Defining best practice for microarray analyses in nutrigenomic studies

Paola Garosi[1]*, Carlotta De Filippo[2], Marjan van Erk[3], Philippe Rocca-Serra[4], Susanna-Assunta Sansone[4] and Ruan Elliott[1]

[1]*Institute of Food Research, Norwich Research Park, Norwich NR4 7UA, UK*
[2]*Department of Pharmacology, University of Florence, Florence, Italy*
[3]*TNO Nutrition and Food Research, PO Box 360, 3700 AJ Zeist, The Netherlands*
[4]*EMBL-EBI, The European Bioinformatics Institute, Hinxton, Cambridge, CB10 1SD, UK*

Microarrays represent a powerful tool for studies of diet–gene interactions. Their use is, however, associated with a number of technical challenges and potential pitfalls. The cost of microarrays continues to drop but is still comparatively high. This, coupled with the complex logistical issues associated with performing nutritional microarray studies, often means that compromises have to be made in the number and type of samples analysed. Additionally, technical variations between array platforms and analytical procedures will almost inevitably lead to differences in the transcriptional responses observed. Consequently, conflicting data may be produced, important effects may be missed and/or false leads generated (e.g. apparent patterns of differential gene regulation that ultimately prove to be incorrect or not significant). This is likely to be particularly true in the field of nutrition, in which we expect that many dietary bioactive agents at nutritionally relevant concentrations will elicit subtle changes in gene transcription that may be critically important in biological terms but will be difficult to detect reliably. Thus, great care should always be taken in designing and executing microarray studies. This article seeks to provide an overview of both the main practical and theoretical considerations in microarray use that represent potential sources of technical variation and error. Wherever possible, recommendations are made on what we propose to be the best approach. The overall aims are to provide a basic framework of advice for researchers who are new to the use of microarrays and to promote a discussion of standardisation and best practice in the field.

Microarrays: Standardisation: Statistical analysis: Experimental design

Microarray technology has been available for several years. Expression profiling using microarrays offers a powerful tool to gain a comprehensive view of biological systems by measuring the expression of thousands of genes simultaneously. The increasing interest in microarrays in recent years has led to a few attempts to standardise aspects of the protocols used (Duggan *et al.* 1999; Hedge *et al.* 2002). However, even though production capabilities and the use of microarrays are becoming increasingly well established and widespread, variation still arises due to the overall complexity of the experimental approach. Significant differences exist with regard to fabrication techniques and user protocols. Such differences make the comparison of results across platforms very difficult. All parts of the protocol (i.e. array production, RNA extraction, cDNA labelling and hybridisation, and data analysis techniques) include a multitude of parameters that need to be optimised to reach stable experimental results. The absence of approaches that are based on 'best practices' for the design, fabrication and end use of microarrays makes comparative data analysis problematic.

For commercial arrays, the use of the manufacturers' protocols already helps to achieve a degree of standardisation, at least in the labelling, hybridisation and some or all of the analytical steps. For this reason, in-house arrays form the primary focus of this review. Reference is, however, also made to a number of issues that are specific to the widely used Affymetrix GeneChips (Santa Clara, CA, USA). Furthermore, the discussion of experimental design, sample preparation and data interpretation is a generic matter that is largely independent of the platform used.

Practical aspects

RNA extraction

The quality and quantity of RNA are vitally important for assuring the reliability and reproducibility of microarray results. RNA is the material that is used to generate the labelled extract. The composition of the different species within the labelled extract should accurately reflect the gene expression profile in a tissue or cell sample. Poorly processed or partially degraded samples may not faithfully represent the true transcription profile (Hedge *et al.* 2002).

Important issues related to RNA extraction include variability in RNA yields from various tissues and states of cell activation, the processing of different quantities of tissue, and the purity of the extracted RNA. The end product of a successful RNA isolation procedure is minimally degraded total, or polyA-enriched, RNA free from contaminating agents such as RNases and other proteins, and genomic DNA.

Several conditions have been proposed as crucial for the control of variability:

1. standardisation of cultures and purification methods to achieve consistency (Benes & Muckenthaler, 2003; Forster *et al.* 2003);
2. the performance of at least three biological replicates (Benes & Muckenthaler, 2003; Tzu-Ming *et al.* 2004);
3. the optimisation of purification steps and consistency with the extraction method (Naderi *et al.* 2004);
4. rigorous quality control (Naderi *et al.* 2004; Forster *et al.* 2003; Ryan & Huffaker, 2004).

To protect RNA from degradation, tissues can be treated in two different ways: snap-freezing in liquid N_2 (particularly suitable for larger tissue samples, e.g. >50 mg) or immersion in a powerful RNase inhibitor (ideal for smaller tissue sections, in which the solution can penetrate cell membranes rapidly to inactivate cellular and other contaminating RNases). Reagents such as TRIzol (Gibco BRL; Paisley, Scotland, UK) or RNA-Later (Ambion; Huntingdon, Cambridgeshire, UK) buffers are used for this purpose and have been found to be very effective for the preservation and recovery of cellular RNA (Baker *et al.* 2004; Tzu-Ming *et al.* 2004; Ulrich *et al.* 2004; Waring *et al.* 2004).

The quality of the RNA isolated should always be assessed. This can be achieved by spectrophotometry and gel electrophoresis, but a more thorough analysis is recommended by using a system such as the Agilent Bioanalyser (Palo Alto, CA, USA). This gives accurate information about sample concentration, constituents, possible contaminants and degree of degradation.

RNA amplification

In situations in which the quantity of RNA available is limiting, the options for enabling accurate expression profiling are (1) to pool RNA samples from different sources within the same treatment group, (2) to increase cDNA labelling and hybridisation efficiency, and (3) to use an RNA/cDNA amplification strategy. At present, there are two main approaches for RNA amplification *in vitro* based on the use of T7 RNA polymerase and PCR.

