ELSEVIER

European Journal of Cancer 40 (2004) 2592–2603

European Journal of Cancer

www.ejconline.com

Interpretation of expression-profiling results obtained from different platforms and tissue sources: examples using prostate cancer data

G. Chiorino [a,*], F. Acquadro [a], M. Mello Grand [a], S. Viscomi [a,b], R. Segir [a], M. Gasparini [c], P. Dotto [d]

[a] *Laboratory of Cancer Pharmacogenomics, Fondo "Edo Tempia", via Malta 3, Biella 13900, Italy*
[b] *Institute of Medical Statistics and Biometry, University of Milan, via Venezian 1, Milan, Italy*
[c] *Department of Mathematics, Politecnico di Torino, C.so Duca degli Abruzzi 24, 10129 Torino, Italy*
[d] *Institute of Biochemistry, University of Lausanne, Chemin de Bovaresses 155 CH-1066 Epalinges, Lausanne, Switzerland*

Received 6 February 2004; received in revised form 16 June 2004; accepted 7 July 2004
Available online 22 September 2004

Abstract

The analysis of expression signatures is a powerful tool for the classification of cancer and other tissue samples. Several protocols and platforms are available on the market, and these lead to both confirmatory and complementary results. We review the main processing techniques for cross-platform comparisons and the different tissue sources for cancer profiling. Some examples and the cross-interpretation of bibliographic data related to prostate cancer are also presented.

Keywords: Expression signatures; Processing techniques; Platforms; Data standardisation

1. Introduction

In recent years, new and very effective techniques have appeared that allow the detection and global analysis of gene expression. Several protocols for RNA extraction and processing are commercially available. These have different platforms that rely on the same basic principles, but lead to slightly different results that need correct processing and interpretation. The power of high-throughput techniques is unquestionable, and interesting and astonishing results have already been obtained. However, procedures are not yet standardised, and reproducing data within different laboratories requires rigorous guidelines for both the experimental and analytical processes. These should allow the achievement of platform-specific standards and enable data to be compared across different studies. For recent examples of gene-expression analyses, see Refs. [1–4].

In this paper, we describe the more diffuse platforms that are available for global gene-expression analyses. Some data-processing techniques with a focus on normalisation that can make the data comparable between arrays and platforms are also detailed. Whatever the technology chosen for global gene-expression analyses, replication is both important and necessary. Therefore, we focus on assessing replication reproducibility, within and between arrays. The next section is dedicated to cross-platform comparisons between single- and double-channel data. Then, we review the main sources of tissue, protocols and reference samples that can be used for the global gene-expression profiling of prostate cancer, and we explain the concepts by some examples. The

* Corresponding author. Tel.: +39 015 351830; fax: +39 015 21116.
E-mail address: giovanna.chiorino@fondoedotempia.it (G. Chiorino).

0959-8049/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ejca.2004.07.029

first refers to a cross-platform comparison of gene-expression profiles that we obtained from two prostate cell lines (LNCaP and PC-3) using Affymetrix chips and Agilent oligo-arrays. In the second example, we focus on the cross-interpretation of gene-expression data using online databases and data-mining tools.

We believe that cross-platform experiments, in particular those performed in different laboratories with standardised protocols, are extremely powerful in controlling for the main sources of variability and obtaining more reliable results.

2. Description of the different platforms for global gene-expression analysis

To be able to choose the tool that is the most suitable for a gene-expression study from the different technologies, it is necessary to evaluate critically several platforms that allow gene-expression analysis on a global scale.

Different array providers offer 'know-how' and continuously improved instruments for the determination of gene-expression profiles. A technology often identified by the name of the company that was the first to develop is Affymetrix. The principle of this platform is to hybridise biotin-labelled antisense cRNA amplified from sample mRNA on a probe array. The probe array consists of thousands of probe sets. Each set corresponds to a unique transcript, and contains 16 to 20 probe cells that are 25 mer oligonucleotides designed at several locations of the transcript. In order to quantify the non-specific hybridisation on each probe cell, a miss-match (MM) probe cell is coupled to the perfect-match (PM) one. The second most diffuse technology is the glass array one.

With the Affymetrix technology; one oligo-chip per sample is used, while the glass-array technology utilises one glass array for the simultaneous hybridisation of two populations of antisense cRNAs or cDNAs obtained from two samples (reference and assay) and labelled with the fluorescent dyes Cy3 and Cy5. A glass array consists of thousands of spots that contain millions of copies of a 60–90 mer oligonucleotide whose sequence is unique and specific to the gene of interest.

2.1. Single-channel and double-channel data

The fluorescent molecules hybridised on a spot or a probe set are excited by a laser, and thereby emit light whose intensity is proportional to their concentration on the spot or set. Unlike the Affymetrix technology, fluorescent molecules on glass arrays are excited by two lasers and emit lights at two wavelengths, which are both recorded. Therefore, single- and double-channel data are obtained from Affymetrix and glass-array platforms, respectively. All the signals from all the spots/probe sets are collected and registered to generate an array image.

Therefore, besides different hardware, the images also have different characteristics, and need to be analysed by suitable software to extract the expression information. The result of the image analysis is a text file in which each gene present on the array is linked to some bibliographical annotations and to one or two quantified fluorescence levels (intensity signals).

Affymetrix files also contain information about the robustness of the data, such as a detection value (P for 'present', A for 'absent' or M for 'marginal') and a statistical indicator. Equivalent information could be obtained from glass-array platforms. After applying filters (as will be described later on), the degree of robustness of data from both platforms is similar, and only 'sure' genes will be further analysed. More than one hybridisation is usually performed, normally with a dye-swap scheme for glass arrays, to increase the data significance and eliminate biases.

3. Data processing

3.1. Quality filtering and systematic effects

The raw data from glass arrays or oligo chips need to be filtered in order to remove low-quality signals.

Quality filtering is usually implemented during the image-analysis step, using criteria such as intensity level and homogeneity or, for glass arrays only, spot-size parameters (diameter, circularity and perimeter) [5,6]. If intensity signals are below zero (after local background subtraction) for the reference and assay samples, the spot/probe set is removed from analysis since it contains no information and may affect the normalisation. All the signals below or very close to zero (marginal) in only one sample are generally replaced by one arbitrary minimum intensity value. This procedure dramatically improves data homogeneity.

Systematic effects are effects intrinsic to a technology that may perturb the acquisition and/or the analysis of data. For instance, dye-incorporation efficiency depends on the dye steric dimensions and on the transcript sequences; hybridisation efficiency is affected by the amount of oligo probes immobilised on an array during its fabrication, the amount of labelled molecules used per array, the local temperature and humidity in the hybridisation chamber, and/or by the preferential hybridisation of Cy3- or Cy5-labelled molecules on some spots; and image acquisition may be biased by the local curvature of the array surface and/or by the detection efficiencies between the fluorescent dyes.

