REVIEW ARTICLE

Microarrays: new tools to unravel parasite transcriptomes

G. N. GOBERT[1,*], L. P. MOERTEL[1,2] *and* D. P. McMANUS[1]

[1] *Molecular Parasitology Laboratory, Australian Centre for International and Tropical Health and Nutrition, Queensland Institute of Medical Research, 300 Herston Road, Brisbane, QLD 4029 Australia*
[2] *The School of Chemical and Biomedical Sciences, Central Queensland University, Bruce Hwy, North Rockhampton, QLD 4702 Australia*

SUMMARY

The ability to monitor the expression levels of thousands of genes in a single microarray experiment is a huge progression from conventional Northern blot analysis or PCR-based techniques. Microarrays can play a pivotal role in the mass screening of genes in a wide range of fields including parasitology. The relatively few parasites that can be readily cultured or isolated from a host, as compared with cell lines or tissue sources, makes microarray technology ideal for maximizing experimental results from a limiting source of starting material. Khan *et al.* (1999*a*) commented in an early review of microarray technology "*With this system in place, one can anticipate a time when data from thousands of gene expression experiments will be available for meta-analysis…….. leading to more robust results and subtle conclusions*". Now in 2005, microarrays represent a very powerful resource that can play an important role in the characterization and annotation of the transcriptomes of many parasites of medical and veterinary importance.

Key words: microarray, gene expression, genomic sequencing, transcriptome.

INTRODUCTION

Microarrays are specially produced slides that have thousands of individual DNA probes attached in an ordered array to the surface. They provide the user with the ability to monitor the expression level of thousands of genes simultaneously. In 1995, Schena and co-workers reported the first complementary DNA microarray analytical procedure using 45 genes from the plant *Arabidopsis* (Schena *et al.* 1995). Since then, this technology has greatly expanded the applications of genomic research. In addition, entry-level array probe printing machines have made the production of chips less expensive (Khan *et al.* 1999*b*). Microarray technology is a new discipline that has generated its own terminology and acronyms. Some of the key terminology is presented in Table 1.

MICROARRAY PRODUCTION AND EXPERIMENTAL CONDITIONS

There are multiple methods in which microarrays may be manufactured, but all share similar characteristics. (1) Complementary DNA (cDNA) array chips. The cDNA probe is transferred to a glass slide by an array-printing machine and stored until use (see Clontech 2001). (2) Simple oligonucleotide spotting, in which the manufacture of oligos is performed separately and then chips are facilitated by simple array printing machines; this makes this method inexpensive compared to others (Street, 2002). (3) Photolithography is used to produce high density microarray chips, in a similar way to the production of computer chips (Hughes *et al.* 2001). Lithographic masks are required to control the exposure of light for each round of oligonucleotide synthesis. High costs are usually associated with this method especially if a custom designed array is utilized. (4) Ink-jet chips, which like photolithographic chips, are high in density. This method utilizes a robotic spotting method to deposit individual nucleotides onto the specially prepared surface, one layer at a time, building up an oligonucleotide probe (Hughes *et al.* 2001).

The longer DNA lengths of cDNA arrays enhance the specificity of hybridization, but require significantly more time and associated costs of the pre-printing setup. Oligonucleotides have less printing set up costs, with acceptable specificity (Hughes *et al.* 2001; Summan *et al.* 2003), but require an additional design process that can guarantee appropriate thermodynamic properties for hybridization and to ensure no cross-hybridizations occur. There are a number of software systems that can help in the process including OligoWiz (Nielsen, Wernersson and Knudsen, 2003) or Array Designer (http://

Table 1. Key terminology used in microarray design, construction, utilization and analysis

(Adapted from Rosetta BioSoftWare: http://www.rosettabio.com/tech/geml/omg/lsr_ge_glossary.doc.)

Array	Array refers to the physical substrate to which DNA probes are attached to create features. In gene expression profiling experiments, arrays are hybridized with labelled sample and then scanned and analysed to generate data.
Array Design	The layout or blueprint of one or more arrays. An Array Design is conceptual and an Array is a physical object.
Background	Background is the measured signal outside of a feature on an array. In many gene expression analysis methods, background subtraction is performed to correct measured signals for observed local and/or global background.
Channel	A channel is an intensity-based portion of expressed data that consists of the set of signal measurements across all features on an array for a particular labelled preparation used in hybridization. In some cases, such as Cy3/Cy5 array hybridizations, multiple channels (one for each label used) may be combined in a single expression profile to create ratios.
Chip	The physical medium of many arrays used in gene expression.
Control	The reference for comparison when determining the effect of some procedure or treatment. (Deletion, mismatch, positive, negative).
Error Model	An error model is an algorithm that computes quality statistics such as P-values and error bars for each gene expression measurement. Error models are typically specific to a particular expression profiling technology.
Experiment	An experiment studies some observable system under controlled conditions while some variable is manipulated in hopes of understanding the effects of the variable on the system.
Expression	The conversion of the genetic instructions present in a DNA sequence into a unit of biological function in a living cell. Typically involves the process of transcription of a DNA sequence into an RNA sequence followed by translation of the RNA into protein.
Extract	Preparation taken from a biological source such as a tissue or cultured cells. It could consist of purified protein, RNA, DNA, or other cellular material. For microarray gene expression experiments, mRNA or total RNA preparations are labelled and then hybridized to arrays.
Feature	A feature refers to a hybridized probe. Commonly referred to as a spot in a microarray experiment.
Feature Extraction	Quantitative analysis of an array image or scan to measure the expression values.
Fluor	Also, fluorophore. Short for fluorescent label. Such a tag is bound to mRNA or cDNA extracted from a sample. When properly excited the fluor produces measurable fluorescence which is the observable in an experiment.
Hybridization	Treating an array with one or more labelled preparations under a specified set of conditions.
Label	Label refers to a fluorescent label, for example Cy3 and Cy5, commonly used to distinguish baseline and experimental preparations in gene expression microarray hybridizations.
Normalization	Mean signal intensity can vary dramatically among expression profiles or channels. Normalization is the procedure by which signal intensities from two ore more expression profiles (or channels) are made directly comparable through application of an appropriate algorithm.
Oligo/Oligonucleotide	Usually short strings of DNA or RNA to be used as probes [features] or spots. These short stretches of sequence are often chemically synthesized.
Probe	In some organizations, a 'probe' is used as a synonym for Feature (see Feature above).
Ratio	Also referred to as 'fold change'. A ratio refers to a normalized signal intensity generated in a Feature given channel divided by a normalized signal intensity generated by the same Feature in another Channel. The channels compared are typically baseline versus experimental, e.g. normal versus diseased or untreated *vs* treated.
Replicate Experiments	A replicate set refers to repeated experiments where the same type of array is used, and the same probe isolation method is used. The intent is to arrive at separately obtained results that can be compared with one or more other sets of observations to arrive at a consensus or more statistically meaningful interpretation of results.
Target	Material that may hybridize to the probe, usually containing all of the mRNA (cDNA or cRNA) or gDNA (genomic) of the subject organism.
Treatment	A treatment is the experimental manipulation of a sample such as a cell culture, tissue, or organism prior to extraction of a preparation.

