Sachs, K. V., Reichner, J. S. and Chin, Y. E. (2004) Antibodies immobilized as arrays to profile protein post-translational modifications in mammalian cells. Mol. Cell. Proteomics 3, 788–795.

22 Keller, E. T. and Yao, Z. (2002) Applications of high-throughput methods to cancer metastases. J. Musculoskelet. Neuronal Interact. 2, 575–578.

23 Hoos, A. and Cordon-Cardo, C. (2001) Tissue microarray profiling of cancer specimens and cell lines: opportunities and limitations. Lab. Invest. 81, 1331–1338.

24 Nilsson, P., Paavilainen, L., Larsson, K., Odling, J., Sundberg, M., Andersson, A. C., Kampf, C., Persson, A., Al-Khalili Szigyarto, C., Ottosson, J. et al. (2005) Towards a human proteome atlas: high-throughput generation of mono-specific antibodies for tissue profiling. Proteomics 5, 4327–4337.

25 Taussig, M. J., Stoevesandt, O., Borrebaeck, C. A., Bradbury, A. R., Cahill, D., Cambillau, C., de Daruvar, A., Dubel, S., Eichler, J., Frank, R. et al. (2007) ProteomeBinders: planning a European resource of affinity reagents for analysis of the human proteome. Nat. Methods. 4, 13–17.

26 Davis, M. A. and Hanash, S. (2006) High-throughput genomic technology in research and clinical management of breast cancer. Plasma-based proteomics in early detection and therapy. Breast Cancer Res. 8, 217.

27 Lau, T. Y., O'Connor, D. P., Brennan, D. J., Duffy, M. J., Pennington, S. R. and Gallagher, W. M. (2007) Breast cancer

proteomics: clinical perspectives. Expert Opin. Biol. Ther. 7, 209–219.

28 Claudino, W. M., Quattrone, A., Biganzoli, L., Pestrin, M., Bertini, I. and Di Leo, A. (2007) Metabolomics: available results, current research projects in breast cancer, and future applications. J. Clin. Oncol. 25, 2840–2846.

29 Shoemaker, R. H. (2006) The NCI60 human tumour cell line anticancer drug screen. Nat. Rev. Cancer. 6, 813–823.

30 Weinstein, J. N. and Pommier, Y. (2003) Transcriptomic analysis of the NCI-60 cancer cell lines. C. R. Biol. 326, 909–920.

31 Roschke, A. V., Tonon, G., Gehlhaus, K. S., McTyre, N., Bussey, K. J., Lababidi, S., Scudiero, D. A., Weinstein, J. N. and Kirsch, I. R. (2003) Karyotypic complexity of the NCI-60 drug-screening panel. Cancer Res. 63, 8634–8647.

32 Reinhold, W. C., Reimers, M. A., Maunakea, A. K., Kim, S., Lababidi, S., Scherf, U., Shankavaram, U. T., Ziegler, M. S., Stewart, C., Kouros-Mehr, H. et al. (2007) Detailed DNA methylation profiles of the E-cadherin promoter in the NCI-60 cancer cells. Mol. Cancer Ther. 6, 391–403.

33 Nishizuka, S., Charboneau, L., Young, L., Major, S., Reinhold, W. C., Waltham, M., Kouros-Mehr, H., Bussey, K. J., Lee, J. K., Espina, V. et al. (2003) Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. Proc. Natl. Acad. Sci. USA. 100, 14229–14234.

34 Ross, D. T., Scherf, U., Eisen, M. B., Perou, C. M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S. S., Van de Rijn, M., Waltham, M. et al. (2000) Systematic variation in gene expression patterns in human cancer cell lines. Nat. Genet. 24, 227–35.

35 Neve, R. M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F. L., Fevr, T., Clark, L., Bayani, N., Coppe, J. P., Tong, F. et al. (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell 10, 515–527.

36 Fournier, M. V., Martin, K. J., Kenny, P. A., Xhaja, K., Bosch, I., Yaswen, P. and Bissell, M. J. (2006) Gene expression signature in organized and growth-arrested mammary acini predicts good outcome in breast cancer. Cancer Res. 66, 7095–7102.

37 Hu, M., Yao, J., Cai, L., Bachman, K. E., van den Brule, F., Velculescu, V. and Polyak, K. (2005) Distinct epigenetic changes in the stromal cells of breast cancers. Nat. Genet. 37, 899–905.