To assess the dye effect on the labelling efficiency (for a given sequence, the incorporation of the fluorescent dye varies according to the type of dye: Cy3 or Cy5)

and the gene effect on the labelling efficiency (for a given fluorescent dye, the fluorescent dye incorporation varies according to the gene sequences) in the glass-array technology, self-to-self and/or dye-swap experiments could be performed. In the self-to-self experiment, the same cRNAs are labelled with both fluorescent dyes and hybridised on one glass array. When there are no dye or gene effects on labelling efficiency, data are distributed as shown in Fig. 1(a). In dye-swap experiments, cRNAs from reference and assay samples are first labelled with both fluorescent dyes. Then, the reference Cy3-labelled and the assay Cy5-labelled cRNAs are hybridised on one glass array, while the reference Cy5-labelled and the assay Cy3-labelled cRNAs are hybridised on a second array. In the absence of a dye effect on labelling efficiency, the distribution of data should be the same in the two experiments (Figs. 2(a) and (b)). All the genes whose labelling efficiency is biased by the dye or gene effect are filtered out.

3.2. Normalisation

In addition to systematic effects, experimental effects may interfere with the data analysis. For instance, unequal amounts of labelled cRNAs/cDNAs, or laser voltage settings [7], may modify intensity measurements. To eliminate these biases and to compare data obtained from replicate experiments or from two different platforms, data need to be normalised.

Several normalisation methods using different approaches have been proposed. They are based on simple assumptions, such as an equal amount of starting material (i.e., mRNA) from all the samples; the array contains a gene collection representative of the human genome. As pointed out by Yang [8], the first step is to decide which genes will be used for normalisation. If the numbers of up- and down-regulated genes are similar and represent a small proportion of the entire gene collection of the array, the normalisation factor is calculated by using that entire gene collection. Otherwise, it could be calculated by using a selected subset of the arrayed genes, such as housekeeping genes or a subset of non-differentially expressed genes (rank-invariant approach) [5,9,10]. Two main normalisation approaches have been proposed: global normalisation [8] and intensity-dependent normalisation, with linear [8,11] or non-linear [8,12] dependence.

In the global approach, normalisation factors are calculated assuming that the total signal intensity summed over all the selected elements in the arrays should be the same for each sample, or that the mean or median intensities should be the same within each sample on a single array or across all arrays. In a logarithmic scale, this means that the log-ratio distribution centre is shifted towards zero, but its spread is not affected by normalisation.

Intensity-dependent normalisation is based on the fact that the M values ($M = \log 2$ ratio of assay over reference signal intensity) systematically depend on the signal intensities. This fact most commonly appears as a deviation from zero for low-intensity spots, as evidenced by several reports [5,8,13] and shown in Fig. 3(a). Therefore, these approaches aim to return to the condition of the ideal experiment, when it is expected that $M = 0$ for non-differentially expressed genes and that data distribution is as in Fig. 3(b).

In general, these methods calculate regression functions written as $M = \alpha_0$ for global normalisation and $M = c(A)$ for intensity dependent normalisation (A = mean of $\log 2$ signal intensity).

$c(A)$ can be estimated by using the locally linear regression function Loess [8,11,14] (linear approach),

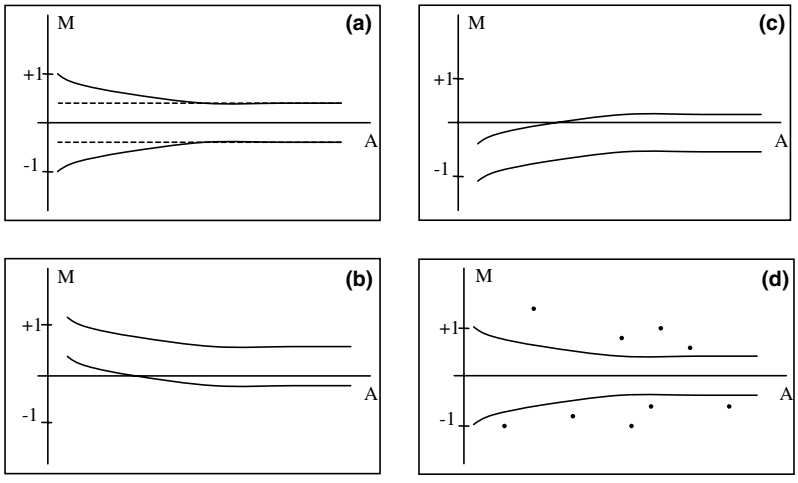

Fig. 1. Plots of $\log 2$ ratios (M) *vs.* mean $\log 2$ intensities (A) in self-to-self experiments. (a) The data distribution should ideally be symmetrical with respect to the x-axis and within the two dashed lines of the M–A plot; however, a wider scatter is often seen at low intensity values. (b) and (c) Dye effects can cause the data distribution to shift and no longer be symmetrical. (d) Gene effects can cause some M values to fall outside the lines.

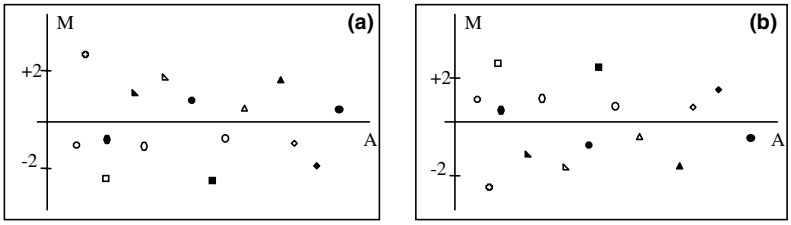

Fig. 2. Schematic *M–A* plots of a dye-swap experiment: when labelling inversion works, data plots are symmetrical.