www.premierbiosoft.com/dnamicroarray/) for those wanting to design oligo-probes for a micro-array. Various considerations are required in the oligonucleotide probe design process, including melting temperature, GC content, self-annealing and secondary structure potential, all of which are discussed in a paper by Tolstrup *et al*. (2003), who also described another design program, OligoDesign.

The difficulties of designing cDNA probes is addressed in a paper by Chen *et al.* (2004*a*), who made special reference to the problems associated with incomplete genomes, a common limitation with parasites. The authors developed a 'sequence diversity index, SDI' which monitored the diversity that was present between and within dynamic clustering. This method was under-scored with broad Gene Ontology (GO) allocations that allowed cross-hybridization to be addressed in relation to varying degrees of clustering. The ability to design microarray probes without the need to cluster ESTs enables the construction of multi-species or cross-species microarrays, an important application in parasitology.

MIAME

The need for a standard of Minimum Information About a Microarray Experiment (MIAME) was first highlighted during a meeting organized by the European Bioinformatics Institute in 1999. After development and discussion, MIAME was proposed as standard practice and presented in *Nature Genetics* in 2001 (Brazma *et al.* 2001). MIAME is a detailed list of information that describes the experiment from construction of the chip to data analysis.

MIAME is made up of 2 major sections:

1. Array design description

    i. Array related information, including design name, platform type, and number of features.
    ii. Information about the probes, sequence, type, attachment, location on array and controls.

2. Gene expression experiment description

    i. Experimental design, including authors, type of experiment, experimental factors (time dose), and quality controls.
    ii. Samples used, extract preparation and labeling, sex, developmental stage, type, biomaterial manipulations, protocol, conditions, treatments, hybridization extract preparation protocol, external controls.
    iii. Hybridization procedures and parameters, batch serial numbers, blocking agent, wash procedure, quantity of target.

(Modified from MIAME; http://www.mged.org/Workgroups/MIAME/miame.html).

Most of these experimental conditions represent good general research practice. However, the consensus and formatting presented in MIAME makes it much easier to access information, adding to the design and utilization of larger databases that correlate with individual microarray datasets.

GENOMICS, EXPRESSED SEQUENCE TAGS AND MICROARRAY PRODUCTION

Advances in genomics and particularly in sequencing methods (see Knox, 2004) have enabled the establishment of many sequencing projects focusing on a range of parasites of medical and veterinary importance. This wealth of new data provides the basis for the design and construction of microarrays from a wide range of parasite taxa.

The most high-profile sequencing program in parasitology to date is that involving *Plasmodium* (Carucci *et al.* 1998; Wilson, 2004). Huge insights have been gleaned from numerous EST and genomic sequencing projects, and the subsequent microarray analysis has followed to help better understand the biology and pathogenesis of malaria. Another prominent parasite group subject to large-scale sequencing efforts are the filarial nematodes, and these studies have led to major advances in our understanding of the genome of *Brugia malayi* (Williams *et al.* 2000), particularly in the area of chromosome mapping and functional genomics (Blaxter *et al.* 2002; Foster *et al.* 2004).

One of the priorities of the World Health Organization has been to address the major problem of human schistosomiasis through the formation of the Schistosoma Genome Network (see http://www.nhm.ac.uk/hosted_sites/schisto/index.html). This program consists of a number of laboratories that aim to use expressed sequence tags (EST), and genomic sequencing strategies to identify novel genes. These sequencing efforts have culminated in the release of large amounts of EST sequence for *Schistosoma japonicum* (Hu *et al.* 2003) and *Schistosoma mansoni* (Verjovski-Almeida *et al.* 2003).

A summary of major sequence projects involving parasites is presented in Table 2. All of the databases listed are available in the public domain databases and, as such, are accessible for future clustering, probe design and microarray construction. The power of *in silico* design and automated oligonucleotide synthesis and spotting makes the transition of raw sequence to a laboratory tool a quick, though relatively expensive proposition.