38 Chang, H. Y., Sneddon, J. B., Alizadeh, A. A., Sood, R., West, R. B., Montgomery, K., Chi, J. T., van de Rijn, M., Botstein, D. and Brown, P. O. (2004) Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. PLoS Biol. 2, E7.

39 Yang, J., Mani, S. A., Donaher, J. L., Ramaswamy, S., Itzykson, R. A., Come, C., Savagner, P., Gitelman, I., Richardson, A. and Weinberg, R. A. (2004) Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. Cell 117, 927–939.

40 McGee, S. F., Lanigan, F., Gilligan, E. and Groner, B. (2006) Mammary gland biology and breast cancer. Conference on Common Molecular Mechanisms of Mammary Gland Development and Breast Cancer Progression. EMBO Rep. 7, 1084–1088.

41 Howlin, J., McBryan, J., Napoletano, S., Lambe, T., McArdle, E., Shioda, T. and Martin, F. (2006) CITED1 homozygous null mice display aberrant pubertal mammary ductal morphogenesis. Oncogene 25, 1532–1542.

42 Shan, L., Yu, M. and Snyderwine, E. G. (2005) Global gene expression profiling of chemically induced rat mammary gland carcinomas and adenomas. Toxicol. Pathol. 33, 768–775.

43 Ma, X. J., Salunga, R., Tuggle, J. T., Gaudet, J., Enright, E., McQuary, P., Payette, T., Pistone, M., Stecker, K., Zhang, B. M. et al. (2003) Gene expression profiles of human breast cancer progression. Proc. Natl. Acad. Sci. USA 100, 5974–5979.

44 Kristensen, V. N., Edvardsen, H., Tsalenko, A., Nordgard, S. H., Sorlie, T., Sharan, R., Vailaya, A., Ben-Dor, A., Lonning, P. E., Lien, S., Omholt, S. et al. (2006) Genetic variation in putative regulatory loci controlling gene expression in breast cancer. Proc. Natl. Acad. USA 103, 7735–7740.

45 Turashvili, G., Bouchal, J., Baumforth, K., Wei, W., Dziechciarkova, M., Ehrmann, J., Klein, J., Fridman, E., Skarda, J., Srovnal, J. et al. (2007) Novel markers for differentiation of lobular and ductal invasive breast carcinomas by laser microdissection and microarray analysis. BMC Cancer 7, 55.

46 Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A. et al. (2000) Molecular portraits of human breast tumours. Nature 406, 747–752.

47 Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S. et al. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. USA 98, 10869–10874.

48 Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S. et al. (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc. Natl. Acad. USA 100, 8418–8423.

49 Carey, L. A., Perou, C. M., Livasy, C. A., Dressler, L. G., Cowan, D., Conway, K., Karaca, G., Troester, M. A., Tse, C. K., Edmiston, S. et al. (2006) Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA 295, 2492–2502.

50 Sorlie, T., Wang, Y., Xiao, C., Johnsen, H., Naume, B., Samaha, R. R. and Borresen-Dale, A. L. (2006) Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms. BMC Genomics 7, 127.

51 Sotiriou, C., Neo, S. Y., McShane, L. M., Korn, E. L., Long, P. M., Jazaeri, A., Martiat, P., Fox, S. B., Harris, A. L. and Liu, E. T. (2003) Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc. Natl. Acad. USA 100, 10393–10398.

52 Rouzier, R., Perou, C. M., Symmans, W. F., Ibrahim, N., Cristofanilli, M., Anderson, K., Hess, K. R., Stec, J., Ayers, M., Wagner, P. et al. (2005) Breast cancer molecular subtypes respond differently to preoperative chemotherapy. Clin. Cancer Res. 11, 5678–5685.

53 Kapp, A. V., Jeffrey, S. S., Langerod, A., Borresen-Dale, A. L., Han, W., Noh, D. Y., Bukholm, I. R., Nicolau, M., Brown, P. O. and Tibshirani, R. (2006) Discovery and validation of breast cancer subtypes. BMC Genomics 7, 231.