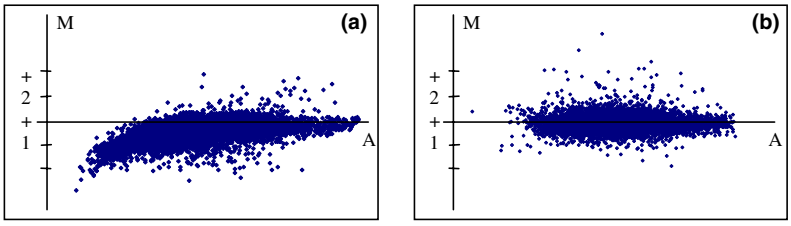

Fig. 3. Examples of *M–A* plots related to assayreference gene-expression comparisons. (a) data distribution is affected by a Cy3 dye effect at low intensity values that can be corrected by applying an intensity-dependent normalisation algorithm; (b) data need no transformation.

or by the locally-weighted linear robust smoother Lowess [8,12,14] (non-linear approach). The underlying idea is to approximate an unknown function as a polynomial regression with gradually varying coefficients, generalising the intuitive idea of smoothing by using a moving average. This method prevents deviant points from distorting the regression. Lowess normalisation differs from Loess because it de-emphasises the contributions made by data from array elements that are far (on the *M–A* plot) from the interpolated mean curve.

All these normalisation methods can also be applied separately to each grid on the microarray to keep in account spatial effects (within print-tip group normalisation) [15]. This approach assumes that differentially expressed genes are scattered all across the array and not concentrated within specific areas. It corrects for systematic spatial variation in the array related to spotting pens, variability in the array surface, and slight local differences in hybridisation conditions, and thereby shifts the centres of the subarray log-ratio distributions to zero.

In general, intensity-dependent normalisation reduces the spread of the log ratios compared with global normalisation. Combined with a within print-tip group scale normalisation, the spread is further reduced, but it can still be non-homogeneously distributed. In fact, whereas the normalisation approach adjusts the mean of the log-ratio measurements, local random variations can cause the variance of the measured log ratios to differ from one region of an array to another or from one intensity range to another. Therefore, Yang and colleagues [15] have proposed scaling methods to reduce variability within and between arrays. Basically the log ratios are divided by scaling factors in order to obtain the same variance of log ratios all over the array [8,16]. This approach assumes that the within print-tip group or intensity-dependent log ratios follow a normal distribution with a mean equal to zero and a variance equal to the product of the variance of the true log ratios multiplied by the within-tip group or the intensity-dependent scaling factor. This method may also be used for comparing data from replicate experiments or across platforms, since the spread of the log ratios may differ from one array to another.

3.3. Recent normalisation approaches

Starting from these first-developed algorithms (global, intensity-dependent, variance regularisation), more complex approaches were developed such as the window-iterative normalisation [17], the robust non-linear method [18], or an iterative application of the non-linear function Loess combined with a smoothing function for background correction [19]. Statistical models have also been proposed to normalise data [20–22]. They could be powerful tools only if the assumptions of the chosen model are verified. In their recent review, Park and colleagues [23] compared several normalisation methods. They showed that the intensity-dependent normalisation performs better than simpler global normalisation in many cases, but the non-linear intensity-dependent procedures and the linear intensity-dependent methods share similar performances. Park and colleagues conclude that complex methods do not necessarily perform better than simpler ones when they are used for microarray studies. In fact, they may add bias if the assumptions are not satisfied and, more in general, the fact that a non-linear approach adjusts better an *M–A* plot is not really evidence that the method is not just fitting 'noise'.

However, several groups have developed new methods that improve data normalisation. For more details on glass-microarray normalisation, Smyth and Speed [24] have recently described a variation on Loess normalisation that consists of a global Loess normalisation, and a two-dimensional normalisation and composite normalisation methods. Among the other novelties on normalisation, Faller and colleagues [25] propose an iterative non-parametric non-linear normalisation scheme called simultaneous Alternating Conditional Expectation (sACE), which reduces the number of false-positive calls (by 57% in 12 experiments with six repeats). Wilson and colleagues [26] present non-linear normalisation techniques dealing also with smooth spatial trends in intensity. They test this method on a new type of glass-microarray experiment that is increasingly used, the small-scale speciality or 'boutique array', where a large proportion of the genes on the microarray are expected to be highly differentially expressed.

4. Replication

Several studies have emphasised the need for technical and biological replicates to increase the reliability of microarray results [27,28]. Indeed the design of experiment (DOE) is one of the first and most relevant steps within the planning of genome analyses [29–32]: concepts of randomisation and replication [33,34] should guide the array layout (the spatial arrangement of the sequences) and the sample allocation.

4.1. Replicated spots

Multiple spotting of an oligonucleotide on an array provides a means of assessing the data quality for the sequence of that oligonucleotide. In an analogous way, multiple nucleotide sequences referring to the same gene (as in Affymetrix probe sets) provide an internal check for the specificity of the hybridisation binding. On glass-slide platforms, the basic idea is to detect and filter out poor-quality genes using measurements from multiple spots [5,6,8]. In general, if each gene is spotted more than once on the slide, the coefficient of variation (CV) (i.e., standard deviation (SD) divided by the mean) of the set of Cy5–Cy3 ratios may be used as an index of quality for each spot, with spot quality being inversely related to its CV. An interesting windowing procedure [5] suggests to drop out unreliable spots within replicates and keep reliable ones for subsequent analysis, therefore avoiding any loss of information.

4.2. Replicated arrays

Lee and colleagues [28] point out that many genes are misclassified in replicated experiments and that false-positive results usually dominate. They also show how the classification precision varies with the number of replicates. A single replicate may appear to have a low misclassification percentage relative to other replicates, but it is evident that the reliability cannot be anticipated in advance. So, it is very important to ensure there are enough replicated arrays within the experimental design [20,21,35].

However, microarray log ratios from replicated experiments have to be treated in a slightly different way than with the traditional statistics applied to replication. As evidenced by Lonnstedt and Speed [36], large standardised averages (t) could arise because of small standard errors (that is a desirable endpoint), even though the mean itself is small. In reality, because of the huge number of genes on each array, it is quite probable to find genes with very small standard errors, but only some of them have non-small means. Tusher and colleagues [37] therefore propose a refinement of t, by adding a constant term to the denominator of the standardised average, thus avoiding the issue that the denominator gets too small. In one report [38], the 90th percentile of the standard errors of all the genes is suggested as a constant term to add.

Bayesian approaches have also been used to develop hierarchical linear models of observed log ratios. They incorporate prior knowledge generated from calibration experiments to construct prior distributions of the model parameters [5,28] and obtain a different adjustment of the standardised average.

5. Common-sequence alignment

Each support available for microarray analysis contains specific transcript sets, both for the kind of genes and for the sequences that represent them. To compare results from two platforms, it is necessary to be able to identify the genes represented on both.

Helpful bioinformatics instruments (http://pga.tigr.org/tigr-scripts/magic/r1.pl) allow us to interrogate specific databases for the comparison of different kinds of array.