APPLICATIONS OF MICROARRAYS

Some of the uses of microarrays for the study of parasite transcriptomes are detailed in Table 3, some specific examples of which follow.

(1) Comparison of gene expression during the parasite life-cycle

A very good example of the use of microarrays in parasitology has been the monitoring of the transformation of *Trypanosoma cruzi* from the trypomastigote to the amastigote (Minning *et al.* 2003).

Table 2. Summary of ongoing parasite sequencing projects

Major sequencing efforts focusing on

- *Plasmodium falciparum* (see Verma and Sharma, 2003; Carlton *et al.* 2005) with programs at
    - TIGR (http://www.tigr.org/tdb/edb/pfdb/pfdb. html),
    - Stanford (http://swquence-www.stanford.edu/group/ malaria/index.html)
    - NCBI (http://www.ncbi.nlm.nih.gov/projects/ Malaria/)

- *Schistosoma sp.* sequencing and analysis repositories
    - *S. japonicum* http://schistosoma.chgc.sh.cn/
    - *S. mansoni*
        - http://bioinfo.iq.usp.br/schisto/
        - http://www.tigr.org/tigr-scripts/tgi/ T_index.cgi?species-s_mansoni

Other prominent projects sequencing important parasites include,
- *Cryptosporidium hominis* (Xu *et al.* 2004)
- *Cryptosporidium parvum* including both
    - EST (http://www.ebi.ac.uk/parasites/cparv.html)
    - Genome (http://www.parvum.mic.vcu.edu/) projects

- Filarial worm projects (http://helios.bto.ed.ac.uk/ mbx/fgn/filgen.html)

- *Giardia lamblia* Informatics Support and Gene Expression Projects
    - http://gmod.mbl.edu/perl/site/giardia? Page=intro

- *Leishmania sp.* Genome project at the
    - Sanger Institute (http://www.sanger.ac.uk/Projects/ L_major/)

- *Trypanosoma cruzi*
    - Genomic (http://www.tigr.org/tdb/mdb/tcdb/)
    - EST (http://www.genpat.uu.se/tryp/tryp.html),

- *Trypanosoma brucei*: genome initiative promoted and partly sponsored by the World Health Organization
    - The Trypanosoma brucei Genome Network http:// parsun1.path.cam.ac.uk/
    - Sanger Institute (http://www.sanger.ac.uk/Projects/ T_brucei/)
    - TIGR (http://www.tigr.org/tdb/mdb/tbdb/index. html)

- *Toxoplasma gondii* (Kissinger *et al.* 2003)
    - http://ToxoDB.org

Table 3. Some applications of microarrays in parasitology

- Life-cycle development. To identify genes that are needed for developmental events in response to environmental changes, deriving nutrition from the host, avoidance of the host immune response.
- In dioecious parasites, differences between male and female: to identify sex-specific expression of genes involved in reproduction.
- Parasites recovered from different experimental and naturally infected hosts: to identify host-specific adaptations to nutritional sources and immunological responses.
- Drug resistant and susceptible strains of parasites: to investigate mechanisms of drug resistance
- Challenge by drug or vaccine. To identify genes involved in damage compensation/healing or avoidance/neutralization of the insult.

This study utilized 4400 probes principally of partly sequenced genomic material and some material from an ORF library that was made from 200–400 bp size DNA. The library used a GFP (Green Fluorescent Protein) expression-based bacterial system for selection. The *T. cruzi* DNA fragments were inserted upstream from the GFP gene lacking a start codon. Thus, clones expressing GFP were selected for use on the microarray. Most of the differential expression was due to the up-regulation of 60 genes in the developing amastigote. Of these, 14 were found to have been characterized previously, 25 were novel, with the remainder being redundant. A subset of genes was validated using real-time PCR that correlated well with the microarray results. The authors aimed to find vaccine targets and identify amastigote-specific genes using this approach. While a relatively large amount of novel genes were identified, the nature of the microarray construction hampered some of the goals of the project. Probe material on the chip could have been more effective if amastigote cDNA had been utilized in an EST strategy study, to enable the maximal chance of finding stage-specific gene expression. No doubt limitations in sample size and the amount of mRNA that can be isolated from *T. cruzi* were a considerable problem, but since this microarray was probably produced in 2002, if not earlier, recently improved cDNA library construction kits may allow this research to be further developed. This in turn may produce a significantly improved microarray. Technological advances in RNA processing, including amplification kits (such as Ambion MessageAmp™ aRNA Kit) that can start with as little as 10 ng of mRNA and library construction kits (such as Invitrogen CloneMiner) starting with only 1–5 μg of mRNA. In addition, improved selection markers,result in larger insert sizes and more rapid library construction, from small amounts of material, a common limitation when working with parasites.

The strategy of using more specific cDNA probes was utilized in a later study by Baptista *et al.* (2004). This group constructed a relatively small cDNA microarray (665 genes; 730 sequences) to examine gene expression and genomic organization in different isolates of *T. cruzi*. Firstly a comparison of the genome between strains was carried out using genomic DNA that detected differential hybridization in 68 genes, of which a subset was validated by Southern blot analysis. The Southern analysis

revealed variations in gene copy number and sequence differences that contributed to the hybridization differences evident in the microarray experiments. These researchers then tested cDNA for hybridizations that yielded information on gene expression between the two strains. They found that 84 probes were differentially expressed, of which a subset was validated by Northern blot analysis. Only 20% of the genes overall had variations in both expression profile and genomic copy number (and/or sequence variation), when these two groups of differentially hybridized genes were compared. This indicated that there was no correlation between the abundance of a particular gene within the genome and the expression of that gene as RNA, demonstrating a high degree of gene expression control and regulation.