RNA amplification with T7 RNA polymerase is, unlike PCR, linear and not biased by the size of the template (under optimal conditions). Consequently, most microarray amplifications are based on this method (Van Gelder *et al.* 1990). In this strategy, polyA RNA is primed for cDNA synthesis by a polyT oligonucleotide containing the 17 Bp sequence for the T7 RNA polymerase promoter at or near its $5'$ end. After second-strand synthesis has been completed, the template is then transcribed by a highly concentrated T7 polymerase, resulting in a 1000-fold amplification of RNA that can be used for hybridisation analysis. If necessary, a second round of amplification can be performed simply by repeating the procedure (Baugh *et al.* 2001). Subsequent cycles of aRNA amplification can be successfully performed, and the products can be used as labelled extracts on microarrays (Xiang *et al.* 2003).

The PCR-based strategy has been applied to evaluate the gene expression profile of single cells (Chiang & Melton, 2003). Through sequential reverse transcription and PCR amplification of the whole-cell transcriptome, the authors typically generated $10-20\,\mu g$ single-cell cDNA, demonstrating that although some distortion of the original single-cell transcript distribution cannot

be excluded, a high degree of confidence can be granted to the comparison of single-cell profiles (Chiang & Melton, 2003). RNA amplification can be used as an effective tool in microarray analysis, but it is recommend that the degree of amplification should be kept to a minimum. The possibility that amplification may introduce some skewing of the RNA or cDNA population should be compensated for by ensuring that all samples under study are processed in the same manner.

Labelling methods

The reproducibility, sensitivity and accuracy of a selection of different labelling methods in cDNA microarray hybridisation have been compared (Manduchi *et al.* 2002; Badiee *et al.* 2003). For Affymetrix GeneChips, there are two standard labelling protocols: one-cycle; and two-cycle target labelling assays (www.affymetrix.com) with robust performance for $1-15\,\mu g$ and $10-100$ ng total RNA as starting material, respectively. Additionally, the control of cRNA fragmentation is a critical step of the Affymetrix labelling protocol. Failing to perform this step consistently results in an inconsistent hybridisation signal.

For other array formats, there are three common labelling methods adopted by most scientists: direct labelling, indirect labelling and dendrimer labelling. The direct labelling method incorporates a deoxyribonucleoside triphosphate fluorescently labelled with a bulky dye adduct (Cy3 or Cy5) during reverse transcription. The indirect labelling method incorporates a reactive amine derivative of 5-(3-aminoallyl)-2'- deoxyuridine $5'$-triphosphate during reverse transcription. Subsequent to the reverse transcriptase reaction, succinimidyl esters of Cy3 or Cy5 are covalently coupled to the amino-allyl-labelled cDNAs. Both of these methods are dependent upon the efficiency of the incorporation of modified dNTPs and the sequence of the clone itself for the amount of label incorporated.

In contrast, the dendrimer labelling method is entirely dependent upon nucleic acid hybridisation kinetics. The initial reverse transcriptase reaction is primed with an oligonucleotide containing a specific dendrimer 'capture' sequence. The cDNAs containing the 'capture' sequences are first hybridised to fluorescently labelled dendrimers and then to the array. A dendrimer is a complex nucleic acid structure created by hybridising nucleotide oligomers to specifically promote the formation of a complex branched structure (Nilsen *et al.* 1997). Dendrimer labelling has the advantages of requiring less starting material and exhibiting minimal sequence or length dependency (Badiee *et al.* 2003). The disadvantage is presented by the cost.

The direct incorporation method has the advantage of being easy and quick to perform, but some researchers believe that it might introduce sequence-specific artefacts (Baugh *et al.* 2001; Kerr *et al.* 2001; Taniguchi *et al.* 2001), probably caused by the variable and differing rates with which these bulky nucleotide analogues are incorporated into the synthesised DNA molecules by reverse transcriptase (the direct incorporation protocol having a somewhat higher incorporation rate for the Cy3 dye).

The indirect labelling technique, despite being laborious and time-consuming, is still popular for several reasons:

1. The yield of cDNA is higher with the post-labelling method than with direct incorporation.
2. Indirect incorporation of the reactive dyes of Cy3 and Cy5 is greater than direct CyDye-labelled nucleotide incorporation.

3. The post-labelling method is less prone to the production of artefacts caused by the size of Cy dye-nucleotides (e.g. chain termination, proximity quenching, sequence-specific bias).
4. Post-labelling produces longer cDNAs, which is beneficial when array targets have been derived from the 5′ ends of cDNAs.

Indirect labelling is proposed as a preferable option by Wurmbach *et al.* (2003) and Benes & Muckenthaler (2003). When the amount of RNA available is limited and the RNA amplification is not feasible, the dendrimer labelling method is recommended (Badiee *et al.* 2003).

Microarray construction

Microarray slides can contain either oligonucleotides or cDNAs corresponding to the genes to be studied. Recent studies (Cao *et al.* 2003; Tan *et al.* 2003) report the results of the comparison between different microarray platforms, including the Affymetrix GeneChip, custom cDNA arrays and custom oligo arrays. They evaluated the internal consistency within each of the platforms and compared the results across the platforms. Replicate analyses on the same platforms proved highly reproducible, with correlations of 0·93–0·99. However, direct comparisons between different platforms revealed a marked lack of agreement, making it difficult to establish which kind of platform gives the best performance.

