An oligonucleotide microarray for transcriptome analysis of Schistosoma mansoni and its application/use to investigate gender-associated gene expression
Fitzpatrick, Jennifer M.; Johnston, David A.; Williams, Gary W.; Williams, Debbie J.; Freeman, Tom C.; Dunne, David W.; Hoffmann, Karl F.

Published in: Molecular and Biochemical Parasitology
DOI: 10.1016/j.molbiopara.2005.01.007
Publication date: 2005

Citation for published version (APA):
An oligonucleotide microarray for transcriptome analysis of
Schistosoma mansoni and its application/use to investigate
gender-associated gene expression

Jennifer M. Fitzpatrick, David A. Johnston, Gary W. Williams, Debbie J. Williams, Tom C. Freeman, David W. Dunne, Karl F. Hoffmann

Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK
Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK
Rosalind Franklin Centre for Genomics Research, Hinxton, Cambridge CB10 1SR, UK

Received 14 September 2004; received in revised form 12 December 2004; accepted 4 January 2005

Abstract

Global profiling transcriptomes of parasitic helminths offers the potential to simultaneously identify co-ordinately expressed genes, novel genetic programs and uniquely utilized metabolic pathways, which together provide an extensive and new resource for vaccine and drug discovery. We have exploited this post-genomic approach to fabricate the first oligonucleotide DNA microarray for gene expression analysis of the parasitic trematode Schistosoma mansoni. A total of 17,329 S. mansoni DNA sequences were used to design a microarray consisting of 7335 parasite elements or approximately 50% of this parasite’s transcriptome. Here, we describe the design of this new microarray resource and its evaluation by extending studies into gender-associated gene expression in adult schistosomes. We demonstrate a high degree of reproducibility in detecting transcriptional differences among biologically replicated experiments and the ability of the microarray to distinguish between the expression of closely related gene family members. Importantly, for issues related to sexual dimorphism, labor division, gamete production and drug target discovery, 197 transcripts demonstrated a gender-biased pattern of gene expression in the adult schistosome, greatly extending the number of sex-associated genes. These data demonstrate the power of this new resource to facilitate a greater understanding into the biological complexities of schistosome development and maturation useful for identifying novel intervention strategies.

Keywords: Schistosoma; Helminth; DNA microarray; Gene expression

1. Introduction

Parasite expressed sequence tag (EST) and genomic sequencing projects have proven to be an invaluable resource for parasite gene discovery and have led to the identification of numerous putative gene products [1–3]. Despite the large availability of DNA sequence information, true potential of this resource will only be realized upon the assignment of gene function within an actual biological and cellular context, thus leading to the possible functional annotation of many important parasitic genomes. Towards this goal, several investigators have developed DNA microarrays to probe and begin to elucidate the role of specific gene products in the lifestyle, pathogenicity and fundamental biology of multiple parasites [4–6]. This approach, in combination with continued wide ranging genomic and EST sequencing has brought together genomic and functional-genomic data...

© 2005 Published by Elsevier B.V.

Corresponding author. Tel.: +44 1223 333 338; fax: +44 1223 333 741.
E-mail address: kfh24@cam.ac.uk (K.F. Hoffmann).

Microarray data reported in this paper is available in the ArrayExpress database at EBI under the reference numbers A-MEXP-134 (description of S. mansoni oligonucleotide microarray) and E-MEXP-223 (all microarray data).

10166-6851/$ – see front matter © 2005 Published by Elsevier B.V.

doi:10.1016/j.molbiopara.2005.01.007
Gene discovery and functional analysis of schistosomes has also benefited from such an approach [5,7]. These studies have revealed cDNA microarrays to be particularly useful and reliable in identifying gender-associated transcripts in both Schistosoma mansoni [5] and S. japonicum [7]. However, while small sample size subsets proved fruitful for the initiation of these investigations, utilizing a high-throughput approach on a whole genome scale will dramatically increase the transcriptional understanding of parasite sexual biology where current thinking proposes that males and females evolved to maximize independent functional roles important to the survival of the parasite. Presently, the S. mansoni genome contains approximately 14,000 predicted genes, and like many other parasites, most of these genes display no database homology and therefore, have no functional annotation [8–10]. Assigning some putative function or association, based on expression profiling by DNA microarray analysis, may provide some of the most promising research areas for elucidating the molecular basis of parasite biology.