(2) *Comparative gene expression between related species and evolutionary considerations*

There have been few investigations of inter-specific genomic variation using microarray technology generally, and, specifically, in parasitology. There are significant limitations in using a microarray platform with limited sequence homology or incomplete sequence, that result in low efficiency hybridizations.

The limited physical length of sequence available in short cDNA or oligonucleotide probe microarrays requires a high degree of homology with the RNA that is probed in order to obtain a hybridization signal. A low level of homology will result in a limited hybridization, which may be misinterpreted as a low gene expression level. However, in either situation, differential expression on the microarray will highlight the probability of a variation in sequence identify or gene expression, which can be further investigated by sequencing efforts. Two examples of successful interspecies analysis include cDNA-based microarrays from fish – *Astatotilapia* (Renn, Aubin-Horth and Hofmann, 2004) and *Salmo* (Rise *et al.* 2004). Gene expression between the different fish species was, as expected, more successful in the closer-related species, than with others further diverged. In both studies the principle of using a microarray platform to examine a wide range of species and present evolutionary and ecologically relevant data were demonstrated. Other microarray studies have examined different species of *Drosophila* (Watanbe *et al.* 2004) and interspecies variations in yeast (Gu *et al.* 2004). The large-scale examination of genes can identify small variations that occur between closely related species or strains, making microarray analysis an excellent method for taxonomic classification. The conclusions from these inter-species microarray studies can readily be adapted to parasitology. While, as discussed previously, a number of parasites are currently subject

to major genome sequencing projects, a large number of species will not have such thorough examination, at least in the medium term. Microarray analysis between species may provide useful insights into evolutionary traits. A strategy to achieve this would be to use using data reduction techniques such as principle component analysis (Randall *et al.* 2003) or multidimensional scaling (Dugas *et al.* 2004). By simplifying the microarray data, broad relationships between species could then be demonstrated. As previously mentioned, extensive differential expression may either identify genes that have diverged between species or are even absent in one of the examined species. Such divergence or total absence of genes may reflect key evolutionary differences between species. The presentation of unique genes within a species may be indicative of relatively newer evolutionary adaptations to environmental pressures.

In regard to the malaria parasites, information regarding life-cycle and inter-species studies are presented in the informative database 'PlasmoDB – The *Plasmodium* Genome Resource' [http://plasmodb.org/] (see Bahl *et al.* 2003 for review). In addition to presenting *Plasmodium* proteomics, primary sequence, and both cDNA and oligonucleotide microarray data, the web site also provides inter-correlations tools, such as plotting taxonomic divergence against sequence similarity, giving a representation of inter-species variation. Presently, the microarray data contained within the site is from *P. falciparum* and there are no reports of cross-species/strain hybridizations. The only comparisons made have been performed *in silico* from proteomic or sequence data. However, the framework and organization provided does present a suitable format for future microarray data that will arise from other species of *Plasmodium* that can be subjected to further analysis.

(2) *Plasmodium microarray studies, leading the way in parasitology*

Two landmark papers in 2003 described the construction of an oligonucleotide microarray for *P. falciparum* (Bozdech *et al.* 2003; Le Roch *et al.* 2003). Bozdech *et al.* (2003) developed software to aid in the design of 70mer oligonucleotide probes to 6000 Open Reading Frames from the parasite genome. The microarray was then used to analyse the gene expression of the trophozoite and schizont stages. As expected, major differences in the transcriptional profiles of these stages were detected. This study represents a good example where public domain sequences could be accessed, a design process followed, and a microarray produced by a non-commercial entity. The work also emphasized the advantage of *in silico* design of probes (as compared to using cDNA clones), so that by controlling

GC content and other parameters, hybridization efficiencies can be enhanced. The work of Le Roch *et al*. (2003) again utilized an oligonucleotide design strategy for microarray construction but presented this on a much larger scale. In this study, a microarray (multiple chips) containing 260 596 features from predicted coding regions (genomic) and 106 630 probes from non-coding regions, was constructed. One of the distinct advantages of using probes designed from genomic sequences, as opposed to ESTs, is that the detection of gene expression is not limited to the life-cycle stages from which the ESTs were originally derived. This was an important consideration in this study since 9 different life-cycle stage of *P. falciparum* were examined using the microarray, and gene expression across the life-cycle was more effectively monitored with genome-derived probes. The authors were able to correlate and cluster genes based on similar expression levels with specific life-cycle stages to group genes with assumed similar functions. A major drawback of this study was the use of 25-mer oligonucleotides which may have had a deleterious effect on the specificity of the probes used in the hybridizations. The use of 60 mers has been reported to provide much higher specificity in hybridization for microarray analysis (Hughes *et al*. 2001; Summan *et al*. 2003).

Recent microarray studies in *P. falciparum* have become more focused. Specific cellular activities including cell cycle elements, overall gene regulation and factors involved in transcriptional control of gene expression, have all be examined in a recent publication by Gissot *et al*. (2004). PCR products ranging in size from 300–1500 bp were printed from 150 genes with homology to known proteins involved in signal transduction or the cell cycle. Differential expression of these prospective candidates was compared between a 'normal' clone of *P. falciparum* and a subclone with a defect in gametocytogenesis. When the two strains were examined during a gametocytogenesis time-course experiment, 114 genes consistently provided clear microarray data, of which 106 genes were identical in expression profile between the two clones. Eight genes were differentially expressed between the two strains during the time-course to differing extent. These experiments identified genes important in the dynamic events leading up to gametocytogenesis in *P. falciparum*.