54 Bergamaschi, A., Kim, Y. H., Wang, P., Sorlie, T., Hernandez-Boussard, T., Lonning, P. E., Tibshirani, R., Borresen-Dale, A. L. and Pollack, J. R. (2006) Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. Genes Chromosomes Cancer 45, 1033–1040.

55 Perreard, L., Fan, C., Quackenbush, J. F., Mullins, M., Gauthier, N. P., Nelson, E., Mone, M., Hansen, H., Buys, S. S., Rasmussen, K. et al.(2006) Classification and risk stratification of invasive breast carcinomas using a real-time quantitative RT-PCR assay. Breast Cancer Res. 8, R23.

56 Hennessy, B. T., Gonzalez-Angulo, A. M. and Hortobagyi, G. N. (2005) Individualization of neoadjuvant therapy for breast cancer according to molecular tumor characteristics. Nat. Clin. Pract. Oncol. 2, 598–599.

57 Sotiriou, C., Wirapati, P., Loi, S., Harris, A., Fox, S., Smeds, J., Nordgren, H., Farmer, P., Praz, V., Haibe-Kains, B. 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.

58 Hannemann, J., Velds, A., Halfwerk, J. B., Kreike, B., Peterse, J. L. and van de Vijver, M. J. (2006) Classification of ductal

carcinoma in situ by gene expression profiling. Breast Cancer Res. 8, R61.

59 van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T. et al. (2002) Gene expression profiling predicts clinical outcome of breast cancer. Nature 415, 530–536.

60 Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F. L., Walker, M. G., Watson, D., Park, T. et al. (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N. Engl. J. Med. 351, 2817–2826.

61 Paik, S., Tang, G., Shak, S., Kim, C., Baker, J., Kim, W., Cronin, M., Baehner, F. L., Watson, D., Bryant, J. et al. (2006) Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. J. Clin. Oncol. 24, 3726–3734.

62 van de Vijver, M. J., He, Y. D., van 't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J. et al. (2002) A gene-expression signature as a predictor of survival in breast cancer. N. Engl. J. Med. 347, 1999–2009.

63 Buyse, M., Loi, S., van 't Veer, L., Viale, G., Delorenzi, M., Glas, A. M., d'Assignies, M. S., Bergh, J., Lidereau, R., Ellis, P. 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.

64 O'Brien, S. L., Fagan, A., Fox, E. J., Millikan, R. C., Culhane, A. C., Brennan, D. J., McCann, A. H., Hegarty, S., Moyna, S., Duffy, M. J. et al. (2007) CENP-F expression is associated with poor prognosis and chromosomal instability in patients with primary breast cancer. Int. J. Cancer 120, 1434–1443.

65 Simon, R., Radmacher, M. D. and Dobbin, K. (2002) Design of studies using DNA microarrays. Genet. Epidemiol. 23, 21–36.

66 Ein-Dor, L., Kela, I., Getz, G., Givol, D. and Domany, E. (2005) Outcome signature genes in breast cancer: is there a unique set? Bioinformatics 21, 171–178.

67 Desmedt, C., Piette, F., Loi, S., Wang, Y., Lallemand, F., Haibe-Kains, B., Viale, G., Delorenzi, M., Zhang, Y., d'Assignies, M. S. 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.

68 Wang, Y., Klijn, J. G., Zhang, Y., Sieuwerts, A. M., Look, M. P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M. E., Yu, J. et al. (2005) Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 365, 671–679.

69 Habel, L. A., Shak, S., Jacobs, M. K., Capra, A., Alexander, C., Pho, M., Baker, J., Walker, M., Watson, D., Hackett, J. et al. (2006) A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. Breast Cancer Res. 8, R25.

70 Esteva, F. J., Sahin, A. A., Cristofanilli, M., Coombes, K., Lee, S. J., Baker, J., Cronin, M., Walker, M., Watson, D., Shak, S. et al. (2005) Prognostic role of a multigene reverse transcriptase-PCR assay in patients with node-negative breast cancer not receiving adjuvant systemic therapy. Clin. Cancer Res. 11, 3315–3319.

71 Kreike, B., Halfwerk, H., Kristel, P., Glas, A., Peterse, H., Bartelink, H. and van de Vijver, M. J. (2006) Gene expression profiles of primary breast carcinomas from patients at high risk for local recurrence after breast-conserving therapy. Clin. Cancer Res. 12, 5705–5712.