The Affymetrix method is considered by most authors to provide the most reliable and reproducible results, but the high cost induces many laboratories to prefer the printed arrays option, which, if performed carefully, also can give a satisfactory performance. The choice of home-made microarrays has led researchers to focus on the development of strategies seeking to ensure manufacture of high-quality microarrays.

Part of the variation in microarray experiments is introduced systematically during the spotting, or deposition, of the DNA on to the slide surface. The printing method is a very important issue in maintaining consistent spot morphology from array to array and in minimising variation between signal intensities.

Rickman *et al.* (2003) and Wrobel *et al.* (2003) investigated the effect of a variety of detergent additives used in the printing process on spot morphology and size, signal intensity and reproducibility. Other works describe optimised methods for the fabrication of high-quality cDNA microarrays, on the basis of key factors including printing device, spotting pins, surface chemistry (Van Hal *et al.* 2002), spotting solution, temperature and humidity (Taylor *et al.* 2003; Park *et al.* 2004).

DNA oligonucleotide microarrays have become increasingly popular. Some reports have focused on the sensitivity and specificity of different oligonucleotide-based microarray platforms, demonstrating the advantages of oligonucleotide arrays compared with cDNA arrays (Kane *et al.* 2000; Wang *et al.* 2003). The source of oligonucleotides is another important factor. Differences in reliability have been observed, suggesting that the design and possibly the quality of synthesised oligonucleotides or methods used to create them are crucial for obtaining accurate transcript profiles (Cao *et al.* 2003; Tolstrup *et al.* 2003).

Although it is true that shorter oligonucleotides (15–20 mers) promise more specific binding with labelled extracts, some reports indicate that 50–70 mers provide improved sensitivity while maintaining the excellent specificity of shorter sequences (Kane *et al.* 2000).

Classic microarrays are based on the deposition of representative sequences deduced from genome assemblies. These array designs, therefore, are plagued with decay as an increasing number of reporter sequences are determined to be inaccurate. In order to circumvent this problem, researchers have recently reported the validation of a universal array that is species and/or genome independent (Roth *et al.* 2004). The technology relies on a combination of enzymatic methods to digest and treat the transcriptome before running it on the universal microarrays. Such a method is able to survey gene expression without *a priori* knowledge of the sequence. Data can therefore be reanalysed and yield additional results as genome annotation evolves without having to print a new array design and perform new experiments. If proven scalable and effective in spite of the extensive liquid-handling steps required by the method, the universal array may well supersede current arrays in its ability to generate data that can be reprocessed easily.

Gene focus v. genome arrays

Moving away from full (or near full) genome coverage arrays, a growing number of biologists are favouring 'focused arrays', devised to follow a selected number of genes involved in a pathway of interest. The main advantage of such an approach is the ability to retain the power of microarray technology for monitoring a set of genes selected for their involvement in a specific pathway, but with higher throughput capacity and at a fraction of the cost of regular microarray experiments. However, users of such arrays should be aware of the pitfalls and limitations of the methods. The main issue when using focused arrays is that of determining a reliable baseline signal. This can be achieved only by providing significant technical replications of the genes of interest on the arrays and including a large number of external controls. Using spiked-in controls has proven to be an efficient way of monitoring signal variations in microarrays and assessing the overall quality of the experiment (van de Peppel *et al.* 2003).

Theoretical aspects

Scanning

This procedure can have a significant effect on data analysis and interpretation. Once an array has been scanned, all data, whether high or poor quality, are essentially fixed. The choice of scanning settings is very important. The major scanner settings for determining the spot intensities are the laser power and the voltage of the photomultiplier tube. A low laser power minimises photo-bleaching. Forster *et al.* (2003) describe a procedure for identifying scanner settings that provide the best representation of signal distribution on the array. Such a procedure involves the collection of a representative series of scatter plots and the selection of the scanner setting producing the optimal signal distribution.

Lyng *et al.* (2004) have investigated the relationship between the photomultiplier tube voltage, spot intensities and expression ratios for three different scanners, in order to define an optimal scanning procedure. All the scanners used showed an almost log-linear relationship between intensity and photomultiplier tube voltage within the intensity range 200–50 000 (mean spot intensity). The use of spot and background intensities outside this range led to errors in the ratios. At intensities above

50 000, error occurs due to the saturation of pixel intensities within the spots.

This usable intensity range, present in most commercial scanners, is considerably less than the maximum detection range of the photomultiplier tubes and is not suitable for an accurate and complete analysis. In fact, scanning performed by adjusting the photomultiplier tube voltage to avoid saturation may not be the optimal strategy due to the risk of achieving spot and background intensities below the usable range. In order to increase the accuracy of the data as well as the number of data points retained in each experiment, a procedure has been proposed in which two scans are obtained for each of the red and green channels. The primary scan is set up so that the lowest feature signal intensities are within the usable range of the scanner, while the secondary scan is set up so that the intensities of the brightest spots are just below the saturation level. The primary scan forms the basis of the analysis, while the secondary scan is used to correct the intensities of spots with saturation in the primary scan (Lyng *et al.* 2004).

Image analysis

Many current image analysis software packages offer automatic image-processing and spot-finding. Although useful, such software processes are not entirely accurate and can introduce errors via misalignment for spotted arrays. A common example is misplacement of the grids by full-row or column offsets. Since annotations are tracked by grid coordinates, this kind of error can lead to incorrect annotations for a large number of array features. Consequently, it is necessary to perform visual inspections and, if necessary, a manual correction of grid placement on each microarray image.

Regardless of the software package used, the accuracy of image analysis depends ultimately on the combination of two processes: segmentation and background subtraction. Segmentation is a process used to differentiate the foreground pixels (i.e. the true signal) in a spot grid from the background pixels. This is a tricky computational problem because the spot morphology in a poor-quality image can vary substantially and the background can be very high. Furthermore, array images often contain other imperfections.

