A guide to issues in microarray analysis: application to endometrial biology

Christine A White1,2 and Lois A Salamonsen1

1Prince Henry’s Institute of Medical Research, PO Box 5152, Clayton, Victoria, 3168, Australia and 2Dept of Obstetrics & Gynaecology, Monash University, Clayton, Victoria, 3168, Australia

Correspondence should be addressed to C A White; Email: christine.white@phimr.monash.edu.au

Abstract

Within the last decade, the development of DNA microarray technology has enabled the simultaneous measurement of thousands of gene transcripts in a biological sample. Conducting a microarray study is a multi-step process; starting with a well-defined biological question, moving through experimental design, target RNA preparation, microarray hybridisation, image acquisition and data analysis – finishing with a biological interpretation requiring further study. Advances continue to be made in microarray quality and methods of statistical analysis, improving the reliability and therefore appeal of microarray analysis for a wide range of biological questions. The purpose of this review is to provide both an introduction to microarray methodology, as well as a practical guide to the use of microarrays for gene expression analysis, using endometrial biology as an example of the applications of this technology. While recommendations are based on previous experience in our laboratory, this review also summarises the methods currently considered to be best practice in the field.

Reproduction (2005) 130 1–13

Principles of microarray analysis

As for many other techniques used in molecular biology, microarrays rely on the complementarity of the DNA duplex, i.e. that the two strands will always reassemble with base pairing A to T and C to G. In addition, single-stranded DNA will bind strongly to a solid support, where it is available for hybridisation with complementary DNA (cDNA). In a DNA microarray experiment, many gene-specific ‘probes’ are immobilised on a solid support (usually nylon membrane or glass) and the array is exposed to labelled cDNA ‘targets’ derived from one or more biological samples (Schena et al. 1995). Nylon membrane arrays are normally hybridised with a single cDNA population, labelled with a radioactive or chemiluminescent tag, so that the intensity of the signal generated by each bound probe indicates the abundance of that transcript in the sample. Transcript abundance in different samples (e.g. treatment and control) can then be compared across serial or parallel array hybridisations. The advantage of the glass DNA microarray is that cDNA from two or more biological samples can be labelled with different fluorescent dyes and competitively hybridised, so that the relative abundance of gene transcripts can be determined by the fluorescent signal obtained (Figure 1). This review will focus on the techniques and data analysis associated with the two most common cDNA microarray platforms; dual colour fluorescence on glass and radioactively labelled nylon membranes. A summary of some of the key issues involved in microarray analysis is provided in Table 1.

Choice of microarray

As most laboratories are not equipped with robotic printers, the investigator must usually obtain microarrays from commercial or other sources. There are a range of mouse, rat and human microarrays available, and they fall into two broad categories; cDNA and high-density synthetic oligonucleotide (reviewed in Barrett & Kawasaki 2003). As described above, cDNA microarrays can be based on either nylon membrane or dual colour fluorescence on glass. Both formats of cDNA microarray involve the deposition of purified and/or PCR-amplified DNA in solution onto the solid support at defined locations using a robotic printer (Figure 1). The quality of the microarray is therefore dependent on the performance of the print tips, which must deliver reproducible volumes and uniform spot sizes to enable effective data analysis. The main advantages of cDNA microarrays are their relatively low cost (~US $150 per slide) and greater flexibility in terms of producing and spotting custom-made clone sets.

Manufacturers of high-density synthetic oligonucleotide microarrays, such as Affymetrix (Santa Clara, CA, USA;
Data validation

† Data analysis

Target RNA preparation

† Experimental design

Table 1 Some key issues involved in microarray analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental design</td>
<td>• Consider the biological question(s) and the ability to achieve</td>
</tr>
<tr>
<td></td>
<td>statistical significance</td>
</tr>
<tr>
<td></td>
<td>• Seek expert statistical advice during the early planning stages</td>
</tr>
<tr>
<td></td>
<td>• Microarray experiments have multiple sources of variation and must</td>
</tr>
<tr>
<td></td>
<td>be carefully controlled</td>
</tr>
<tr>
<td></td>
<td>• Biological and technical replication are essential</td>
</tr>
<tr>
<td></td>
<td>• Sample pooling should be avoided if accurate sample synchronisation is not possible</td>
</tr>
<tr>
<td></td>
<td>• Microarray analysis of purified cells will only reveal genes expressed by these cells, but removal from the in vivo microenvironment may alter gene expression</td>
</tr>
<tr>
<td></td>
<td>• There are limitations in the use of both whole tissue and purified cells, which may necessitate the use of microdissection and RNA amplification techniques</td>
</tr>
<tr>
<td></td>
<td>• When using clinical samples, detailed patient history and tissue histopathology are critical to the interpretation of gene expression profiles</td>
</tr>
<tr>
<td>Target RNA preparation</td>
<td>• The quality of the target RNA is one of the most important factors in the success or failure of a microarray experiment</td>
</tr>
<tr>
<td>Data analysis</td>
<td>• While critical to the outcome of a microarray experiment, statistical analysis of microarray data is not well understood by many biologists and expert advice should be sought</td>
</tr>
<tr>
<td>Data validation</td>
<td>• The biomedical research community does not yet accept that microarray data can stand alone without independent validation</td>
</tr>
<tr>
<td></td>
<td>• The investigator must decide which genes to examine further, and those with larger fold changes and statistical significance are often the best candidates</td>
</tr>
<tr>
<td></td>
<td>• To describe a biological event or system, gene expression data obtained by microarray analysis must be extended to the study of protein products</td>
</tr>
</tbody>
</table>