Before any comparison, it is necessary to explain the meaning of two codes that usually accompany a given transcript. The UniGene ID (es. Hs.191446) is the most general identifier for a gene, while the GenBank Accession Number (es. AB005217) is an alpha-numerical code representing a specific nucleotide sequence of a gene. More than one GenBank Accession Number could indicate the same gene.

When we consider data from two different technologies, it is very important to decide at what level the comparison has to be made. If any difference is found between the results obtained from the comparison made at the UniGene ID level and those found by comparing

unique Accession Numbers, this could be due to low-quality sequence selection in any of the platforms or it may be a reflection of differential splicing.

6. Different tissue sources, references and protocols

6.1. Tissue sources

Sample choice is the first point to define in the DOE, since each sample must be representative of one population, homogeneous and stable.

Tissue samples from patients are not always easy to collect, but they are one of the solutions mainly adopted because they represent the situation *in vivo*.

In order to avoid rapid degradation of the material, samples are usually snap-frozen in liquid nitrogen, even though this procedure does compromise the fine morphology of the tissue. Therefore, depending on the use of material (for example, laser-capture microdissection; LCM), alternative procedures (such as RNA later or zinc fixation) may be adopted that still prevent RNA degradation and preserve tissue morphology.

Studies on whether the processing time may influence the gene-expression profile for prostate tissue samples have indicated that ischaemia, which occurs during the interval between radical prostatectomy and sample collection, can alter gene expression-activating pathways involved in hypoxia [39].

Paraffin blocks are also preserved for many years in tissue archives. However, their use for global gene-expression analyses is not recommended, since RNA can be degraded by cross-linkage with formalin [40], even though new protocols and *ad hoc*-designed oligos recently set conditions that might overcome this problem, and Schoor and colleagues [41] demonstrate that moderate RNA degradation does not prevent microarray analysis of small amounts of RNA.

The heterogeneity of cellular populations within some tumours, such as prostate cancer, which is characterised by multifocality and different degrees of malignancy, is the main limitation on the use of tissue samples. It is possible to solve this problem either by establishing a required threshold that the tumour sample must satisfy to be included in a study (for instance, it should contain more than 75% of malignant cells) or by using manual or computerised LCM to isolate homogeneous subpopulations. Both solutions have pros and cons.

A good approach is to couple global gene-expression studies with a rather new technology, called *tissue array*, that joins traditional techniques (immunohistochemistry and *in situ* hybridisation) to the simultaneous analysis of hundreds of samples that represent different organs, tumoral and normal areas of same tissue or various stages of the same pathology [42]. Small cylinders from archival, paraffin-embedded tissues are included in a new paraffin block according to a precise plan, cut in sections of variable thickness and deposited on slides. The main limit is the small dimension of the sections, which may not portray the whole tissue [43,44].

On the other hand, cell lines or primary cultures represent easily available sample sources for *in vitro* studies. In prostate carcinoma, the more used cancer cell lines are: LNCaP (an androgen-sensitive cell line established from the left supraclavicular lymph node metastasis of a 50-year-old man with prostate carcinoma in 1977) [45], PC3 (an androgen-insensitive cell line established from the bone marrow metastasis isolated post-mortem from a 62-year-old Caucasian man with grade IV poorly differentiated prostate adenocarcinoma after androgen suppression therapy) [46], and DU145 (established from the tumour tissue removed from the metastatic central nervous system lesion of a 69-year-old man with prostate carcinoma in 1975) [47]. These cancer cell lines represent progressive stages of the disease, and are reproducible and comparable models. However, it is important both to be aware that cell-culture conditions do not reflect the cell environment within a tissue and to be sure about the origin and characteristics of the cells [48].

6.2. Reference samples

The choice of a reference sample fundamentally depends on the biological question upon which the experiment is based, but also has more practical considerations such as its homogeneity, stability and abundance. Either commercial or laboratory-prepared references of various natures and origins (surgical specimens, biopsies, cell lines) can be used.

In the DOE for a tumour-profiling study, it must be established whether to compare various stages of the same pathology (for example, androgen-sensitive and -insensitive samples of prostate carcinoma) or to compare tumour samples with normal ones. In this case, it is necessary to define what is a normal sample: a sample from a healthy man (difficult to find) or from the same patient but near to the carcinoma, or a sample affected by hyperplasia or benign tumour (easier to find, but not 'normal'). Furthermore, there is another issue to consider, tissue complexity and heterogeneity: within the tumour there are cells with different levels of maturity (such as tumour stem cells or low/highly differentiated cells). Various strategies can be adopted in tumour-profiling studies and RNA samples can be compared with:

- a pool of RNA obtained from all the samples (cell lines or tissues) tested in the experiment – in this way the different expression of a sample with respect to the mean is evidenced;

- a pool of RNA obtained from normal tissue collected in the study – the main limit is that the reference does not have standard characteristics and derives from a small number of samples [49];
- a commercial organ-specific reference – a pool of RNA obtained from a specific organ of healthy donors (information about donors may not be exhaustive, but results can be easily compared between laboratories);
- a human universal reference – a pool of total RNA, extracted from a collection of different tissues representing the whole organism, or from cell lines representing different human tissues to ensure a broad coverage of human genes (the pool does not allow a specific organ analysis; in this way specific organ genes can be understimated).

6.3. Protocols

The choice of a protocol depends principally on the aim of the study, but also on duration and economic considerations.

In gene expression-profiling studies with oligo-microarrays the main steps are: RNA extraction (where the kind and the amount of sample influence the yield of reaction), RNA amplification (which must be linear to maintain proportions between different mRNAs), labelling (direct or indirect, or with dendrimer) [50,51], qualitative and quantitative RNA analysis (to verify RNA integrity, and estimate amplification, labelling and sample concentrations), and hybridisation. Slight changes between protocols or within the same one (for instance, different amplification protocols, simultaneous sample labelling or not, etc.), could lead to different results [52]. However, we recently made an interesting comparative hybridisation of two aliquots of the same sample, one amplified for 4 h and the other for 14 h, and found that no significant changes in gene expression could be detected. As shown in Fig. 4, genes with an expression ratio larger than 2 can only be seen for low intensity values, where variability is usually higher. By applying local variance regularisation, significant changes are no longer detected.

In any case, data analysis and experimental procedures are still not standardised, and therefore it is necessary to report carefully all the steps adopted. Minimal information about a microarray experiment (MIAME) guidelines [53] are very useful for comparability and reproducibility of experiments within different laboratories.

7. Cross-platform comparison of LNCaP and PC-3 data

All the techniques described above (replication, combination of different normalisation algorithms, intensity-dependent variance regularisation, common-sequence alignment, platform-correlation assessment) were applied to the comparison of gene-expression levels between two prostate cancer cell lines.