Host responses to malaria were examined in an animal model, using *P. cynomolgi bastianellii* and rhesus monkeys (Ylostalo *et al*. 2005). A commercial oligonucleotide chip of the human genome (Affymetrix HG-U133A) was used to follow a time-course of infection, with RNA isolated from whole blood, before infection, during the initial liver phase, during peak parasitaemia and at the first/second relapses. A wealth of information was obtained but the most significant findings included a general down-regulation of genes during the initial liver phase of parasite infection, demonstrating a host response to the infection at the transcriptional level. In addition, 'defensive-response' genes were also identified by distinctive up-regulation. Some examples of these genes include α enolase (*ENO1*) which regulates inflammatory responses, and a *NFE2L3* which regulates many erythroid specific genes.

(4) *Investigation of schistosome transcriptomes with microarray technology*

Highly informative microarray studies have been published for schistosomes (Hoffmann, Johnston & Dunne, 2002; Fitzpatrick *et al*. 2004). This group utilized cDNA microarrays consisting of ESTs from adult *S. mansoni* (576 features) and *S. japonicum* (457 features) to identify sex-specific gene expression; sex-associated genes were identified by microarray analysis and confirmed by RT-PCR. In the *S. japonicum* study (Fitzpatrick *et al*. 2004) functional studies (enzymatic assays and localization) were also presented to further demonstrate the gender-specific differences occurring at the transcriptional level. The report also presented some sex-specific gene differences that were evident between geographical isolates (Anhui and Zhejiang) of the Chinese strain of *S. japonicum*. More differential expression may have been detected in comparing geographical strains with more pronounced phenotypic difference such as the Philippine and Chinese strains of *S. japonicum*.

A recent addition to this work includes a report of the construction of a 7335 *S. mansoni* element microarray using a spotted oligonucleotide approach (Fitzpatrick *et al*. 2005). The oligo-probes were designed from 17 329 *S. mansoni* EST sequences and represent approximately 50% of the estimated schistosome transcriptome. Again the focus of the study was to identify sex-biased genes, which resulted in the identification of 141 genes up-regulated in the female adult *S. mansoni* and 86 genes up-regulated in the male. Good correlations with previous data from this group were made, in addition to greatly extending the list of genes for analysis. Although the authors demonstrated strong reproducibility between their array experiments, as reflected in the R coefficient values, the confidence levels (or error modelling) for individual genes in individual microarray experiments were not presented. The authors further mentioned that all included data points were within 90% confidence limits in 3 of 5 experiments, but a higher confidence level would have provided more depth to their analysis and in the microarray in general.

Two almost simultaneous releases of EST data from *S. japonicum* (Hu *et al*. 2003) and *S. mansoni* (Verjovski-Almeida *et al*. 2003) have provided the

platform for study of the schistosome transcriptome. A large proportion of the EST sequences generated still require characterization as only 45% of the *S. mansoni* and 65% of the *S. japonicum* ESTs showed similarity to sequences already in GenBank. The *S. japonicum* EST data were structured into 13 131 clusters, including 8527 singletons of varying sizes and probably represent the majority of the transcriptome, estimated to contain approximately 15 000 genes (Johnston *et al.* 1999). We have recently designed and utilized an oligonucleotide microarray based on the *S. japonicum* and *S. mansoni* EST data sets. The designed microarray contains 7055 *S. japonicum* and 12 166 *S. mansoni* probes, effectively providing a much wider coverage of the schistosome transcriptome, which will facilitate in-depth comparative studies of the two species represented on the microarray.

(5) *Microarray analysis of infected host tissues and the effects of parasitism*

Another use for microarrays in parasitology is to examine the differential gene expression of host tissues in response to parasitic infections. A report by Sexton *et al.* (2004) described the use of a commercial oligonucleotide mouse array to monitor the effects of malaria infection. They identified hundreds of genes that were altered in response to the parasite. Modification in specific gene profiles that were related to glycolysis, immunology, and erythropoiesis were correlated with the underlying pathology of the infected host. While the study was comprehensive in terms of the microarray used, information of sampling numbers and whether tissues were pooled or not, was not provided. Significant variation in gene expression can occur between individuals, and this would be more pronounced with out-bred mice due to the wider genetic profile in the population under investigation. Fortunately, Sexton *et al.* (2004) used inbred C57BL/6 that would have given a more even genetic response. In pooling samples, subtle fluctuations in gene expression can be lost, or, alternatively, a single large variation in gene expression in one individual may present a misleading indication when the pooled group is examined. It is fully acknowledged that microarray experiments are expensive, but the study would have benefited if gene expression variation between individuals could have been determined. A compromise may have been to use a validation method (such as real-time PCR) to investigate individual RNA samples, in order to determine if any significant inter-individual variation occurred.

A similar study by Hoffmann *et al.* (2001) examined gene expression in schistosome-infected mouse liver using a cDNA microarray. They monitored various genes encoding immunological proteins in response to granuloma formation caused by *S. mansoni* infection. The unique pattern of gene regulation that was apparent, allowed a profile to be formulated that was able to differentiate a healthy host from an infected host by the gene expression within the liver.