There are several segmentation methods: fixed circle, adaptive circle, adaptive shape and histogram segmentation (Yang *et al.* 2001*a,b*). Whatever segmentation method is used, it is crucial to produce high-quality microarrays and collect high-quality images: when the image quality is high, the algorithm can predict the size of the spots and segment their signal accurately. If there is dust contamination or a high background signal, the algorithm not only rejects poor-quality spots, but might also incorrectly assign contaminating particles as spots. In this case, both the true signal and the background signals will be erroneously estimated (Ahmed *et al.* 2004).

After the segmentation process, the pixel intensities within the foreground and background masks (i.e. the areas in the image defined as foreground and background by the software, respectively) are averaged separately to give foreground and background intensities, respectively. Median or other extraction methods can be used when there are extreme values in the spots that skew the distribution of the pixel intensities. Subtracting the background intensity from the foreground intensity in each channel gives the spot intensity for calculating the expression ratio between the two channels. The most commonly used background

subtraction methods are local background, morphological opening and constant background. A fourth option is not to correct for background at all.

Comparisons between image analysis procedures (Yang *et al.* 2001*a*) found that the choice of background correction method has a large impact on the log-intensity ratios. It is suggested that the morphological opening method provides a better estimate of background than other methods (Yang *et al.* 2001*b*). This method is an intermediate adjustment that provides less variable estimates than local background methods and more accurate estimates than intensities calculated with background subtraction. However, in software packages in which morphological opening is unavailable, calculating log-ratios without background subtraction is suggested to be better than subtraction of a local background (Yang *et al.* 2001*b*).

Data analysis

Data normalisation

Extracting biological information from microarray requires two important steps: normalisation and statistical analysis. The first step after data acquisition aims to determine the significance of the data and to normalise the data. It is important to be able to determine the quality of every single spot and to normalise the signal. There are many sources of systematic variation in microarray experiments that affect the measured gene expression levels. The main source of variation, if the array is printed and hybridised properly, is accounted for by differences in labelling efficiency between the two fluorescent dyes. A fundamental assumption is that the amount of cDNA used to hybridise to the array was the same. According to this, the overall signal for the Cy3 and Cy5 should be the same.

The normalisation process aims to balance many of the systematic variations present in an array experiment (Quackenbush, 2002). In recent years, several normalisation techniques have been developed (Kroll & Wolfl, 2002). The underlying principle behind the different methods requires the identification of genes that are not affected by experimental conditions and thus show a ratio between the reference and the experimental sample equal to 1. To compare measurements from gene expression array experiments, quantitative data can be normalised using reference genes or global normalisation methods based on mean or median values. These methods are based on the assumption that (1) selected reference genes are expressed at a standard level in all experiments, or (2) that the mean or median signal of expression will give a quantitative reference for each individual experiment (Quackenbush, 2001).

In addition to these methods, there are a number of alternative approaches, including linear regression analysis, log-centering, rank invariant methods and Chen's ratio statistics, among others. However, none of these approaches takes into account systematic biases that may appear in the data. Several reports have indicated that the $\log_2$(ratio) values can have a systematic dependence on intensity. Lowess normalisation uses locally weighted linear regression to smooth the data. The smoothing process is considered local because each smoothed value is determined by neighbouring data points defined within the span. The process is weighted because a regression weight function is defined for the data points contained within the span. In addition to the regression

weight function, one can use a robust weight function, which makes the process resistant to outliers.

The use of a normalisation reference has a very important effect on the entire process. Housekeeping genes whose expression is expected to be constant across samples are often adopted as internal controls. However, recent studies have demonstrated that genes that do not change their expression levels in response to a variety of experimental conditions simply do not exist. Another approach to control variability in microarray experiments is to spot genomic DNA in multiple dilutions on the array and to use the signal obtained after hybridisation for normalisation purposes (Benes & Muckenthaler, 2003).

A universal approach to normalising gene expression data is the application of exogenous control genes. External controls can be used as negative controls if no corresponding mRNA is present in the RNA samples to be analysed; negative controls help to determine the 'noise' of a microarray experiment. In some cases, internal controls are a used as positive or 'spike-in' controls: exogenous RNAs are added to the reference and the experimental samples in predetermined concentrations before the synthesis of fluorescent-labelled cDNAs. Spike-in controls need to be titrated to cover the entire range of signal intensities obtained in a microarray experiment to be representative for all detectable genes. Spike-in RNAs added in equal amounts to the reference and experimental sample can serve as normalisation controls (Benes & Muckenthaler, 2003).

Statistical analysis

Replication of a microarray experiment is essential to define the variation in gene expression for statistical calculation. Every microarray experiment should be performed in at least triplicate to increase data reliability. There are two types of replication: biological and technical. Biological replication refers to the analysis of multiple independent biological samples (e.g. one tissue type obtained from different patients with the same disease, or individual samples of a particular cell line under the same treatment), whereas technical replication refers to the repetition of microarray experiment using the same extracted RNA samples. Technical replication provides a precise measurement of gene expression for a particular sample and eliminates many technical variations introduced during the experiment. Unfortunately, merely obtaining a precise expression measurement of a tissue by technical replication will not resolve the problem of biological variation. It is therefore usually preferable to have biological replication.

In the end, an average fold change will be calculated for each gene. By applying a cut-off value to the fold change, a set of differentially expressed genes can be selected. However, this method does not take into account the variance of expression of each gene between replicate arrays.

When comparing two conditions, Student's t test can be used to find significantly differentially expressed genes (Cui & Churchill, 2003). By using Student's t test, only genes showing a large change in gene expression relative to the within-treatment variance are considered to be significantly differentially expressed. Student's t test would perform well in the case of highly replicated experiments since such replication would allow accurate estimates of the variance within experimental treatments.