72 Nuyten, D. S., Kreike, B., Hart, A. A., Chi, J. T., Sneddon, J. B., Wessels, L. F., Peterse, H. J., Bartelink, H., Brown, P. O., Chang, H. Y. et al.. (2006) Predicting a local recurrence after breast-conserving therapy by gene expression profiling. Breast Cancer Res. 8, R62.

73 Minn, A. J., Gupta, G. P., Siegel, P. M., Bos, P. D., Shu, W., Giri, D. D., Viale, A., Olshen, A. B., Gerald, W. L. and Massague, J. (2005) Genes that mediate breast cancer metastasis to lung. Nature 436, 518–524.

74 Minn, A. J., Gupta, G. P., Padua, D., Bos, P., Nguyen, D. X., Nuyten, D., Kreike, B., Zhang, Y., Wang, Y., Ishwaran, H. et al. (2007) Lung metastasis genes couple breast tumor size and metastatic spread. Proc. Natl. Acad. USA 104, 6740–6745.

75 Chang, H. Y., Nuyten, D. S., Sneddon, J. B., Hastie, T., Tibshirani, R., Sorlie, T., Dai, H., He, Y. D., van 't Veer, L. J., Bartelink, H. et al. (2005) Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. Proc. Natl. Acad. USA 102, 3738–3743.

76 Mazumder, A. and Wang, Y. (2006) Gene-expression signatures in oncology diagnostics. Pharmacogenomics 7, 1167–1173.

77 Lonning, P. E., Sorlie, T. and Borresen-Dale, A. L. (2005) Genomics in breast cancer-therapeutic implications. Nat. Clin. Pract. Oncol. 2, 26–33.

78 Brennan, D. J., O'Brien, S. L., Fagan, A., Culhane, A. C., Higgins, D. G., Duffy, M. J. and Gallagher, W. M. (2005) Application of DNA microarray technology in determining breast cancer prognosis and therapeutic response. Expert Opin. Biol. Ther. 5, 1069–1083.

79 Eden, P., Ritz, C., Rose, C., Ferno, M. and Peterson, C. (2004) "Good Old" clinical markers have similar power in breast cancer prognosis as microarray gene expression profilers. Eur. J. Cancer 40, 1837–1841.

80 Nimeus-Malmstrom, E., Ritz, C., Eden, P., Johnsson, A., Ohlsson, M., Strand, C., Ostberg, G., Ferno, M. and Peterson, C. (2006) Gene expression profilers and conventional clinical markers to predict distant recurrences for premenopausal breast cancer patients after adjuvant chemotherapy. Eur. J. Cancer 42, 2729–2737.

81 Bogaerts, J., Cardoso, F., Buyse, M., Braga, S., Loi, S., Harrison, J. A., Bines, J., Mook, S., Decker, N., Ravdin, P. et al. (2006) Gene signature evaluation as a prognostic tool: challenges in the design of the MINDACT trial. Nat. Clin. Pract. Oncol. 3, 540–551.

82 Siziopikou, K. P. and Cobleigh, M. (2007) The basal subtype of breast carcinomas may represent the group of breast tumors that could benefit from EGFR-targeted therapies. Breast 16, 104–107.

83 Afshari, C. A., Nuwaysir, E. F. and Barrett, J. C. (1999) Application of complementary DNA microarray technology to carcinogen identification, toxicology, and drug safety evaluation. Cancer Res. 59, 4759–4760.

84 Nuwaysir, E. F., Bittner, M., Trent, J., Barrett, J. C. and Afshari, C. A. (1999) Microarrays and toxicology: the advent of toxicogenomics. Mol. Carcinog. 24, 153–159.

85 Cooper, C. S. (2001) Applications of microarray technology in breast cancer research. Breast Cancer Res. 3, 158–175.

86 O'Driscoll, L. and Clynes, M. (2006) Biomarkers and multiple drug resistance in breast cancer. Curr. Cancer Drug Targets 6, 365–384.

87 Fan, M., Yan, P. S., Hartman-Frey, C., Chen, L., Paik, H., Oyer, S. L., Salisbury, J. D., Cheng, A. S., Li, L., Abbosh, P. H. et al. (2006) Diverse gene expression and DNA methylation profiles correlate with differential adaptation of breast cancer cells to the antiestrogens tamoxifen and fulvestrant. Cancer Res. 66, 11954–11966.