http://www.affymetrix.com), use photolithography and solid-phase DNA synthesis to generate synthetic 25 base polynucleotides (25mers) directly on the glass surface (Lipschutz et al. 1999). Each gene is represented by 11 to 20 different 25mers, in ‘perfect match’ or ‘mismatch’ sequence pairs. Probes can be generated representing a unique part of a gene transcript, enabling discrimination between closely related genes or splice variants and the ‘mismatch’ sequences provide an internal control for every gene. Longer oligonucleotide (50 to 100mers) microarrays are also available, which provide even greater hybridisation specificity (Barrett & Kawasaki 2003). Oligonucleotide microarrays are hybridised with a single fluorescently labelled sample and gene expression in different samples compared across multiple microarrays. The main disadvantage of these microarrays is their high cost (up to US$800 per slide), so their appeal is likely to increase as they become more economical.

Commercially-available microarrays are printed or synthesised with a particular clone set (reviewed in Bowtell 1999) and these differ in the proportion of known genes and expressed sequence tags (ESTs). Some ESTs correspond to a segment of a known gene, but most represent partially sequenced novel genes. A number of groups including those at Merck, Washington University, the IMAGE (Integrated Analysis of Genomes and their Expression) Consortium and the Cancer Genome Anatomy Project (CGAP) have been responsible for sequencing over one million human ESTs (Bowtell 1999). The online database dbEST (a division of GenBank; http://www.ncbi.nlm.nih.gov/dbEST/index.html) houses all these EST sequences and the automated process known as UniGene assigns overlapping sequences to a single cluster, which may or may not have a known identity (http://www.ncbi.nlm.nih.gov/UniGene/index.html). As the human genome sequence is effectively complete (Venter et al. 2001) it is expected that all ESTs will progressively be assigned an identity. Until then, careful consideration should be given to whether identifying differentially expressed ESTs is a priority in any particular microarray experiment. If not, then using a more tailored array specific to a cellular process or pathway may be more appropriate and cost-effective.

Experimental design

The most important considerations in microarray experimental design are the biological question under study and the ability to achieve statistical significance (reviewed in Churchill 2002, Kerr & Churchill 2001, Smyth et al. 2003, Yang & Speed 2002). Answering the biological question may require identification of downstream target genes, and may also involve a time course. Designing a microarray experiment with sufficient statistical power requires input from a statistician or bioinformatician with experience in microarray technology. The right statistical advice during early planning can save vast amounts of time and money. The reason for their complexity is that microarray experiments have multiple sources of variation, each of which must be considered in the experimental design (reviewed in Churchill 2002, Chen et al. 2004). Firstly, there is biological variation between animals or patients. Secondly, technical variation arises from the RNA extraction, reverse transcription, label incorporation and hybridisation steps. Thirdly, measurement errors occur due to differences in hybridisation efficiencies between spots, between different print-tip groups across the array and between slides in the same and different print runs. Technical variation and measurement error can also interact. For example, the scanning properties of the fluorescent dyes can vary with the spot intensity and spatial position on the slide (Smyth & Speed 2003).

A conventional power analysis requires prior knowledge of the variance of individual measurements, the
magnitude of the effect to be detected, the acceptable false-positive rate and the desired 'power' of the calculation; that is, the probability of detecting an effect of the specified or greater magnitude (Yang & Speed 2002, Yang et al. 2003, Chen et al. 2004). In a microarray experiment, two of these components are unknown; both the variance of the expression ratio measurements and the magnitude of the effects of interest will be different for every gene on the microarray. To overcome this, power calculations can be performed using the median variance across all of the genes in a previous microarray hybridisation (Yang & Speed 2002).

It is critical that microarray experiments are carefully controlled, particularly when using dual colour fluorescence microarrays in which the endpoint is a ratio of expression between two or more samples. As in any experiment, treatment controls must be carefully incorporated into the study design. To ensure that there is only one source of experimental variation, consistency must also be applied to tissue collection, processing and RNA extraction, as well as the microarray hybridisations. Even with a single variable, such as a differentiation stimulus, it is possible to end up comparing cells or tissues in completely different physiological states. In this situation, differentially expressed genes will likely be the consequence, rather than the cause, of the differences in phenotype. This problem can be minimised by using carefully controlled inducible systems and examining early rather than later time points.