In microarray studies, however, samples are often available in limited supply, and thus the level of replication is low, resulting in poor estimates of variance. Furthermore, the performance of

thousands of t tests (since microarrays usually consist of thousands of genes) will lead to a large number of false positives among the genes detected as significantly differentially expressed. This multiple testing problem can be approached by calculating or estimating the false discovery rate, i.e. the proportion of false positives in the total set of differentially expressed genes (Cui & Churchill, 2003; Reiner et al. 2003). Certain statistical methods that attempt to correct for the false-discovery rate can be used (e.g. Benjamini and Hochberg, Bonferroni and Bonferroni step-down [Holm]). However, none of these provides a perfect solution as they overcompensate, leading to truly differentially regulated genes being overlooked. Nevertheless, a statistical method that takes into account the false-discovery rate, by either attempting to correct for it or estimating its magnitude, is advisable.

Use of a Bayesian statistical approach is another method to improve confidence in the interpretation of DNA microarray data with a low number of replicates. This approach is closely related to the false-discovery rate approach (Efron et al. 2001). The Bayesian statistical approach incorporates prior information of within-treatment measurements by assuming that genes of similar expression levels have similar measurement errors. Thus, the variance of a single gene can be estimated from the variance from a number of genes with a similar expression level (Long et al. 2001).

The ANOVA approach (Kerr & Churchill, 2001; Kerr et al. 2001) is a generalisation of the t test that can be used when comparing more than two conditions. The idea behind ANOVA is to build a model that takes into account the sources of variance that affect measurements, and then to use the data to estimate the variance of each individual variable in the model. By using this approach, one can distinguish between interesting variations, such as gene regulation, and side-effects, such as variations caused by different dyes or arrays (Draghici, 2002).

As far as data analysis is concerned, independently developed methods have been reported to outperform Affymetrix default software in term of sensitivity and specificity. Robust Multichip Analysis (Affymetrix, Santa Clara, CA, USA; Bolstad et al. 2003) and Match-only Integral Distribution (Novartis Research Foundation, San Diego, CA, USA; Zhou & Abagyan, 2002) have become the methods most broadly applied by Affymetrix users for processing and normalising data. These methods have now been incorporated in Affymetrix GREX software, but open source packages written in R, as bioconductor modules are still available (Ihaka & Gentleman, 1996).

Data interpretation

Selecting (groups of) differentially expressed genes from a microarray experiment should be followed by the interpretation of the data in a biological context. For successful data interpretation, it is important regularly to update information on the genes present on the arrays, like gene names and annotations. A number of Internet databases are available for this purpose. Information on function of genes can be derived from the Gene Ontology database (http://www.geneontology.org), which provides a structured annotation of genes with respect to molecular function, biological process and cellular component.

Principal component analysis can be used to reduce the high-dimensional microarray data to two or three dimensions, in which the data can then be visualised (Misra et al. 2002; Slonim, 2002). This analysis can be used to assess the similarity or dissimilarity

of expression profiles between samples, and it facilitates the selection of (groups of) relevant genes from the data.

Another interesting approach in data interpretation is pathway analysis. The visualisation of gene expression changes in (metabolic) pathways can help to evaluate which pathways are most affected in the experiment. Tools for pathway analysis include GenMAPP (Dahlquist *et al.* 2002; http://www.genmapp.org), Pathway Miner (Pandey *et al.* 2004; http://www.biorag.org/pathway.html) and Pathway Processor (Grosu *et al.* 2002; http://cgr.harvard.edu/cavalieri/pp.html).

Probably the largest amount of information on genes that could be useful for data interpretation is present in the literature. Searching information on genes and protein in literature, such as PubMed abstracts, will be facilitated by so called text-mining tools, for example iHOP (Hoffmann & Valencia, 2004; http://www.pdg.cnb.uam.es/UniPub/iHOP/).

Experimental design

Two basic types of microarray experimental designs exist: loop and reference designs (Kerr & Churchill, 2001; Hwa Yang & Speed, 2002; Kerr, 2003; Fig. 1). In the loop design, each sample is compared directly with other samples in a circular or multiple-pairwise (interwoven loop) fashion. This approach has greater statistical power than the reference design (described below) so the ability to detect differences is maximised. Also, all of the data produced in a loop design experiment are derived from test samples and are therefore useful experimental information, whereas half the data produced in a reference design experiment are redundant. The loop design may be very useful when small numbers of samples are compared, but it can

become difficult to correctly design and execute loop studies for a large number of samples. Another disadvantage of this method is that ratios observed across different pairwise comparisons are not immediately comparable, and visualisation of the results in a form that is easy to interpret is made more difficult (Townsend, 2003).

When a large number of samples are analysed, the reference design is preferable. In this method, each sample is compared with a common RNA reference sample, serving as a common denominator between different microarray hybridisations (Kim *et al.* 2002; Novoradovskaya *et al.* 2004). Reference RNA can be used for time course experiments in which the response of cells to drugs or other perturbations to the biological system is monitored over time. In addition, comparing microarray datasets produced in different laboratories will be made more reliable by employing the use of reproducible common reference RNA and can also be used to normalise data from one set of Affymetrix experiments to another.