A recent article by Diez-Tascon *et al.* (2005) has adapted a microarray resource to investigate the mechanism that leads to host immunity to parasitic infections. They utilized a bovine microarray to examine the transcriptional profile between genetically resistant and susceptible lambs infected with a mixture of naturally occurring nematode infections (including *Haemonchus contortus*, *Trichostrongylus colubriformis* and *Ostertagia* sp. although the precise composition of the initial infection was not clearly reported). The authors compared gene expression within the duodenum of the infected host, using a 10 204-element cDNA microarray derived from a previously undescribed bovine library. It can only be assumed that a microarray derived from ovine material was not constructed since the bovine microarray was available and homology between the two species is high (average nucleotide homology 96%, Diez-Tascon *et al.* 2005) in the coding region. Differential expression in 126 genes was detected, with 112 of these having reduced expression in the resistant host breed compared to susceptible animals. These differentially expressed genes were subjected to lexical analysis (aligned into known pathways) and then, assigned Gene Ontology. Two distinct metabolic pathways were identified one of which was involved in immune response acquisition and the other in intestinal smooth muscle neogenesis.

Other studies that have investigated the host-parasite interplay include, monitoring the mosquito host of the malaria parasite, for susceptibility, using a 1200 cDNA microarray (based on *Aedes aegypti* ESTs), with 28 genes identified that were differentially expressed between susceptible and resistant mosquito strains (Chen *et al.* 2004*b*). In addition, the responses to *Eimeria acervulina* and *Eimeria maxima* infections were analysed in avian hosts using a 400 element cDNA microarray consisting of unique chicken genes, revealed both up and down-regulation of a number of genes (Min *et al.* 2003). Both studies demonstrate the utility of microarrays to address very specific and unique questions in parasitology.

CONCLUDING REMARKS

Microarrays provide the perfect tool for the monitoring of gene expression in parasites given the inherent technical difficulties associated with tissue collection and the complexity of many parasite life-cycles. The formulation of a complete transcriptional profile of a parasite during its development and differentiation throughout the life-cycle,

or in response to external insults such as a drug or vaccine, can reveal many facets of its adaptive biology including the development of resistance (see Table 3).

REFERENCES

Bahl, A., Brunk, B., Crabtree, J., Fraunholz, M. J., Gajria, B., Grant, G. R., Ginsburg, H., Gupta, D., Kissinger, J. C., Labo, P., Li, L., Mailman, M. D., Milgram, A. J., Pearson, D. S., Roos, D. S., Schug, J., Stoeckert, C. J. Jr **and Whetzel, P.** (2003). PlasmoDB: the Plasmodium genome resource. A database integrating experimental and computational data. *Nucleic Acids Research* **31**, 212–215.

Baptista, C. S., Vencio, R. Z., Abdala, S., Valadares, M. P., Martins, C., De Braganca-Pereira, C. A. and Zingales, B. (2004). DNA microarrays for comparative genomics and analysis of gene expression in *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology* **138**, 183–194.

Blaxter, M., Daub, J., Guiliano, D., Parkinson, J. and Whitton, C. (2002). Filarial Genome Project. The Brugia malayi genome project: expressed sequence tags and gene discovery. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **96**, 7–17.

Bozdech, Z., Zhu, J., Joachimiak, M. P., Cohen, F. E., Pulliam, B. and Derisi, J. L. (2003). Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. *Genome Biology* **4**, R9.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C. J., Aach, J., Ansorge, W., Ball, C. A., Causton, H. C., Gaasterland, T., Glenisson, P., Holstege, F. C., Kim, I. F., Markowitz, V., Matese, J. C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R. and Vingron, M. (2001). Minimum information about a microarray experiment (MIAME) – toward standards for microarray data. *Nature Genetics* **29**, 365–371.

Carlton, J., Silva, J. and Hall, N. (2005). The genome of model malaria parasites, and comparative genomics. *Current Issues in Molecular Biology* **7**, 23–37.

Carucci, D. J., Gardner, M. J., Tettelin, H., Cummings, L. M., Smith, H. O., Adams, M. D., Hoffman, S. L. and Venter, J. C. (1998). The malaria genome sequencing project. *Expert Reviews in Molecular Medicine* **1**, 1–9.

Chen, H., Wang, J., Liang, P., Karsay-Klein, M., James, A. A., Brazeau, D. and Yan, G. (2004*b*). Microarray analysis for identification of *Plasmodium*-refractoriness candidate genes in mosquitoes. *Genome* **47**, 1061–1070.

Chen, Y. A., McKillen, D. J., Wu, S., Jenny, M. J., Chapman, R., Gross, P. S., Warr, G. W. and

Almeida, J. S. (2004*a*). Optimal cDNA microarray design using expressed sequence tags for organisms with limited genomic information. *BMC Bioinformatics* **5**, 191.

Clontech Laboratories (2001). Creator SMART cDNA library construction kit user manual, [available at http://www.bdbiosciences.com/clontech/techinfo/manuals/PDF/ PT3577-1.pdf]. BD Biosciences.

Diez-Tascon, C., Keane, O. M., Wilson, T., Zadissa, A., Hyndman, D. L., Baird, D. B., McEwan, J. C. and Crawford, A. M. (2005). Microarray analysis of selection lines from outbred populations to identify genes involved with nematode parasite resistance in sheep. *Physiological Genomics* **21**, 59–69.

Dugas, M., Merk, S., Breit, S. and Dirschedl, P. (2004). MDCLUST – exploratory microarray analysis by multidimensional clustering. *Bioinformatics* **20**, 931–936.