When an experiment involves comparisons across multiple dual colour fluorescence microarrays, there are a number of possible design matrices (Figure 2). Hybridising an appropriate reference sample to each microarray (Figure 2A) can provide a consistent control across multiple slides. The ideal reference contains all possible mRNA transcripts present in the experimental samples, so that fluorescence ratio measurements less than zero cannot occur. This is usually achieved by generating a pool of RNA from multiple samples of the tissue or cell type under study. Another approach is to create a reference mixture of all the PCR products spotted on the microarray (Sterrenburg et al. 2002). Although there are some advantages to using a reference sample, it is more precise and economical to make the critical comparisons directly on the same microarray (Figure 2B; Kerr & Churchill 2001, Churchill 2002). As a reference design requires more complicated statistical analysis, it should only be used for a well-defined purpose. It may also be useful to use a common reference if a large number of experimental samples are to be collected and analysed over a long period of time. A saturated design (Figure 2C) may be used when an experiment has more than two treatments, and all comparisons are of interest in answering the biological question. The efficiency of time course experiments can be maximised by a loop design (Figure 2D). Regardless of the design matrix, care should be taken that all pairings are biologically relevant and controlled. For example, pairs could be wild type and knockout littermates, or isolated cells from the same endometrial biopsy with and without treatment.

Replication is critical in a microarray experiment, as it enables the data to be effectively analysed using formal

---

**Figure 1** Dual colour fluorescence cDNA microarray analysis. Clones of interest (probes) are amplified by PCR and printed onto treated glass slides using a robotic printer. Total RNA samples extracted from treated and control cells/tissues (targets) are reverse transcribed and labelled with either Cy3 (green) or Cy5 (red). The samples are combined and competitively hybridised to the microarray under stringent conditions. Following washing to remove non-hybridised target, laser excitation is applied and the emissions measured in each colour channel. Specialised software is used to attach gene names, fluorescence intensity values and intensity ratios to each spot, which are then exported for advanced statistical analysis to identify differentially expressed genes.
statistical methods, and the results of that analysis to be broadly applicable to the sampled population. Given the sources of variation described above, replication should be both biological and technical. Biological replication is the use of RNA samples from multiple animals or patients. Technical replication (i.e. repeated measures) includes the presence of duplicate DNA probes on the microarray and hybridisation of the same RNA sample on multiple microarrays. As biological variability is usually greater than technical variability, the hybridisation of independent RNA samples should be prioritised. Replication allows the investigator to identify and remove false-positives and false-negatives, so only reproducible data are considered for further analysis. A single hybridisation may be justified if the aim is to generate hypotheses for further testing, but advanced statistical analysis will be more productive on 3 or more replicates per treatment (reviewed in Sasik et al. 2004). In the case of dual colour fluorescence microarrays, dye swap replicates should also be performed to account for unequal dye incorporation and quenching (Figure 2). To avoid or minimise bias, the assignment of dye labels should be randomised.

Pooling samples is often considered when RNA is in limited supply, or to minimise the effects of biological variation. In addition, it has been demonstrated that pooling RNA from an increased number of subjects can reduce the number of microarrays required, without any loss of precision (Kendziorski et al. 2003, Peng et al. 2003). Sample pooling should be avoided when it is not possible to accurately synchronise the samples (Sasik et al. 2004). For example, when using pseudopregnant mice on the same day after vaginal plug detection, there is likely to be variation in the time at which mating occurred and therefore in the physiological state of the uterus. In this situation, it is better to maintain sample independence and find common gene expression features at the data analysis stage.

Careful consideration should be given to the use of whole tissue or purified homogeneous cell populations in a microarray study. The advantage of using whole tissue is that there is a greater amount of RNA available for technical replicates and subsequent validation studies (see ‘Data validation’ below). It is important to consider, however, that tissues contain a range of different cell types. Whole endometrial biopsies, for example, contain luminal

---

**Figure 2** Microarray experimental designs (adapted from Churchill 2002, Smyth et al. 2003). Letters refer to different treatments/genotypes and subscripts indicate biological replicates. Each arrow represents one microarray, with the arrow pointing away from the Cy3 (green) labelled sample and towards the Cy5 (red) labelled sample. Double arrows indicate dye-swap pairs. A, Indirect comparison with a common reference; B, Direct comparison; C, Saturated design and D, Time course experiment loop design.
and glandular epithelium, non-decidualised and decidualised stroma, endothelial cells, smooth muscle cells and leukocytes. If the cell type of interest makes up a small proportion of the total tissue, whole tissue gene expression data may not be particularly informative. Microarray analysis of purified cells will only reveal genes expressed by those cells, but they will also have been removed from their in vivo microenvironment and cultured under conditions which are likely to alter gene expression. The limitations of both approaches need to be balanced against the aims of the experiment and perhaps additional technologies such as laser capture microdissection (Emmert-Buck et al. 1996) considered. New RNA amplification methods (see ‘Target RNA preparation’ below) have improved the feasibility of using microdissected tissue for gene expression studies and this approach has the advantage of maintaining close to in vivo cellular context.

The parallel measurement of other biological parameters can be used to assist the interpretation of microarray data. For example, variation in prolactin secretion levels from different preparations of decidualising human endometrial stromal cells may correlate with variations in the expression levels of other genes. When using clinical samples, patient history and tissue histopathology are also critical in the final interpretation of gene expression profiles.