In studies evaluating expression differences in perturbed and non-perturbed systems that are otherwise identical, RNA isolated from the non-perturbed state generally serves as an excellent reference as one can easily see changes in the perturbed state in pairwise comparisons. However, when microarray studies involve complex sets of comparisons between large numbers of samples, the identification of a single reference RNA source can be difficult. One solution that has been widely used is to create an RNA pool derived either from all the samples under study or from independent collections of samples selected to achieve the widest possible representation of gene expression. Yang *et al.* (2002) showed that RNA pools from cell lines can be used to make efficient reference samples. A problem is the difficulty in identifying samples that provide complete coverage for all of the genes represented on the array. Also, should the initial RNA reference sample be exhausted, it is often difficult to precisely reconstruct a new supply. The use of a universal commercial reference may solve the problem and provide a solution for the standardisation and cross-referencing of microarray experiments by offering a high-quality standard for accurate and consistent data comparison (Novoradovskaya *et al.* 2004).

Recently, the External RNA Control Consortium (http://www.cstl.nist.gov/biotech/biotech/workshops.html), an industry-led consortium, has evolved from a meeting sponsored by the US National Institute of Standards and Technology. The goal of the External RNA Control Consortium is to develop a universal (platform-independent) set of RNA spike-in control materials for performance evaluation and monitoring of gene expression experiments (using microarray and/or RT-PCR).

Cross-site comparisons of gene expression data

The issues described above demonstrate that microarray assays are complex, multistep procedures involving array fabrication, RNA extraction, cDNA labelling, hybridisation and data analysis. Laboratories around the world have developed a variety of protocols for each of these steps. It is obvious that this range of available methodology results in data variability between laboratories. However, the many different technologies available may all be valid, and it would be unrealistic to impose on users any particular platform, software or methods of data analysis.
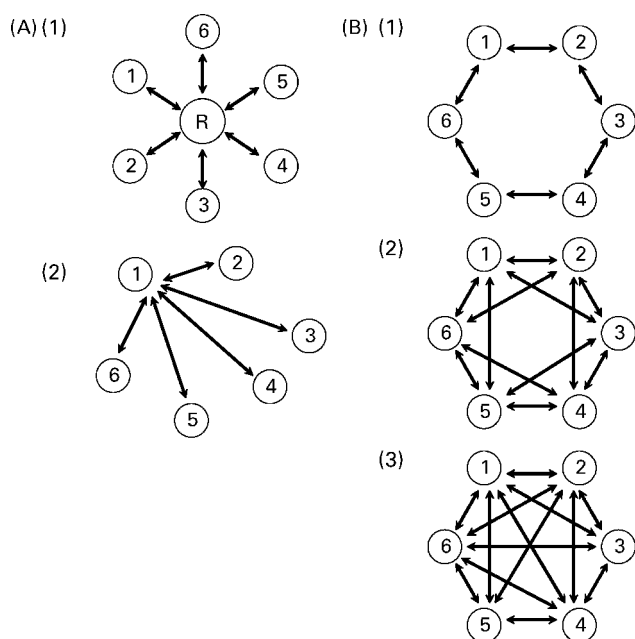

Fig. 1. Common strategies for microarray experimental design. (A) Reference design: (1) Each sample is tested against a single, common standard. This standard can be a commercial or in-house reference, a pooled sample made up of equal parts of all the text samples, or (2) one of the test samples. (B) Loop design: (1) Simple loop and (2, 3) interwoven loop design. Each sample is compared head-to-head with the other samples in a circular fashion. Each comparison is performed with a dye swap.

The standardisation of microarray experiment reporting has been proposed by the Microarray Gene Expression Data Society (www.mged.org) with the creation of the Minimum Information About a Microarray Experiment (MIAME) project, which describes the minimum information required to ensure that microarray data can be easily interpreted and that results derived from its analysis can be independently verified (Brazma *et al.* 2001). The aim of this project is to facilitate the integration of microarray data generated by different research groups on different array platforms. Several journals have already followed the Microarray Gene Expression Data Society call for the standardisation of microarray data publications in that, upon submission, datasets must be MIAME compliant and made publicly available in one of the main public repositories (Array Express: www.ebi.ac.uk/microarray/ArrayExpress; National Center for Biotechnology Information Gene Expression Omnibus: www.ncbi.nlm.nih.gov/geo). Currently, the minimal information includes six parts:

1. experiment goals and design;
2. array design and manufacture;
3. biological materials used and a description of their creation;
4. hybridisation and wash procedure;
5. scan information and image analysis;
6. normalisation controls.

It is clear that, without this information, the meaningful comparison and integration of data generated in different laboratories on different platforms will be impaired, and errors or misunderstanding could go undetected. A detailed description of each part and convenient checklist are available on the MIAME website (www.mged.org/Workgroups/MIAME/miame.html).

Another project promoted by Measurements for Biotechnology (www.mfbprog.org.uk) is evaluating the normalisation strategies currently in use with the aim of setting up recommendations for best practice in array normalisation. Data from this project will be publicly available in Array Express.

The Hepatotoxicity Working Group of the International Life Sciences Institute recently developed a collaborative scientific programme consisting of a cross-site microarray study aiming at evaluating the consistency of data generated across multiple sites (Baker *et al.* 2004; Chu *et al.* 2004; Ulrich *et al.* 2004; Waring *et al.* 2004). The goal of this project was to evaluate and compare biological and gene expression responses in rats to two well-studied hepatotoxins (clofibrate and methapyrilene), using a standard experimental protocol, and to address the following issues: (1) how comparable are the biological and gene expression data from different laboratories running identical *in vivo* studies; (b) how reproducible are the data generated across laboratories using the same microarray platform; and (c) how do data compare using different platforms? In the study conducted by Waring *et al.* (2004), a single platform was used at different sites, while Ulrich *et al.* (2004) described a comparison between different array platforms (membrane and glass). The results showed a generally high level of concordance, although with some variability, which appears to be due largely to differences in the experimental and data analysis procedures used in each laboratory and platform differences.