The differential expression of the two cell lines (LNCaP and PC-3) was evaluated by two separate laboratories using the same RNA samples on two different microarray supports: Agilent Oligo Human 1.0 A Glass Array (22 575 spots) and Affimetrix Human U95 Av2 Biochip (12 626 probe sets). Without considering control spots or replicated sequences, the number of unique represented UniGene Clusters is 15 444 in Agilent and 8305 in Affymetrix, while the number of different GenBank Accession numbers is 15 939 in Agilent and 11 334 in Affimetrix.

This comparison was made because LNCaP cells represent an early stage of prostate cancer, in which cells are still sensitive to androgens, while PC-3 cells represent a more aggressive tumour that does not respond

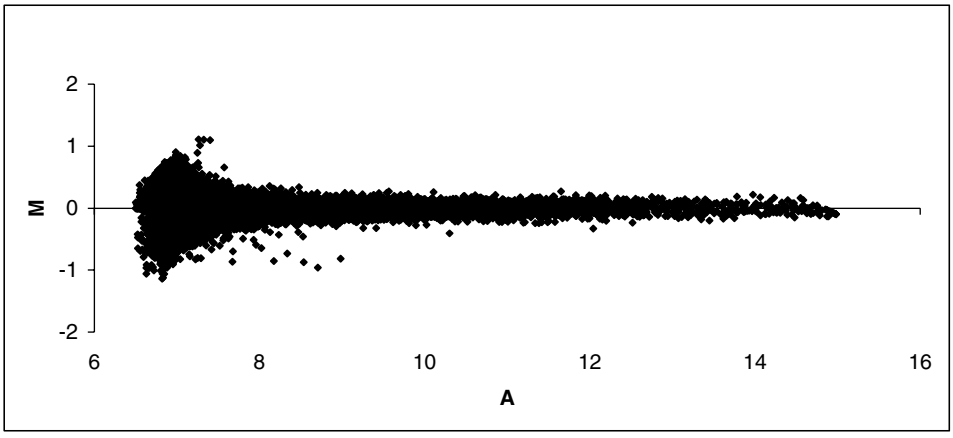

Fig. 4. *M–A* plot of a comparative hybridisation, in which one sample has been amplified for 14 h and the other for only 4 h. Only low-intensity spots, which usually are the most variable, show a slight differential expression. No data normalisation has been applied.

to hormone therapy. This study was part of a larger experiment in which the LNCaP and PC-3 cells were used as the reference, the assays being LNCaP and PC-3 modified cells in which the expression of a selected gene was amplified by viral transfection.

Therefore, while the Affymetrix data derived from separate chips for LNCaP and PC-3 cells, the Agilent data derived from result files (related to comparative hybridisations) from which only reference intensity levels were extracted.

However, it was, possible to separate the two fluorescence channels and analyse them as if they came from different chips. The fundamental idea supporting this approach is that there are no interactions between the two labelled probes when they are put on the same slide for the hybridisation. This is possible because the number of oligo sequences in each spot is higher than the number of labelled targets that could match them (non-competitive reaction).

Therefore, we collected the separated fluorescence values of the reference channels and applied a series of filters to eliminate genes that did not respect quality criteria. Global normalisation was performed to get the intensity medians to a common value, then an intensity-dependent normalisation using the Loess function and local data scaling was applied. The 'span' variables that determine the smoothness of the robust local regressions (which estimate the local mean log ratio and SD) were set to 0.6, whereas the 'trim' variables that determine the proportion of the most extreme values in the data (which are ignored in this estimation) were set to 0.1.

The transformed expression values of replicates (two for Affymetrix and two, with dye swap, for Agilent) were then averaged, and the subset of common genes between Agilent and Affymetrix was considered for the cross-platform comparison study.

The research for common genes was done using the *RESOURCERER* software (http://pga.tigr.org/tigr-scripts/magic/r1.pl): 6876 UniGene Clusters present on both platforms were found.

To estimate the correlation between Agilent and Affymetrix results, we used two different methods.

The first was the analysis of the concordance between the normalised expression values (Z-scores) obtained from the two technologies. As we see in Fig. 5, the scatter plot of Affymetrix *vs.* Agilent Z-scores shows a fair concordance between the results for the two platforms. This is a good result, considering that the experiments were done in different laboratories and that two channel intensities were taken separately.

A more thorough analysis of the concordance between the two platforms would involve assessing the significance and the estimated order of magnitude of the platform effect in a linear model accounting for all different sources of variability and verifying that the genes identified as differentially expressed by the two platforms are mostly the same. Therefore, we evaluated the number and kind of genes with significantly different expression in the two cell lines, extracted using the same extraction criteria for Agilent and Affymetrix data. Extracted genes with equal behaviour (total agreement) in the two platforms, different behaviour (total disagreement) or found to be modulated by one platform only were evaluated. The threshold Z-scores used to consider a gene as being differentially expressed in PC-3 *vs.* LNCaP were 1.96 for upregulated genes and −1.96 for downregulated. The results of this analysis are presented in Fig. 6: for only three genes is there complete disagreement, while 207 genes were found to be differentially expressed by both technologies, 369 were modulated by Agilent and borderline (same sign but below the threshold) with Affymetrix, whereas 251 were modulated by Affymetrix and borderline with Agilent.

As a control, we checked the gene coding for the androgen receptor (AR) and found, by both technologies, that it was downregulated in PC-3 cells. This result was consistent, since PC-3 cells are effectively negative for AR, while LNCaP are positive.

We can conclude that there is reasonably good agreement between the two platforms used for this comparison and that they gave rise to both confirmatory and complementary results. Previous reports [54] did not show such a concordance and one of the reasons could be that oligo-chips were compared with cDNA instead of specifically designed oligo-glass arrays.

8. Comparative bibliographic analysis

When researchers obtain a list of significantly modulated genes from an experimental dataset, they usually look for interesting information about them and their products, such as protein function (if it is known), gene-ontology terms, chromosome location, related disease information and so on. After choosing the kind of organism, search options include gene symbol/name, or gene-related codes such as clone ID, GenBank Accession Number, UniGene Cluster ID and LocusLink ID (see http://harvester.embl.de for a detailed list of the main online gene search engines).