Target RNA preparation

The target in microarray analysis is a labelled population of cDNAs representing the mRNA repertoire of a cell type or tissue. The purity of the extracted total RNA is one of the most critical factors in the success or failure of a microarray experiment. Residual chloroform, phenol or ethanol from the extraction process can interfere in the efficiency of reverse transcription to cDNA, and other contaminants such as cellular protein, lipids and carbohydrates can cause non-specific binding of fluorescent cDNAs to the glass surface (Duggan et al. 1999). For most tissues, including human endometrium, optimal quality total RNA can be extracted using TRIzol Reagent (Invitrogen, Carlsbad, CA, USA), followed by column purification using the RNeasy Minikit (Qiagen, Hilden, Germany). Highly pure total RNA can be extracted from isolated cells using the RNeasy kit alone. Both of these methods should be followed by removal of genomic DNA. It is recommended that a number of different RNA extraction methods be compared for the tissue or cell type of interest, to maximise RNA quality and yield.

Particularly in microarray experiments, it is critical to accurately measure RNA quality and quantity, to minimise variation and therefore improve labelling and hybridisation consistency. The standard UV spectrophotometer is useful for an initial estimate of RNA quality and quantity. Optical density (OD) can be measured at 230, 260 and 280 nm and RNA purity considered acceptable at values of OD_{260/230} 1.8–2.0 and OD_{260/280} < 1.0. Agarose gel electrophoresis may then be used to further

confirm RNA integrity. However, other systems such as the RiboGreen Assay (Molecular Probes, Eugene, OR, USA) or the Agilent 2100 Bioanalyzer (Agilent Technologies UK Ltd, Cheadle, Cheshire, UK) are more sensitive and accurate. The Agilent Bioanalyzer requires only 50–500 ng of total RNA and produces a detailed electropherogram which will reveal any RNA degradation or genomic DNA contamination.

The amount of RNA required per hybridisation is the greatest limitation to the use of this technology, particularly when the tissue of interest is in limited supply, or when using isolated cell populations. Until very recently, it was recommended that 50–200 µg total RNA per sample be used for each hybridisation to generate a sufficient signal (Duggan et al. 1999). However, improved RNA purification, fluorescent labelling methods and hybridisation conditions have reduced this requirement to 5–10 µg for both glass and nylon membrane arrays. As this amount of RNA may still be difficult to obtain in some systems, a number of different RNA amplification approaches have been developed. The most commonly used method is T7 polymerase in vitro transcription (IVT; van Gelder et al. 1990). While this method can significantly reduce the RNA requirement for each hybridisation, it is also expensive, time-consuming and labour-intensive. In addition, multiple rounds of amplification may be required, which decreases the linearity of the amplification and may result in a cDNA target which is no longer representative of the original sample (Petalidis et al. 2003). Newly developed PCR-based cDNA amplification techniques can decrease the amount of starting total RNA required to 200 ng, while maintaining amplification linearity (Petalidis et al. 2003). This method vastly improves the feasibility of glass microarray studies on clinical samples.

Microarray hybridisation

Effective hybridisation of the target to the microarray is essential in obtaining high quality data. Coverslips with raised Teflon edging (Lifter Slips; Erie Scientific, Portsmouth, NH, USA) are a useful way of ensuring the full hybridisation volume maintains contact with the arrayed probes. Hybridisation chambers, such as those available from Corning (Corning, NY, USA), allow the microarrays to be submerged in water of a set temperature, greatly improving hybridisation consistency across the slide. Commerciially available buffers such as ExpressHyb Hybridisation Solution (Clontech, Palo Alto, CA, USA) are very effective for nylon membrane microarray hybridisation (Evans et al. 2003). The use of non-specific blocking agents including Cot-1 DNA, polyadenylic acid and salmon sperm DNA ensures that the signal detected from each spot is specific to the particular probe sequence and background is minimised. Most microarrays include positive and negative (printing buffer alone) hybridisation controls, as well as spiked controls and RT efficacy controls. Internal control spots can be useful for assessing
data quality, but it is not recommended that they be used to standardise the data, as a housekeeping gene would be used in quantitative RT-PCR (Churchill 2002). Optimal target quality and hybridisation conditions will ensure that the maximum microarray sensitivity is achieved, allowing even very low abundance genes to be detected and differential expression determined.