Fitzpatrick, J. M., Johansen, M. V., Johnston, D. A., Dunne, D. W. and Hoffmann, K. F. (2004). Gender-associated gene expression in two related strains of *Schistosoma japonicum*. *Molecular and Biochemical Parasitology* **136**, 191–209.

Fitzpatrick, J. M., Johnston, D. A., Williams, G. W., Williams, D. J., Freeman, T. C., Dunne, D. W. and Hoffmann, K. F. (2005). An oligonucleotide microarray for transcriptome analysis of *Schistosoma mansoni* and its application/use to investigate gender-associated gene expression. *Molecular and Biochemical Parasitology* **141**, 1–13.

Foster, J. M., Kumar, S., Ganatra, M. B., Kamal, I. H., Ware, J., Ingram, J., Pope-Chappell, J., Guiliano, D., Whitton, C., Daub, J., Blaxter, M. L. and Slatko, B. E. (2004). Construction of bacterial artificial chromosome libraries from the parasitic nematode *Brugia malayi* and physical mapping of the genome of its *Wolbachia* endosymbiont. *International Journal for Parasitology* **34**, 733–746.

Gissot, M., Refour, P., Briquet, S., Boschet, C., Coupe, S., Mazier, D. and Vaquero, C. (2004). Transcriptome of 3D7 and its gametocyte-less derivative F12 Plasmodium falciparum clones during erythrocytic development using a gene-specific microarray assigned to gene regulation, cell cycle and transcription factors. *Gene* **341**, 267–277.

Gu, Z., Rifkin, S. A., White, K. P. and Li, W. H. (2004). Duplicate genes increase gene expression diversity within and between species. *Nature Genetics* **36**, 577–579.

Hoffmann, K. F., Johnston, D. A. and Dunne, D. W. (2002). Identification of *Schistosoma mansoni* gender-associated gene transcripts by cDNA microarray profiling. *Genome Biology* **3**, 0041.1–0041.11.

Hoffmann, K. F., McCarty, T. C., Segal, D. H., Chiaramonte, M., Hesse, M., Davis, E. M., Cheever, A. W., Meltzer, P. S., Morse, Hc 3rd and Wynn, Ta. (2001). Disease fingerprinting with cDNA microarrays reveals distinct gene expression profiles in lethal type 1 and type 2 cytokine-mediated inflammatory reactions. *FASEB Journal* **15**, 2545–2547.

Hu, W., Yan, Q., Shen, D. K., Liu, F., Zhu, Z. D., Song, H. D., Xu, X. R., Wang, Z. J., Rong, Y. P., Zeng, L. C., Wu, J., Zhang, X., Wang, J. J., Xu, X. N.,

Wang, S. Y., Fu, G., Zhang, X. L., Wang, Z. Q., Brindley, P. J., McManus, D. P., Xue, C. L., Feng, Z., Chen, L. and Han, Z. G. (2003). Evolutionary and biomedical implications of a *Schistosoma japonicum* complimentary DNA resource. *Nature Genetics* **35**, 139–147.

Hughes, T. R., Mao, M., Jones, A. R., Burchard, J., Marton, M. J., Shannon, K. W., Lefkowitz, S. M., Ziman, M., Schelter, J. M., Meyer, M. R., Kobayashi, S., Davis, C., Dai, H., He, Y. D., Stephaniants, S. B., Cavet, G., Walker, W. L., West, A., Coffey, E., Shoemaker, D. D., Stoughton, R., Blanchard, A. P., Friend, S. and Linsley, P. S. (2001). Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nature Biotechnology* **19**, 342–347.

Johnston, D. A., Blaxter, M. L., Degrave, W. M., Foster, J., Ivens, A. C. and Melville, S. E. (1999). Genomics and the biology of parasites. *Bioessays* **21**, 131–147.

Khan, J., Bittner, M. L., Chen, Y., Meltzer, P. S. and Trent, J. M. (1999*a*). DNA microarray technology: the anticipated impact on the study of human disease. *Biochimica et Biophysica Acta* **1423**, M17–28.

Khan, J., Saal, L., Bittner, M., Chen, Y., Trent, J. and Meltzer, P. S. (1999*b*). Expression profiling in cancer using cDNA microarray. *Electrophoresis* **20**, 223–229.

Kissinger, J. C., Gajria, B., Li, L., Paulsen, I. T. and Roos, D. S. (2003). ToxoDB: accessing the *Toxoplasma gondii* genome. *Nucleic Acids Research* **31**, 234–236.

Knox, D. P. (2004). Technological advances and genomics in metazoan parasites. *International Journal for Parasitology* **34**, 139–152.

Kurella, M., Hsiao, L. L., Yoshida, T., Randall, J. D., Chow, G., Sarang, S. S., Jensen, R. V. and Gullans, S. R. (2001). DNA microarray analysis of complex biologic processes. *Journal of the American Society of Nephrology* **12**, 1072–1078.

Le Roch, K. G., Zhou, Y., Blair, P. L., Grainger, M., Moch, J. K., Haynes, J. D., De La Vega, P., Holder, A. A., Batalov, S., Carucci, D. J. and Winzeler, E. A. (2003). Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* **301**, 1503–1508.

Min, W., Lillehoj, H. S., Kim, S., Zhu, J. J., Beard, H., Alkharouf, N. and Matthews, B. F. (2003). Profiling local gene expression changes associated with *Eimeria maxima* and *Eimeria acervulina* using cDNA microarray. *Applied Microbiology and Biotechnology* **62**, 392–399.