A useful clue provided by some gene databases available online, such as Oncomine (http://oncomine.org), concerns gene-expression data associated with the input gene and obtained from published papers with publicly available datasets containing that gene. By clicking on the 'Show gene expression data' link provided by SOURCE (http://source.stanford.edu) [55], for instance, a list of available datasets appears, with dataset descriptions and paper references, together with authors' web pages. The 'View expression' option leads to a row of coloured squares representing differential expression

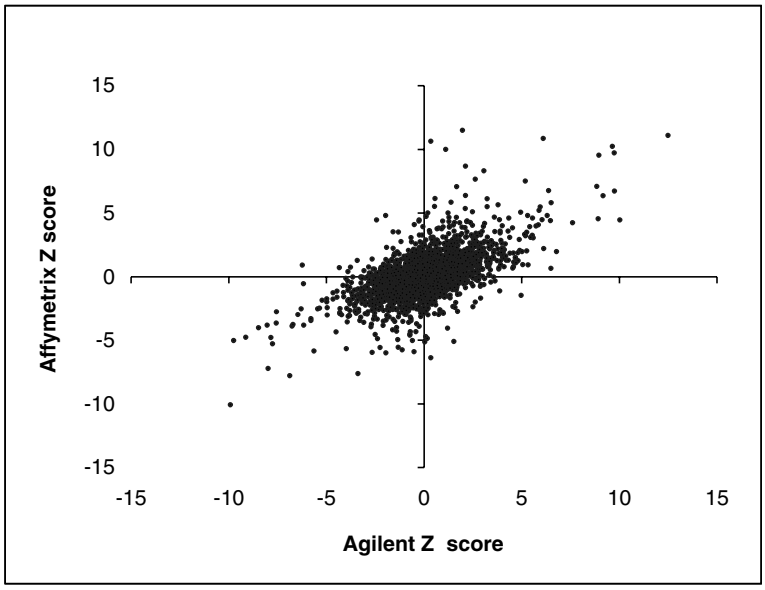

Fig. 5. Scatter plot representing the transformed log-ratio values of Agilent *vs.* Affymetrix data related to the comparison between LNCaP and PC-3 expression. The plot shows a reasonable concordance between the two platforms, as also shown by the Pearson correlation coefficient, which equals 0.6.

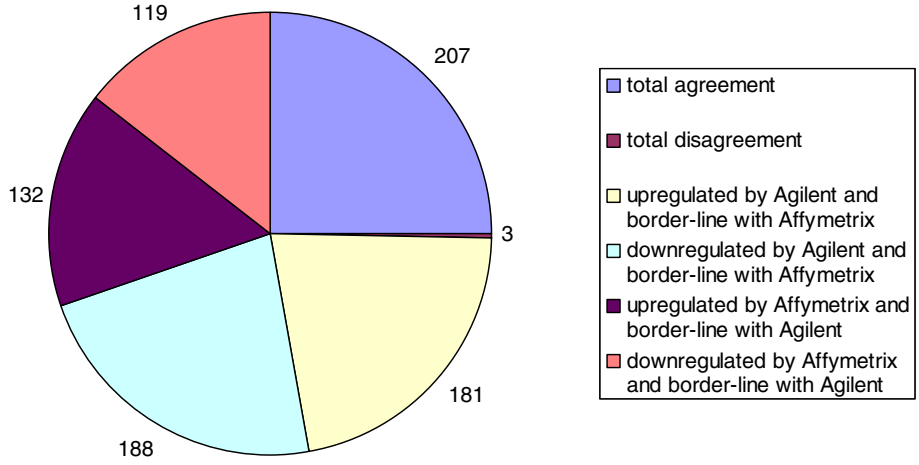

Fig. 6. Distribution of the number of genes differentially expressed in LNCaP cells *vs.* PC3, extracted from the two platforms; *Z*-score thresholds of 1.96 for upregulated genes and 1.96 for downregulated ones were set.

measures associated with the input gene. Each square refers to a sample within the experiment described in the selected paper. The colour is usually associated with the level of up- or down-regulation with respect to the specific reference sample used (as a general rule, the red scale is used for upregulation, the green for downregulation, black for lack of modulation and grey for missing values).

A practical example is useful to clarify this procedure and point out its potential and pitfalls. Let us choose again the gene coding for human AR, with UniGene number Hs. 99915. Three datasets containing that gene are available: the first refers to acute myeloid leukaemia (AML) and acute lymphoblastic leukaemia (ALL) samples leukaemia samples [56] and therefore is not informative for prostate cancer; the second concerns carcinoma classifications using 174 human epithelial tumours [57] including prostate cancer; the third is called normal tissue atlas (from mouse and human samples) and also contains normal and prostate cancer samples [58]. AR is found twice in the second dataset, but only one sequence is red-coloured in prostate tissues and, when clicked, is associated with a gene cluster. That cluster contains 48 genes with the same expression profile across the experimental samples that are upregulated in prostate tissue and down- or not modulated in the other tissues. In particular, the three genes with behaviour most similar to AR are hepsin, *GREB1* and *KIF5c*

(kinesin family member 5c). Hepsin is a transmembrane serine protease that plays an essential role in cell growth and the maintenance of cell morphology. Interestingly, if we look at the 'Unigene & expression sequence tag (EST) expression information' link (another useful tool provided by SOURCE), we note that the normalised expression distribution for hepsin is very high in invasive prostate tumours (33%). *GREB1* is an oestrogen-responsive gene thought to play an important part in hormone-responsive tissues and cancer, while kinesin is a microtubule-associated force-producing protein that may play a role in organelle transport. Its normalised expression distribution is high in melanoma, ovary, fibrotheoma and high-grade prostatic intraepithelial neoplasia (15%). No information is associated to *KIF5c*, but its very similar expression behaviour allows us to suppose that this gene may code for a protein with a role in cell growth.

If the same procedure is applied to the third dataset, only one AR sequence appears and the related gene cluster contains 46 genes that are upregulated; not only in prostate tissues, but also in testis, ovary and uterus.

Now, what is the meaning of all this? It is clear that it depends on all of the considerations previously discussed in this paper. What do the coloured squares hide? Is it a log ratio between assay and reference? Which data preprocessing and transformation procedures have been used? What about the reference? How has normal prostate tissue been obtained? And the cancer tissue?

We know that prostate cancer is multifocal and that AR expression varies greatly, depending on the cell origin and location within the prostate gland [59], on the gene mutation state and the disease stage, with the AR protein having a complex and as yet unclear dual role in prostate cell growth regulation [60]. For example, some androgen-independent prostate cancers are characterised by AR amplification accompanied by an overexpression of AR that is able to promote cell growth, even with very low concentrations of circulating androgens, while some other equally aggressive tumours have mutated AR that can be activated by non-androgens or even by anti-androgens. So, the different levels of AR expression can be associated with very varied factors that are not so easy to detect by just analysing the AR transcript level.