**Image acquisition**

Nylon membrane cDNA microarrays hybridised with a radioactive $^{33}$P-labelled probe are scanned with a phosphorimager screen. Commercially available software such as Imagene (BioDiscovery, CA, USA) is then used to align the specific grid of arrayed DNA spots and quantify the signal intensity at each location. Scanners and software used for fluorescent image acquisition from glass cDNA microarrays are more complex, and there are many different systems available (see http://www.biocompare.com). The Axon GenePix 4000B scanner has two lasers which simultaneously excite a small region of the glass surface ($\sim 100 \mu m^2$), at a focal plane pre-set by the user ($-50 \mu m$ to $+200 \mu m$ relative to the slide). The entire image is obtained by moving the laser lens across the glass slide. Light emitted at the wavelengths of the fluorescent labels (532 nm for Cy3 and 635 nm for Cy5) is converted to an electrical signal with a photomultiplier tube (PMT). These signals are then displayed as a 16-bit tagged image file format (TIFF) image (Cy5 coloured red and Cy3 coloured green) and given numerical values. Scanning and processing images from glass microarrays requires the investigator to perform a number of manual tasks, each of which demands a high level of technical knowledge. The PMT of the GenePix scanner needs to be adjusted to maximise its dynamic range (0 to 65,536 pixels), to prevent signal saturation and balance the intensities of the two excitation wavelengths (Forster et al. 2003). Both red and green foreground and background intensities are measured for each spot. The foreground intensity for the spot is given as the mean intensity of all foreground pixels and this is assumed to be proportional to the number of complementary mRNA molecules present in the sample. The method of background calculation differs between software packages, but local background correction is preferred over global methods (Kim et al. 2002). The default method used by GenePix software is local background subtraction, in which a different background value is computed for each individual spot and the median background pixel intensity is used for correction purposes.

Following image acquisition, the user must align an appropriate grid containing spot identities to the image, as well as identify artefacts of the hybridisation process so that they can be removed from subsequent analyses. As the settings used for background calculation, background thresholds and ratio calculation can greatly influence data quality, the investigator should be aware of the implications of using each of the different methods. Importantly, the methods used for image acquisition can be optimised from slide to slide, but those used for image quantification should be identical for all slides in the experiment (Forster et al. 2003).

**Data analysis**

Following image acquisition and conversion of the image into spot intensity and intensity ratio measures, this large body of data must be stored in spreadsheet form for further analysis. Depending on the number of microarrays processed, a high capacity system such as an SQL server may be required. Data analysis is perhaps the most critical aspect of a microarray experiment and the least understood by the majority of biologists. Many different analytical approaches have been developed to achieve sensitivity in detecting gene changes while also providing a measure of statistical significance and likelihood of error. Measures of statistical significance can be made for a single microarray, or across multiple biological and technical replicate microarrays. The complexity of the data and statistical analysis requires the use of sophisticated visualisation and analysis software.

Graphical displays are useful in determining the overall success of a microarray experiment (Smyth et al. 2003). The red-green image produced during scanning can detect any problems with colour balance, hybridisation, spatial effects, spot quality or artefacts such as scratches and dust. The original red-green image can also be used to check differential expression of a particular gene. Before they are plotted or analysed, the raw intensity data are always log-transformed ($\log_2$) to spread the values more evenly across the scale from 0 to 65,535 pixels. If any negative values for red (R) or green (G) foreground intensity have arisen due to high spot background, these will be removed from the analysis on a log scale. Using these log-transformed values, an informative visualisation tool is the MA-plot (Dudoit et al. 2002b). This scatterplot has M-values (R/G ratio log-transformed to $M = \log_2 R/G$) on the vertical axis and A-values (spot intensity expressed as $A = \log_2 (R \times G)$) on the horizontal axis. Particularly when using large microarrays with thousands of probes, the majority of genes should not be differentially expressed. An MA-plot of good quality microarray data should therefore have an elongated comet shape centred around $M = 0$ (i.e. equal red and green intensities over a wide intensity range). As well as helping to identify spot artefacts and intensity-dependent patterns, MA-plots can be used to display the effects of normalisation on the data (Smyth et al. 2003).

**Normalisation**

Normalisation is essential in microarray experiments to adjust the data for systematic non-biological effects arising from technical variation and measurement error (see ‘Experimental design’ above). The aim of normalisation is to remove the effect of this ‘noise’ from the data, while...
still maintaining the ability to detect significantly differentially expressed genes. When using multiple nylon membrane or other single sample microarrays, each of the arrays must be paired with another and normalised or ‘scaled’ to its pair (reviewed in Evans et al. 2003). Dual colour fluorescence microarrays require normalisation to account for differences between microarrays, print-tips groups and fluorescent dye channels (reviewed in Smyth & Speed 2003). There is no universally accepted method of microarray data normalisation, and a description and comparison of all available methods is beyond the scope of this review. Overall, the literature supports the use of intensity-dependent normalisation methods, such as print-tip loess (local weighted regression) normalisation (Dudoit et al. 2002b, Yang et al. 2002, Park et al. 2003, Smyth & Speed 2003). This method is capable of removing biases without altering the structure of the data. Essentially, print-tip loess normalisation corrects the M-values (log₂R/G ratios) for non-biological spatial and intensity effects.

**Statistical analysis**

Clustering was one of the first methods used to impose order on microarray data (Eisen et al. 1998). This method involves grouping genes on the basis of similar expression patterns, with the assumption that each cluster of genes is co-ordinately regulated, perhaps as part of the same signaling pathway. Clustering can be useful in assigning potential functions to unidentified genes and ESTs, which can then be tested in further studies. Related methods such as supervised clustering, principle component analysis, self-organising maps and linear discriminant analysis are also widely used to discover patterns of gene expression common to a particular physiological state.