88 Amundson, S. A., Bittner, M., Chen, Y., Trent, J., Meltzer, P. and Fornace, A. J., Jr. (1999) Fluorescent cDNA microarray hybridization reveals complexity and heterogeneity of cellular genotoxic stress responses. Oncogene 18, 3666–3672.

89 Park, W. Y., Hwang, C. I., Im, C. N., Kang, M. J., Woo, J. H., Kim, J. H., Kim, Y. S., Kim, J. H., Kim, H., Kim, K. A. et al. (2002) Identification of radiation-specific responses from gene expression profile. Oncogene. 21, 8521–8528.

90 Li, Z., Xia, L., Lee, L. M., Khaletskiy, A., Wang, J., Wong, J. Y. and Li, J. J. (2001) Effector genes altered in MCF-7 human breast cancer cells after exposure to fractionated ionizing radiation. Radiat. Res. 155, 543–553.

91 Guo, G., Yan-Sanders, Y., Lyn-Cook, B. D., Wang, T., Tamae, D., Ogi, J., Khaletskiy, A., Li, Z., Weydert, C., Longmate, J. A. et al. (2003) Manganese superoxide dismutase-mediated gene expression in radiation-induced adaptive responses. Mol. Cell Biol. 23, 2362–2378.

92 Fan, C., Oh, D. S., Wessels, L., Weigelt, B., Nuyten, D. S., Nobel, A. B., van't Veer, L. J. and Perou, C. M. (2006) Concordance among gene-expression-based predictors for breast cancer. N. Engl. J. Med. 355, 560–569.

93 Oh, D. S., Troester, M. A., Usary, J., Hu, Z., He, X., Fan, C., Wu, J., Carey, L. A. and Perou, C. M. (2006) Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. J. Clin. Oncol. 24, 1656–1664.

94 Bacac, M., Provero, P., Mayran, N., Stehle, J. C., Fusco, C. and Stamenkovic, I. (2006) A mouse stromal response to tumor invasion predicts prostate and breast cancer patient survival. PLoS ONE 1, e32.

95 Park, E. S., Lee, J. S., Woo, H. G., Zhan, F., Shih, J. H., Shaughnessy, J. D. and Frederic Mushinski, J. (2007) Heterologous tissue culture expression signature predicts human breast cancer prognosis. PLoS ONE 2, e145.

96 Viksna, J., Celms, E., Opmanis, M., Podnieks, K., Rucevskis, P., Zarins, A., Barrett, A., Neogi, S. G., Krestyaninova, M., McCarthy, M. I. et al. (2007) PASSIM–an open source software system for managing information in biomedical studies. BMC Bioinformatics 8, 52.

97 Brazma, A. and Culhane, A. (2005) Algorithms for gene expression analysis. In: Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics, pp. 3148–3159, Jorde, L., Little, P., Dunn, M. and Subramaniam, S. (eds.) John Wiley & Sons., London.

98 Steinhoff, C. and Vingron, M. (2006) Normalization and quantification of differential expression in gene expression microarrays. Brief Bioinform. 7, 166–77.

99 Jeffery, I. B., Higgins, D. G. and Culhane, A. C. (2006) Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. BMC Bioinformatics 7, 359.

100 Hu, Z., Fan, C., Oh, D. S., Marron, J. S., He, X., Qaqish, B. F., Livasy, C., Carey, L. A., Reynolds, E., Dressler, L. et al. (2006) The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 7, 96.

101 Calza, S., Hall, P., Auer, G., Bjohle, J., Klaar, S., Kronenwett, U., Liu, E. T., Miller, L., Ploner, A., Smeds, J. et al. (2006) Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients. Breast Cancer Res. 8, R34.

102 Barrett, T., Troup, D. B., Wilhite, S. E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I. F., Soboleva, A., Tomashevsky, M. and Edgar, R. (2007) NCBI GEO: mining tens of millions of expression profiles – database and tools update. Nucleic Acids Res. 35, D760–765.

103 Ikeo, K., Ishi-i, J., Tamura, T., Gojobori, T. and Tateno, Y. (2003) CIBEX: center for information biology gene expression database. C. R. Biol. 326, 1079–1082.