However, whatever reference or sample is used, the power of the global analysis of gene expression consists of the detection of 'expression profiles' that are able to cluster different clinical samples or cell lines that eventually have a common response to treatments. Now we are not just looking at one gene at a time, but at expression signatures in which several genes apparently not linked to each other are to be evaluated together.

However, the potential of the information derived from this comparative approach needs to be 'weighted' and the results are yet to be interpreted in a context-dependent way.

For example, the gene *trp-p8*, which was found to be prostate-specific and upregulated in prostate cancer compared with normal prostate tissue [61], appears to be one of the most important genes whose diminution (and eventually loss) of expression was significantly correlated with relapse and negative prognosis in a genome-wide microarray analysis study [62]. Cancer samples were subdivided into non-relapsed and relapsed and the *trp-p8* gene clustered into the group of genes that have a lower level of expression in relapsed cancer than in non-relapsed. For instance, while upregulation compared with that in normal tissue is an easily understandable datum that can be independently evaluated using samples from different patients by investigators from different laboratories, the concept of 'decrease in expression' cannot be translated into prognostic practice, since it has a relative meaning that makes sense only within the context of the study. For instance, referring intensity levels to validated 'internal standards' would improve the translatability of results.

Another important issue concerns cell-type homogeneity within the samples chosen for the analyses [63]. Therefore, to be able to achieve real clinical application, a strong effort should be made to standardise the techniques and increase the availability of clear guidelines for cell-type extraction (and percentage evaluations) and gene-expression comparisons.

9. Conclusions

We have described and commented on the main steps that occur between cancer tissue collection and hybridised oligo-chips or glass arrays, and between scanned oligo-chips or glass arrays and gene-expression clusters.

Cross-platform comparison and data validation with techniques other than microarrays are powerful tools for overcoming variability and obtaining reliable results.

Rapid translation into clinical application is desirable and will soon occur when protocols become standardised and sharable.

The potential of expression 'signatures' for sample classification could then be exploited to reduce overtreatment by improving both diagnosis and prognosis prediction and to select drugs to be used in a 'patient-tailored' way.

Many other aspects, such as gene-network modelling and gene-regulation analysis by promoter evaluation, were not covered in this Review, but are also part of the 'gene-expression world'. All these steps taken together are of extreme importance in providing a clear picture of what happens at the molecular level when gene deregulation, such as in cancer, occurs.

Conflict of interest statement

None declared.

Acknowledgements

C.G., A.F. and S.R. were supported by Compagnia di San Paolo, Torino; M.G.M. was supported by the LATT project of Regione Piemonte; M.G. was supported by Ministero dell'Istruzione, dell'Università e della Ricerca, grant PRIN 2003133820. Dr. Fabien Richard is gratefully acknowledged for his helpful discussions and review of the manuscript.

References

1. *Nat Genet* 2002 Dec, **32** Suppl.
2. Parmigiani G, Garrett ES, Irizarry RAet al.], *et al.*], eds. *The analysis of gene expression data: methods and software*. New York, Springer, 2003.
3. Speed T, ed. *Statistical analysis of gene expression microarray data*. London/Boca Raton, Chapman&Hall/CRC, 2003.
4. Amaratunga D, Cabrera J. *Exploration and analysis of dna microarray and protein array data, Wiley series in probability and statistics* 2003.
5. Tseng GC, Oh MK, Rohlin L, *et al*. Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res* 2001, **29**(12), 2549–2557.
6. Dudoit, Yang YH. Bioconductor R packages for exploratory analysis and normalization of cDNA microarray data. In Parmigiani G, Garrett ES, Irizarry RA, *et al*, eds. *The analysis of gene expression data: methods and software*. New York, Springer, 2003. pp. 73–101.
7. Speed T. *CRC-CGF Bioinformatics Course*. WEHI, 2002.
8. Yang YH, Dudoit S, Luu P, *et al*. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 2002, **30**(4), e15.
9. Schadt EE. Feature extraction and normalization algorithm for high-density oligonucleotide gene expression array data. *J Cell Biochem* 2001, **84**(S37), 120–125.
10. Tseng GC, Oh MK, Rohlin L. Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res* 2001, **29**(12), 2549–2557.
11. Colantuoni C, Henry G, Zeger S, *et al*. SNOMAD (standardization and normalization of microarray data): web-accessible gene expression data analysis. *Bioinformatics* 2002, **18**(11), 1540–1541.
12. Cleveland WS. Robust Locally weighted regression and smoothing scatterplots. *J Am Stat Assoc* 1979, **74**, 829–836.
13. Yang IV *et al*. Within the fold: assessing differential expression measures and reproducibility in microarray assays. *Genome Biol* 2002, **3**, 1–12.
14. Kepler TB, Crosby L, Morgan KT. Normalization and analysis of DNA microarray data by self-consistency and local regression. *Genome Biol* 2002, **3**(7), 1–12.
15. Yang YH, Dudoit SD, Luu P, *et al*. Normalization for cDNA microarray data. *SPIE BioE*.
16. Huber W, von Heidenbreck A, Sultmann H, *et al*. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* 2002, **18**, S96–S104.
17. Wang Y, Lu J, Lee R, *et al*. Iterative normalization of cDNA microarray data. *IEEE Trans Inf Technol Biomed* 2002, **6**, 29–37.
18. Workman C, Jensen LJ, Jarmer H. A new non-linear normalization method for reducing variability in DNA microarray experiments. *Gen Biol* 2002, **3**, 1–16.
19. Edwards D. Non-linear normalization and background correction in one channel cDNA microarray studies. *Bioinformatics* 2003, **19**, 825–833.
20. Kerr MK, Churchill GA. Statistical design and the analysis of gene expression microarrays. *Genet Res* 2001, **77**, 123–128.
21. Kerr MK *et al*. Statistical analysis of a gene expression microarray experiment with replication. *Stat Sinica* 2002, **12**(1), 203–217.
22. Wolfinger RD, Gibson G, Wolfinger ED, *et al*. Assessing gene significance from cDNA microarray expression data via mixed models. *J Comput Biol* 2001, **8**(6), 625–637.
23. Park T, Yi S-G, Kang S-H, *et al*. Evaluation of normalization methods for microarray data. *Bioinformatics* 2003, **4**(1), 33–45.
24. Smyth GK, Speed T. Normalization of cDNA microarray data. *Methods* 2003, **31**(4), 265–273.
25. Faller D, Voss HU, Timmer J, *et al*. Normalization of DNA-microarray data by non-linear correlation maximization. *J Comput Biol* 2003, **10**(5), 751–762.
26. Wilson DL, Buckley MJ, Helliwell CA, *et al*. New normalization methods for cDNA microarray data. *Bioinformatics* 2003, **19**(11), 1325–1332.
27. Nguyen DV, Arpat AB, Wang N, *et al*. DNA microarray experiments: biological and technological aspects. *Biometrics* 2002, **58**(4), 701–717.
28. Lee MT, Kuo FC, Whitmore, *et al*. Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *PNAS* 2000, **97**(18), 9834–9839.
29. Dobbin K, Simon R. Comparison of microarray design for class comparison and class discovery. *Bioinformatics* 2002, **18**(11), 1438–1445.
30. Yang YH, Speed T. Design issues for cDNA microarray experiments. *Nat Rev Genet* 2002, **3**(8), 579–588.
31. Kendziorski CM, Zhang Y, Lan H, *et al*. The efficiency of pooling mRNA in microarray experiments. *Biostatistics* 2003, **4**(3), 465–477.
32. Quackenbush J. Microarray data normalization and transformation. *Nat Genet* 2002, **32**(Suppl), 496–501.
33. Fisher RA. *The design of experiments*. 9th ed. New York, Hafner Publishing Company, 1971.
34. Montgomery DC. *Design and analysis of experiments*. New York, Wiley, 1997.
35. Zien X, Fluck J, Zimmer R, *et al*. Microarrays: How many do you need?. *J Comput Biol* 2003, **10**(3–4), 653–667.
36. Smyth GK, Yang YH, Speed T. Statistical issues in cDNA microarray data analysis. *Methods Mol Biol* 2003, **224**, 111–136.
37. Tusher VG, Tibshirani G, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 2001, **98**(9), 5116–5121.
38. Efron B *et al*. Empirical Bayes analysis of a microarray experiment. *J Am Stat Assoc* 2001, **96**(456), 1151–1160.
39. Dash A, Maine IP, Varambally S, Shen R, *et al*. Changes in differential gene expression because of warm ischemia time of radical prostatectomy specimens. *Am J Pathol* 2002, **161**(5), 1743–1748.
40. Liotta L, Petricoin E. Molecular profiling of human cancer. *Nat Rev Genet* 2000, **1**(1), 48–56.
41. Schoor O, Weinschenk T, Hennenlotter J, *et al*. Moderate degradation does not preclude microarray analysis of small amounts of RNA. *Biotechniques* 2003, **35**(6), 1192–1196, 1198–1201.