The aim of a microarray experiment is usually to identify differentially expressed genes, with a measure of statistical significance (reviewed in Dudoit et al. 2002b; Cui & Churchill 2003). Most microarray experiments are designed with only one categorical factor (e.g. treatment or genotype), so the statistical analysis is based on the paired t-test. Experiments with multiple categorical factors (e.g. genotype and time) require methods based on the analysis of variance (ANOVA). Once the data are appropriately normalised, it is common practice to consider a univariate testing problem for each gene and calculate t-statistics (Dudoit et al. 2002b). The t-statistic tests the null hypothesis of equal mean expression levels in the two samples (e.g. treatment and control). Another useful indicator of differential expression is the B-statistic (Lonnstedt & Speed 2002), which is an estimate of the odds that the gene is differentially expressed. The challenge in assigning statistical significance to a differentially expressed gene is that the often thousands of genes on a microarray result in a high level of multiple testing. Determining the false discovery rate is the most powerful method of controlling for multiple testing (Tusher et al. 2001), but this can also be achieved using adjusted P values (Dudoit et al. 2002b).

Time course experiments require even more specialised statistical analysis (Cui & Churchill 2003) and should only be conducted if the primary biological question is one of time dependence.

Just as diagnostic MA-plots can be invaluable for visualising trends in raw and normalised data, plots of values obtained during statistical analysis are also useful. Both fold change (difference in gene abundance between two samples) and significance measures can be represented graphically in a ‘volcano plot’ (Cui & Churchill 2003), with the log odds of differential expression on the vertical axis and the mean M-value (log₂R/G ratio) on the horizontal axis. Genes with statistically significant differential expression will appear above a horizontal threshold line and those with large fold changes (up- or downregulated) will lie to the far left or right. Differentially expressed genes identified by the B-statistic will appear in the upper left or right quadrants.

There are many different software packages available for performing normalisation, statistical analysis and visualisation with single and dual sample microarrays. Some of the more widely used packages include Cyber-T (Baldi & Long 2001), SAM (Tusher et al. 2001), BRB-ArrayTools (http://linus.nci.nih.gov/BRB-ArrayTools), QVALUE (Storey & Tibshirani 2003) and Focus (Cole et al. 2003). The statistical language R (Ihaka & Gentleman 1996, http://www.r-project.org) has also been used successfully for the analysis of microarray data (Dudoit et al. 2002a) and indeed many of the other packages are based on R commands. Bioconductor (http://www.bioconductor.org) provides a more user-friendly interface for the R statistical language. Although they are often easier for biologists to use, care should be exercised in the choice of commercially available software packages. Some are excellent for data visualisation and normalisation, but cannot assign measures of statistical significance within and across multiple microarrays, or do not handle time course data.

**Data validation**

The biomedical research community does not yet accept that microarray data can stand alone, without independent validation (reviewed in Rockett & Hellmann 2004). There are a number of reasons for this caution, including the relatively recent development of the technology, the lack of standard operating procedures and the potential for errors to exist in the data (Knight 2001, Kothapalli et al. 2002). Hybridisation errors may occur due to cross-hybridisation between transcripts of high homology and data may be misleading if mis-annotation of probe sequences has occurred. Even in well-maintained clone sets, it is estimated that 1–5% of clones do not contain the correct sequence (Knight 2001). In addition, as the statistical tests used for microarray data analysis are yet to be standardised, often several methods are used and the resulting data requires further validation. Best practice in microarray analysis can achieve less than 5–10%
variation in signal intensity from replicate probes on the same microarray, and around 10–30% variation between corresponding probe signal intensities on different microarrays (reviewed in Stears et al. 2003). Despite this, there is an expectation that an additional mRNA quantification method will be used to confirm the differential expression of the genes of interest (Firestein & Pisetsky 2002).

The first important task for the investigator is to decide which genes to investigate further. From experience, genes displaying a large fold change (>2) and statistical significance are the best candidates for validation. Before embarking on additional studies, it is good practice to review the primary red-green image data to confirm differential expression of these genes, and the spotted DNA sequence may also be checked for correct annotation. Comparing data with that obtained from other microarray studies on the same system can also provide ‘in silico’ validation and increase confidence in the data set as a whole (Chuaqui et al. 2002).

Quantitative real-time RT-PCR is commonly used to confirm mRNA levels, as it has higher sensitivity and lower RNA requirements than Northern blot. Previous studies have demonstrated that genes with relatively high expression and at least 2-fold regulation are likely to be validated using real-time RT-PCR (Rajeevan et al. 2001). The advantage of Northern blot and RNase protection assay is that they provide a quantitative measure as well as reveal the number and size of transcripts detected by the particular spotted DNA sequence. Quantitative data obtained with microarray and Northern blot are comparable, while Northern blot slightly more sensitive in detecting differential expression compared with microarray (Taniguchi et al. 2001). In complex tissues such as the endometrium, defining the cellular localisation of mRNA expression using in situ hybridisation can provide important functional information. As it is almost impossible to differentiate between primary and secondary gene expression effects in microarray data, further testing may be required to define the molecular interactions occurring.