Minning, T. A., Bua, J., Garcia, G. A., McGraw, R. A. and Tarleton, R. L. (2003). Microarray profiling of gene expression during trypomastigote to amastigote transition in *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology* **131**, 55–64.

Nielsen, H. B., Wernersson, R. and Knudsen, S. (2003). Design of oligonucleotides for microarrays and perspectives for design of multi-transcriptome arrays. *Nucleic Acids Research* **31**, 3491–3496.

Randall, R. L., Wade, M., Albritton, K. H., Coffin, C. M. and Joyner, D. E. (2003). Validation of cDNA microarray analysis to distinguish tumor type *ex vivo*. *Clinical Orthopaedics and Related Research* **415S**, S110–S119.

Renn, S. C., Aubin-Horth, N. and Hofmann, H. A. (2004). Biologically meaningful expression profiling across species using heterologous hybridization to a cDNA microarray. *BMC Genomics* **5**, 42.

Rise, M. L., Von Schalburg, K. R., Brown, G. D., Mawer, M. A., Devlin, R. H., Kuipers, N., Busby, M., Beetz-Sargent, M., Alberto, R., Gibbs, A. R., Hunt, P., Shukin, R., Zeznik, J. A., Nelson, C., Jones, S. R., Smailus, D. E., Jones, S. J., Schein, J. E., Marra, M. A., Butterfield, Y. S., Stott, J. M., Ng, S. H., Davidson, W. S. and Koop, B. F. (2004). Development and application of a salmonid EST database and cDNA microarray: data mining and interspecific hybridization characteristics. *Genome Research* **14**, 478–490.

Schena, M., Shalon, D., Davis, R. W. and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complimentary DNA microarray. *Science* **270**, 467–470.

Sexton, A. C., Good, R. T., Hansen, D. S., D'Ombrain, M. C., Buckingham, L., Simpson, K. and Schofield, L. (2004). Transcriptional profiling reveals suppressed erythropoiesis, up-regulated glycolysis, and interferon-associated responses in murine malaria. *Journal of Infectious Diseases* **189**, 1245–1256.

Street, M. (2002). DNA microarrays: manufacture and applications. *Australasian Biotechnology* **12**, 38–39.

Summan, M., McKinstry, M., Warren, G. L., Hulderman, T., Mishra, D., Brumbaugh, K., Luster, M. I. and Simeonova, P. P. (2003). Inflammatory mediators and skeletal muscle injury: a DNA microarray analysis. *Journal of Interferon and Cytokine Research* **23**, 237–245.

Tolstrup, N., Nielsen, P. S., Kolberg, J. G., Frankel, A. M., Vissing, H. and Kauppinen, S. (2003). OligoDesign: Optimal design of LNA (locked nucleic acid) oligonucleotide capture probes for gene expression profiling. *Nucleic Acids Research* **31**, 3758–3762.

Verjovski-Almeida, S., Demarco, R., Martins, E. A., Guimaraes, P. E., Ojopi, E. P., Paquola, A.C, Piazza, J. P., Nishiyamamy, J. R., Kitajima, J. P., Adamson, R. E., Ashton, P. D., Bonaldo, M. F., Coulson, P. S., Dillon, G. P., Farias, L. P., Gregorio, S. P., Ho, P. L., Leite, R. A., Malaquias, L. C., Marques, R. C., Miyasato, P. A., Nascimento, A. L., Ohlweiler, F. P., Reis, E. M., Ribeiro, M. A., Sa, R. G., Stukart, G. C., Soares, M. B., Gargioni, C., Kawano, T., Rodrigues, V., Madeira, A. M., Wilson, R. A., Menck, C. F., Setubal, J. C., Leite, L. C. and Dias-Neto, E. (2003). Transcriptome analysis of the acoelomate human parasite *Schistosoma mansoni*. *Nature Genetics* **35**, 148–157.

Verma, P. and Sharma, Y. D. (2003). Malaria genome project and its impact on the disease. *Journal of Vector Borne Diseases* **40**, 9–15.

Watanabe, T., Murata, Y., Oka, S., Iwahashi, H. (2004). A new approach to species determination for yeast strains: DNA microarray-based comparative genomic hybridization using a yeast DNA microarray with 6000 genes. *Yeast* **21**, 351–365.

Williams, S. A., Lizotte-Waniewski, M. R., Foster, J., Guiliano, D., Daub, J., Scott, A. L., Slatko, B. and Blaxter, M. L. (2000). The filarial genome project: analysis of the nuclear, mitochondrial and endosymbiont genomes of *Brugia malayi*. *International Journal for Parasitology* **30**, 411–409.

Wilson, R. J. (2004). The transcriptome: malariologists ride the wave. *Bioessays* **26**, 339–42.

Xu, P., Widmer, G., Wang, Y., Ozaki, L. S., Alves, J. M., Serrano, M. G., Puiu, D., Manque, P., Akiyoshi, D., Mackey, A. J., Pearson, W. R., Dear, P. H., Bankier, A. T., Peterson, D. L., Abrahamsen, M. S., Kapur, V., Tzipori, S. and Buck, G. A. (2004). The genome of *Cryptosporidium hominis*. *Nature* **431**, 1107–1112.

Ylostalo, J., Randall, A. C., Myers, T. A., Metzger, M., Krogstad, D. J. and Cogswell, F. B. (2005). Transcriptome profiles of host gene expression in a monkey model of human malaria. *The Journal of Infectious Diseases* **191**, 400–409.