42. Simon R, Sauter G. Tissue microarrays for miniaturized high-throughput molecular profiling of tumors. *Exp Hematol* 2002, **30**(12), 1365–1372.
43. Hsu FD, Nielsen TO, Alkushi A, *et al*. Tissue microarrays are an effective quality assurance tool for diagnostic immunohistochemistry. *Mod Pathol* 2002, **15**(12), 1374–1380.
44. Dhanasekaran SM, Barrette TR, Ghosh D, *et al*. Delineation of prognostic biomarkers in prostate cancer. *Nature* 2001, **412**(6849), 822–826.
45. Gibas Z, Becher R, Kawinski E, *et al*. A high-resolution study of chromosome changes in a human prostatic carcinoma cell line (LNCaP). *Cancer Genet Cytogenet* 1984, **11**(4), 399–404.
46. Kaighn ME, Narayan KS, Ohnuki Y, *et al*. Establishment and characterization of a human prostatic carcinoma cell line (PC-3). *Invest Urol* 1979, **17**(1), 16–23.
47. Pan Y, Kytola S, Farnebo F, *et al*. Characterization of chromosomal abnormalities in prostate cancer cell lines by spectral karyotyping. *Cytogenet Cell Genet* 1999, **87**(3–4), 225–232.
48. Van Bokhoven A, Varella-Garcia M, Korch C, *et al*. Widely used prostate carcinoma cell lines share common origins. *Prostate* 2001, **47**(1), 36–51.
49. Churchill GA. Fundamentals of experimental design for cDNA microarrays. *Nat Genet* 2002, **32**(Suppl), 490–495.
50. Yu J, Othman MI, Farjo R, *et al*. Evaluation and optimization of procedures for target labelling and hybridization of cDNA microarrays. *Mol Vis* 2002, **8**, 130–137.
51. Manduchi E, Scearce LM, Brestelli JE, *et al*. Comparison of different labelling methods for two-channel high-density microarray experiments. *Physiol Genomics* 2002, **10**(3), 169–179.
52. Spiess AN, Mueller N, Ivell R. Amplified RNA degradation in T7-amplification methods results in biased microarray hybridizations. *BMC Genomics* 2003, **4**(1), 44.
53. Brazma A, Hingamp P, Quackenbush J, *et al*. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 2001, **29**(4), 365–371.
54. Kuo WP, Jenssen TK, Butte AJ, *et al*. Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics* 2002, **18**(3), 405–412.
55. Diehn M, Sherlock G, Binkley G, *et al*. SOURCE: a unified genomic resource of functional annotations, ontologies, and gene expression data. *Nucleic Acids Res* 2003, **31**(1), 219–223.
56. Golub TR, Slonim DK, Tamayo P, *et al*. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999, **286**(5439), 531–537.
57. Su AI, Welsh JB, Sapinoso LM, *et al*. Molecular classification of human carcinomas by use of gene expression signatures. *Cancer Res* 2001, **61**(20), 7388–7393.
58. Su AI, Cooke MP, Ching KA, *et al*. Large-scale analysis of the human and mouse transcriptomes. *Proc Natl Acad Sci USA* 2002, **99**(7), 4465–4470.
59. Culig Z, Klocker H, Bartsch G, *et al*. Androgen receptors in prostate cancer. *J Urol* 2003, **170**(4 Pt 1), 1363–1369.
60. Nelson WG, De Marzo AM, Isaacs WB. Prostate cancer. *N Engl J Med* 2003, **349**(4), 366–381.
61. Tsavaler L, Shapero MH, Morkowski S, *et al*. Trp-p8, a novel prostate-specific gene, is up-regulated in prostate cancer and other malignancies and shares high homology with transient receptor potential calcium channel proteins. *Cancer Res* 2001, **61**(9), 3760–3769.
62. Henshall SM, Afar DE, Hiller J, *et al*. Survival analysis of genome-wide gene expression profiles of prostate cancers identifies new prognostic targets of disease relapse. *Cancer Res* 2003, **63**(14), 4196–4203.
63. Stuart RO, Wachsman W, Berry CC, *et al*. In silico dissection of cell-type-associated patterns of gene expression in prostate cancer. *PNAS* 2004, **101**(2), 615–620.