Conclusion

DNA microarray technology is a powerful technique for monitoring changes in gene expression on a global scale in different systems. Microarray assays are complex, multistep procedures for which different laboratories have developed a variety of protocols. This range of available methodologies results in data variability between laboratories and, in order to make the results comparable across laboratories, great effort will be needed in the careful planning, experimental design, analysis of data and interpretation of microarray experiments.

References

Ahmed AA, Vias M, Iyer NG, Caldas C & Brenton JD (2004) Microarray segmentation methods significantly influence data precision. *Nucleic Acids Res* **32**, e50.

Badiee A, Eiken HG, Steen VS & Lovlie R (2003) Evaluation of five different cDNA labelling methods for microarrays using spike controls. *BMC Biotechnol* **3**, 23.

Baker VA, Harries HM, Waring JF, et al. (2004) Clofibrate-induced gene expression changes in rat-liver: a cross laboratory analysis using membrane cDNA arrays. *Environ Health Perspect* **112**, 428–438.

Baugh LR, Hill AA, Brown E & Hunter CP (2001) Quantitative analysis of mRNA amplification by in vivo transcription. *Nucleic Acid Res* **29**, E29.

Benes V & Muckenthaler M (2003) Standardization of protocols in cDNA microarray analysis. *Trends Biochem Sci* **28**, 244–249.

Bolstad BM, Irizarry RA, Astrand M & Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193.

Brazma A, Hingkamp P, Quackenbush J, et al. (2001) Minimum information about a microarray experiment (MIAME) – toward standards for microarray data. *Nat Genet* **29**, 365–371.

Cao YA, Lee SY, Kim JW, Chang MS & Choi S (2003) Cross comparison of DNA microarray platforms. Brief communication on: www.signaling-gateway.org/reports/v1/DA0010.pdf

Chiang MK & Melton DA (2003) Single-cell transcript analysis of pancreas development. *Dev Cell* **4**, 383–393.

Chu TM, Deng S, Wolfinger R, Paules RS & Hamadeh HK (2004) Cross-site comparison of gene expression data reveals high similarity. *Environ Health Perspect* **112**(4), 449–455.

Cui X & Churchill GA (2003) Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol* **4**, 210–219.

Dahlquist KD, Salomonis N, Iranians K, Lawlor SC & Conklin BR (2002) GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat Genet* **31**, 19–20.

Draghici S (2002) Statistical intelligence: effective analysis of high-density microarray data. *Drug Discov Today* **7**, Suppl., 55–63.

Duggan DG, Bittner M, Chen Y, Meltzer P & Trent JM (1999) Expression profiling using cDNA microarrays. *Nat Genet* **21**, 10–14.

Efron B, Tibshirani R, Storey JD & Tusher V (2001) Empirical Bayes analysis of microarray data. *J Am Stat Assoc* **96**, 1151–1160.

Forster T, Roy D & Ghazal P (2003) Experiments using microarray technology: limitations and standard operating procedures. *J Endocrinol* **178**, 195–204.

Grosu P, Townsend JP, Hart DL & Cavalieri D (2002) Pathway Processor: a tool for integrating whole-genome expression results into metabolic networks. *Genome Res* **12**, 1121–1126.

Hedge PQ, Rong K, Abernathy C, Gay S, Dharap R, Gaspards J, Earle-Hughes E & Snerud E (2002) A concise guide to cDNA microarray analysis. II. *Biotechniques* **29**, 548–562.

Hoffmann R & Valencia A (2004) A gene network for navigating the literature. *Nat Genet* **36**, 664.

Hwa Yang Y & Speed T (2002) Design issues for cDNA microarray experiments. *Nat Rev Genet* **3**, 579–588.

Ihaka R & Gentleman R (1996) A language for data analysis and graphics. *J Comput Graph Stat* **5**, 299–314.

Kane MD, Jatkoe TA, Stumpf CR, Lu J, Thomas JD & Madore SJ (2000) Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acid Res* **28**, 4552–4557.

Kerr MK (2003) Design considerations for efficient and effective microarray studies. *Biometrics* **59**, 822–828.

Kerr MK & Churchill GA (2001) Experimental design for gene expression microarrays. *Biostatistics* **2**, 183–201.

Kerr K, Martin M & Churchill G (2001) Analysis of variance for gene expression microarray data. *J Comput Biol* **7**, 23–27.

Kim H, Zhao B, Snerud EC, Haas BJ, Town CD & Quackenbush J (2002) Use of RNA and genomic DNA references for inferred comparison in DNA microarray analyses. *Biotechniques* **33**, 924–930.

Kroll TC & Wolfl S (2002) Ranking: a closer look on globalisation methods for normalisation of gene expression arrays. *Nucleic Acids Res* **30**, e50.

Long AD, Mangalam HJ, Chann BYP, Tolleri L, Hatfield GW & Baldi P (2001) Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework. *J Biol Chem* **276**, 19937–19944.

Lyng H, Badiee A, Svendsrud D, Hovig E, Myklebost O & Stokke T (2004) Profound influence of microarray scanner characteristics on gene expression ratios: analysis and procedure for correction. *BMC Genomics* **5**, 10–19.

Manduchi E, Scearce LM, Brestelli JE, Grant GR, Kaestner KH & Stoeckert CJ (2002) Comparison of different labelling methods for two-channel high-density microarray experiments. *Physiol Genomics* **10**, 169–179.

Misra J, Schmitt W, Hwang D, Hsiao LL, Gullans S & Stephanopoulos G (2002) Interactive exploration of microarray gene expression patterns in a reduced dimensional space. *Genome Res* **12**, 1112–1120.