While mRNA reflects the functional state of the cell, it is the proteins which ultimately carry out the instructions of the genome. Translation of mRNA into protein may be controlled independently of transcription and proteins may undergo post-translational modifications that alter their function. To describe a biological event or system, therefore, gene expression data obtained by microarray analysis must be extended to the study of protein products. Particularly if target RNA has been prepared from whole tissue, characterising the cellular distribution of the corresponding protein by immunostaining or tissue array is critical to understanding the function of a gene. Protein quantification by Western blot or ELISA will indicate whether transcription and translation are co-ordinately regulated.

Defining the functions of differentially expressed genes may be considered the ultimate validation of microarray data. Functional studies may include in vitro experiments using dominant-negative mutants or RNA interference, or in vivo experiments using antisense morpholino oligonucleotides, knockout or conditional knockout technologies. Though the experiments may be carried out some time later, each level of data validation (mRNA, protein and function) should be considered at the microarray experimental design stage, to allow additional controlled samples to be obtained.

Endometrial gene expression analysis

In the last few years, both cDNA and oligonucleotide microarray technology have been successfully applied to the study of endometrial gene expression (reviewed in Giudice 2003). The endometrium is a uniquely dynamic tissue, with the capacity to undergo dramatic remodelling in response to cyclic variations in steroid hormones and local autocrine and paracrine factors. The mRNA complement of each of its different cell types is altered during different phases of the menstrual cycle, with the onset of decidualisation and in response to an implanting embryo, as well as in pathological conditions such as endometriosis, abnormal bleeding, infection or cancer. Gene expression profiling has the capacity to identify new targets for the manipulation of fertility and the diagnosis and treatment of endometrial abnormalities.

A number of endometrial gene expression studies have been discussed in recent reviews (Giudice 2003, 2004, Horcajadas et al. 2004), so rather than providing a detailed description of their findings, the experimental design features of these studies have been summarised in Table 2. With only three exceptions (Popovici et al. 2000, Martin et al. 2002, Okada et al. 2003), all of these studies included validation of a small number of differentially expressed genes (usually less than 10) by an independent mRNA quantification method (Northern blot, semi-quantitative or quantitative RT-PCR). Less than half also included cellular localisation studies (in situ hybridisation and/or immunohistochemistry).

The molecular events of endometrial stromal cell decidualisation are still not well defined, so microarray analysis is ideally suited to identify genes with important regulatory roles in this process. A number of factors including combined oestrogen and progesterone or cAMP can be used to induce decidualisation in vitro and the comparison of decidualised with non-decidualised cells has revealed many of the genes which are likely to contribute to this process (Popovici et al. 2000, Brar et al. 2001, Tierney et al. 2003). The two studies investigating the window of implantation (Kao et al. 2002, Borthwick et al. 2003) were identical in experimental design except that the microarrays used in the first study were hybridised with RNA samples from individual endometrial biopsies, whereas the second study used pooled samples. It is important to note that there is no consensus on the use of pooled or individual samples...
(see ‘Experimental design’ above). Many of the differentially expressed genes identified by Borthwick et al. (2003) were the same as those reported by Kao et al. (2002), suggesting that different experimental designs may be equally valid. Similarly, there was some consensus between the study by Kao et al. (2002) and that of Carson et al. (2002) examining the transition of the endometrium into receptivity. The study by Riesewijk et al. (2003) was slightly different from that of Carson et al. (2002) in that consecutive endometrial biopsies were taken from the same patient (rather than from different patients) and the timing of the biopsies was more precise. There was good agreement in the data obtained by Riesewijk et al. (2003) and Kao et al. (2002), but less between the two more similar studies (Carson et al. 2002, Riesewijk et al. 2003). Discrepancies in data sets are not unexpected, considering the many sources of variation in microarray experiments, and the differences in methodology used by different laboratories. Any genes which show consistent differential expression under similar experimental conditions can therefore be considered almost certain to play an important biological role.

Genes involved in endometrial receptivity and implantation have also been examined using the progesterone receptor antagonist RU486 (Cheon et al. 2002, Catalano et al. 2003, Tynan et al. 2005). As RU486 is known to inhibit implantation in mice and humans, its downstream target genes are likely to be involved in normal implantation, and have been identified in the whole mouse uterus (Cheon et al. 2002), in human endometrial explants (Catalano et al. 2003) and in cynomolgus monkey endometrial biopsies (Tynan et al. 2005). While no genes were found to be regulated by RU486 in all

Table 2 Summary of studies using DNA microarray analysis to investigate gene expression in the endometrium during normal processes or in response to stimuli