104 Parkinson, H., Kapushesky, M., Shojatalab, M., Abeyguna-wardena, N., Coulson, R., Farne, A., Holloway, E., Kolesnykov, N., Lilja, P., Lukk, M. et al. (2007) ArrayExpress–a public database of microarray experiments and gene expression profiles. Nucleic Acids Res. 35, D747–750.

105 Hanauer, D. A., Rhodes, D. R., Sinha-Kumar, C. and Chinnaiyan, A. M. (2007) Bioinformatics approaches in the study of cancer. Curr. Mol. Med. 7, 133–141.

106 Culhane, A. C., Perriere, G. and Higgins, D. G. (2003) Cross-platform comparison and visualisation of gene expression data using co-inertia analysis. BMC Bioinformatics 4, 59.

107 Larkin, J. E., Frank, B. C., Gavras, H., Sultana, R. and Quackenbush, J. (2005) Independence and reproducibility across microarray platforms. Nat. Methods. 2, 337–344.

108 Lu, J., Lee, J. C., Salit, M. L. and Cam, M. C. (2007) Transcript-based redefinition of grouped oligonucleotide probe sets using AceView: high-resolution annotation for microarrays. BMC Bioinformatics 8, 108.

109 Carter, S. L., Eklund, A. C., Mecham, B. H., Kohane, I. S. and Szallasi, Z. (2005) Redefinition of Affymetrix probe sets by sequence overlap with cDNA microarray probes reduces cross-platform inconsistencies in cancer-associated gene expression measurements. BMC Bioinformatics 6, 107.

110 Hong, F., Breitling, R., McEntee, C. W., Wittner, B. S., Nemhauser, J. L. and Chory, J. (2006) RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. Bioinformatics 22, 2825–2827.

111 Wang, J., Coombes, K. R., Highsmith, W. E., Keating, M. J. and Abruzzo, L. V. (2004) Differences in gene expression between B-cell chronic lymphocytic leukemia and normal B cells: a meta-analysis of three microarray studies. Bioinformatics 20, 3166–3178.

112 Shen, R., Ghosh, D. and Chinnaiyan, A. M. (2004) Prognostic meta-signature of breast cancer developed by two-stage mixture modeling of microarray data. BMC Genomics 5, 94.

113 Grigoryev, D. N., Ma, S. F., Irizarry, R. A., Ye, S. Q., Quackenbush, J. and Garcia, J. G. (2004) Orthologous gene-expression profiling in multi-species models: search for candidate genes. Genome Biol. 5, R34.

114 Shankavaram, U. T., Reinhold, W. C., Nishizuka, S., Major, S., Morita, D., Chary, K. K., Reimers, M. A., Scherf, U., Kahn, A., Dolginow, D. et al. (2007) Transcript and protein expression profiles of the NCI-60 cancer cell panel: an integromic microarray study. Mol. Cancer Ther. 6, 820–832.

115 Fagan, A., Culhane, A. C. and Higgins, D. G. (2007) A multivariate analysis approach to the integration of proteomic and gene expression data. Proteomics. 7, 2162–2171.

116 Pusztai, L., Ayers, M., Stec, J., Clark, E., Hess, K., Stivers, D., Damokosh, A., Sneige, N., Buchholz, T. A., Esteva, F. J. et al. (2003) Gene expression profiles obtained from fine-needle aspirations of breast cancer reliably identify routine prognostic markers and reveal large-scale molecular differences between estrogen-negative and estrogen-positive tumors. Clin. Cancer Res. 9, 2406–2415.

117 Alexander, H., Stegner, A. L., Wagner-Mann, C., Du Bois, G. C., Alexander, S. and Sauter, E. R. (2004) Proteomic analysis to identify breast cancer biomarkers in nipple aspirate fluid. Clin. Cancer Res. 10, 7500–7510.

118 Pawlik, T. M., Hawke, D. H., Liu, Y., Krishnamurthy, S., Fritsche, H., Hunt, K. K. and Kuerer, H. M. (2006) Proteomic analysis of nipple aspirate fluid from women with early-stage breast cancer using isotope-coded affinity tags and tandem mass spectrometry reveals differential expression of vitamin D binding protein. BMC Cancer 6, 68.

To access this journal online:
http://www.birkhauser.ch/CMLS