Naderi A, Ahmed AA, Barbosa-Morais NL, Aparicio S, Brenton JD & Caldas C (2004) Expression microarray reproducibility is improved by optimising purification steps in RNA amplification and labelling. *BMC Genomics* **5**, 9–22.

Nilsen TW, Grayzel J & Prensky W (1997) Dendritic nucleic acid structures. *J Theor Biol* **187**, 273–284.

Novoradovskaya N, Whitfield ML, Basehore LS, *et al.* (2004) Universal reference RNA as a standard for microarray experiments. *BMC Genomics* **5**, 20–32.

Pandey R, Guru RK & Mount DW (2004) Pathway Miner: extracting gene association networks from molecular pathways for predicting the biological significance of gene expression microarray data. *Bioinformatics* **20**, 2156–2158.

Park CH, Jeong HJ, Jung JJ, Lee GY, Kim SC, Kim TS, Yang SH, Chung HC & Rha SY (2004) Fabrication of high quality cDNA microarray using a small amount of cDNA. *Int J Mol Med* **13**, 675–679.

Quackenbush J (2001) Computational analysis of microarray data. *Nat Genet* **2**, 418–427.

Quackenbush J (2002) Microarray normalization and transformation. *Nat Genet* **32**, 496–501.

Reiner A, Yekutieli D & Benjamini Y (2003) Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* **19**, 368–375.

Rickman DS, Herbert CJ & Aggerbeck LP (2003) Optimising solutions for increased reproducibility of cDNA microarrays. *Nucleic Acid Res* **31**, e109.

Roth ME, Feng L, McConnell KJ, *et al.* (2004) Expression profiling using a hexamer-based universal microarray. *Nat Biotechnol* **22**, 418–426.

Ryan MM & Huffaker SJ (2004) Application and optimisation of microarray technologies for human post-mortem brain studies. *Biol Psychiatr* **55**, 329–336.

Slonim DK (2002) From patterns to pathways: gene expression data analysis comes of age. *Nat Genet* **32**, Suppl., 502–508.

Tan PK, Downey TJ, Spitznagel JR, *et al.* (2003) Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acid Res* **31**, 5676–5684.

Taniguchi M, Miura K, Iwao H & Yamanaka S (2001) Quantitative assessment of DNA microarrays – comparison with Northern blot analysis. *Genomics* **71**, 34–39.

Taylor S, Smith S, Windle B & Guiseppi-Elie A (2003) Impact of surface chemistry and blocking strategies on DNA microarrays. *Nucleic Acid Res* **31**, e87.

Tolstrup N, Nielsen N, Kolberg JG, Frankel AM, Vissing H & Kauppinen S (2003) OligoDesign: optimal design of LNA (locked nucleic acid) oligonucleotide capture probes for gene expression profiling. *Nucleic Acid Res* **31**, 3758–3762.

Townsend JP (2003) Multifactorial expression design and transitivity of ratios with spotted DNA microarrays. *BMC Genomics* **4**, 41–49.

Tzu-Ming C, Deng S, Wolfinger R, Paules RS & Hamadeh HK (2004) Cross-site comparison of gene expression data reveals high similarity. *Environ Health Perspect* **112**, 449–455.

Ulrich RG, Rockett JC, Gibson GG & Pettit SD (2004) Overview of an interlaboratory collaboration on evaluating the effects of model hepatotoxicants on hepatic gene expression. *Environ Health Perspect* **112**, 423–427.

van de Peppel J, Kemmeren P, van Bakel H, Radonjic M, van Leenen D & Holstege FC (2003) Monitoring global messenger RNA changes in externally controlled microarray experiments. *EMBO Rep* **4**, 387–393.

Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD & Eberwine JH (1990) Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci USA* **87**, 1663–1667.

Van Hal FNLW, Vorst O, Kramer E, Hall DR & Keijer J (2002) Factors influencing cDNA microarray hybridisation on silylated glass slides. *Anal Biochem* **308**, 5–17.

Wang H-Y, Malek RL, Kwitek AE, Greene AS & Luu TV (2003) Assessing unmodified 70-mer oligonucleotide probe performance on glass slide microarrays. *Genome Biol* **4**, R5.

Waring JF, Ulrich RG, Flint N, Morfitt D, Kalkuhl A, Staedtler F, Lawton M, Beekman JM & Suter L (2004) Interlaboratory evaluation of rat hepatic gene expression changes induced by methapyrilene. *Environ Health Perspect* **112**, 439–448.

Wrobel G, Schlingemann J, Hummerich L, Kramer H, Lichter P & Hahn M (2003) Optimisation of high-density cDNA microarray protocols by 'design experiments'. *Nucleic Acid Res* **31**, e67.

Wurmbach E, Yuen T & Sealfon SC (2003) Focused microarray analysis. *Methods* **31**, 306–316.

Xiang C, Chen M, Ma L, Phan QN, Inman JM, Kozhich OA & Brownstein MJ (2003) A new strategy to amplify degraded RNA from small tissue samples for microarray studies. *Nucleic Acids Res* **31**, e53.

Yang YH, Buckley MJ, Dudoit S & Speed TP (2001a) *Comparison of Methods for Image Analysis on cDNA Microarray Data. Technical report* no. 584. Berkley, CA: Department of Statistics, University of California.

Yang HY, Buckley MJ & Speed TP (2001b) Analysis of cDNA microarray images. *Brief Bioinform* **2**, 341–349.

Yang IV, Chen E, Hasseman JP *et al.* (2002) Within the fold: assessing differential expression measures and reproductability in micro:array assays. *Genome Biol* **24**, 3.

Zhou Y & Abagyan R (2002) Match-only Integral Distribution (MOID) algorithm for high-density oligonucleotide array analysis. *BMC Bioinformatics* **3**, 3–11.