<table>
<thead>
<tr>
<th>Species</th>
<th>Cell/tissue type</th>
<th>Process/stimulus</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Window of implantation vs late proliferative phase</td>
<td></td>
<td></td>
<td>Kao et al. 2002, Borthwick et al. 2003</td>
</tr>
<tr>
<td>Normal vs IVF cycles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laser capture microdissection</td>
<td></td>
<td></td>
<td>Matsuzaki et al. 2004, Yanaihara et al. 2005</td>
</tr>
<tr>
<td>Endometrial explants</td>
<td>Epithelial vs stromal cells</td>
<td></td>
<td>Catalano et al. 2003</td>
</tr>
<tr>
<td>Whole uterus</td>
<td>Across normal oestrous cycle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-vs post-implantation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implantation vs interimplantation sites</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oestrogen</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endometriosis</td>
<td>Hoxa-10 deficiency</td>
<td></td>
<td>Yao et al. 2003, Poggi et al. 2003</td>
</tr>
<tr>
<td>Rhesus monkey</td>
<td>Whole uterus</td>
<td>Ovariectomy and oestrogen treatment</td>
<td>Naciff et al. 2002, Wu et al. 2003</td>
</tr>
<tr>
<td>Cynomolgus monkey</td>
<td>Endometrial biopsies</td>
<td>Pre-receptive vs receptive</td>
<td>Ace &amp; Okulicz, 2004</td>
</tr>
<tr>
<td>Cow</td>
<td>Endometrial biopsies</td>
<td>Progesterin antagonist (RU486)</td>
<td>Tynan et al. 2005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pregnant vs non-pregnant</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
three species, some were identified as downregulated with both RU486 treatment (Cheon et al. 2002) and in post-implantation compared with pre-implantation mouse uterus (Yoshioka et al. 2000). Reese et al. (2001) examined genes involved in mouse implantation using a combined approach of implantation versus interimplantation sites and activated versus delayed implantation. Interestingly, many of the genes regulated in both models were associated with the maternal immune response. Genes found to be regulated in both mouse and human endometrium with the onset of decidualisation and/or receptivity are attractive targets for the manipulation of implantation mechanisms conserved across species.

Microarrays have also been utilised to identify potential markers of endometrial pathologies. Studies exploring differential gene expression in endometriotic lesions versus eutopic endometrium (Eyster et al. 2002, Lebovic et al. 2002, Arimoto et al. 2003) have revealed dysregulation of a number of genes in endometriotic tissue (reviewed in Giudice 2003), which may prove to have functional roles in this disease. Interestingly, a comparison of gene expression in eutopic endometrium from women with and without endometriosis (Kao et al. 2003) has shown that the endometrium of women with endometriosis has an altered transcriptional profile to that of women without the disease. As endometriosis is often associated with infertility, genes with altered expression in endometriosis patients may be involved in endometrial receptivity and embryo implantation (reviewed in Giudice et al. 2002). Similarly, genes found by microarray to be differentially expressed in endometrial tumours compared with normal endometrium (Mutter et al. 2001, Saidi et al. 2004) are likely to provide diagnostic markers and treatment targets in the future (reviewed in Giudice 2003). Another powerful application of microarray technology is the classification of tumour types by their gene expression profiles and a number of studies have successfully utilised this approach in endometrial cancer (Moreno-Bueno et al. 2003, Risinger et al. 2003, Cao et al. 2004, Ferguson et al. 2004, 2005). Molecular classification of tumours using microarray technology has the potential to greatly enhance patient management and improve treatment and prognosis.

Conclusions

Conducting a microarray experiment requires extensive planning and a high level of technical ability. Each of the steps involved, from experimental design through to data analysis, requires careful consideration and greatly benefits from collaboration with a bioinformatician. As the volume of microarray data continues to expand, there is an increasing need for accepted standard operating procedures and coordinated data deposition. The Microarray Gene Expression Data (MGED) Society exists to encourage MIAME (Minimum Information About a Microarray Experiment) compliance, so that microarray data from different laboratories and platforms can be unambiguously interpreted and independently verified. Indeed publication in a number of leading journals, including Nature, is now dependent upon meeting MIAME standards (Rockett & Hellmann 2004). Standardisation of the field could also be achieved by including a Standard Gene Set (SGS) on all microarray platforms (Fryer et al. 2002). Publicly-available databases are also required to enable large-scale data mining. For example, mRNA levels measured in many human tissues using the Affymetrix system are now available online (HuGE Index; http://www.hugeindex.org) and an endometrial database containing gene expression data from many of the studies mentioned above can be accessed at http://endometrium.bcm.tmc.edu/edr.

Despite many recent advances, microarray analysis should not be considered the end-point of an investigation, but rather as a tool to assist in the formulation of hypotheses. With improved microarray quality, standardised data analysis methods and integration with proteomic approaches, gene expression profiling will be an extremely effective tool towards understanding the biology of the reproductive system and in developing diagnostic tests and therapeutic strategies for reproductive abnormalities.

Acknowledgements

The authors would like to thank Dr Garry Myers, Dr Gordon Smyth, Prof Terry Speed and Dr Andrew Sharkey for training and helpful discussions and Sue Panckridge for assistance in the preparation of Figure 1. LAS is supported by the National Health and Medical Research Council of Australia (grants #241000 and #143798) and CAW by an Australian Postgraduate Award. The authors declare that there is no conflict of interest that would prejudice the impartiality of this scientific work.

References

Ace CI & Okulicz WC. 2004 Microarray profiling of progesterone-regulated endometrial genes during the rhesus monkey secretory phase. Reproductive Biology & Endocrinology 2 54.


Received 10 February 2005
First decision 12 April 2005
Revised manuscript received 27 April 2005
Accepted 3 May 2005