Here we describe the design, fabrication and validation of a new DNA microarray for schistosome transcriptome analysis based on the use of long oligonucleotide probes. Long oligonucleotide DNA microarrays were constructed as highly sensitive alternatives to cDNA microarrays due to inherent advantages throughout production and experimental use [11]. The DNA microarray contains 7335 S. mansoni probes covering approximately 50% of the total estimated gene complement and was used to expand upon our previous studies of schistosome conjugal biology. The experiments described provide a dual function: (1) to specifically characterize the fabricated oligonucleotide DNA microarray allowing sensitive, reproducible gene expression results to be generated by multiple users and laboratories; and (2) to further elucidate the expression profiles of male and female adult parasites with the goal of expanding our knowledge relating to sexual maturation, sexual dimorphism, labour division and gamete production. Ultimately, investigations into S. mansoni transcriptional mechanisms will likely generate new insights into the development and maintenance of this helminth’s dioecious lifestyle, leading to the identification of novel drug targets or vaccine candidates.

2. Materials and methods

2.1. Parasites

Adult male and female S. mansoni (NMRI Puerto Rican strain) were perfused from percutaneously infected mice at 7 weeks after challenge with independent batches of 250 cercariae each shed from albino Biomphalaria glabrata. After perfusion, both immature and mature worms were counted and sex-separated. Miracidia used to infect B. glabrata were hatched from eggs collected from mouse livers 7 weeks after infection [12].

2.2. Design of S. mansoni DNA oligonucleotide probes

S. mansoni DNA elements chosen for oligonucleotide design were selected from EST sequences available from public databases as of June 28, 2002, full-length mRNA and genomic DNA (gDNA) sequences (using NCBI Entrez limits excluding ESTs, STNs, GSSs, TPAs, patents and working drafts) available from public databases as of April 2003 and from one full-length mRNA sequence identified in our laboratory (AY267032—S. mansoni arginase). The 16,815 EST sequences were clustered using the CAP3 DNA sequence assembly program [13] into 2076 contigs (representing more than one EST sequence) and 5049 singletons (representing only one EST sequence) for a total of 7125 unique DNA sequence clusters. In addition to these 7125 non-redundant EST clusters, 513 full-length mRNA (some redundancy with respect to EST clusters) and gDNA sequences were included to bring the total number of DNA sequences used as templates for oligonucleotide design to 7638. CAP3 has previously been shown to be tolerant of sequencing errors resulting from single pass sequencing and is effective at differentiating between closely related gene family members [14].

Putative sequence homology of each schistosome DNA element was assigned using the web-based Basic Local Alignment Search Tool (BLASTx) [15] searching against the NCBI protein non-redundant (nr) database. BLASTx hits with an Expect-value (E-value) of $10^{-5}$ were considered significant and the corresponding NCBI protein nr designation was used to annotate the EST contigs and singletons. BLASTx hits with an E-value of $10^{-10}$ were not considered significant and therefore the corresponding schistosome DNA elements obtaining these scores were annotated as 'UNKNOWN'. In a further attempt to annotate these unknown schistosome DNA sequences, they were compared against the S. mansoni EST database compiled by Verjovski-Almeida et al. [9] using BLASTn, where DNA elements generating BLASTn bit score values of $200$ were considered a match. Subsequent annotation was provided then by the EST database assigned through additional BLASTx searches of NCBI (using their applied criteria [9]). Sequence similarity assignment was performed for two reasons: (1) to annotate the DNA sequence representations deposited on each DNA microarray; and (2) to identify which DNA strand to select for oligonucleotide design. For BLASTx searches with no significant similarity, the sense strand was selected for oligonucleotide design, unless there was a poly-T tract incorporated at one end of the parent DNA sequence (anti-sense strand was used in this case). Therefore, a small percentage of oligonucleotides deposited on this DNA microarray may have represented the non-coding DNA strand. The labelling procedure used in this study [16] (Klenow incorpo-
ration of Cy3-dCTP molecules into double stranded cDNA) fluorescently labels both strands of cDNA, which effectively compensates for this small percentage of oligonucleotides.

All oligonucleotides were 50 bases in length, modified by a 6-C linked 5′ amino modification and synthesized at 200 nmol scale (Illumina, SD, CA). Multiple hierarchical tests were performed on each of the 7638 DNA sequences to select the optimum 50-mer oligonucleotide representative for each contig/singleton/mRNA gene. The European Molecular Biology Open Software Suite (EMBOSS [17]) hosted at SourceForge (http://emboss.sourceforge.net/) was used for most of the analyses. Initially, each sequence was required to be a minimum of 50 bases in length. Those sequences passing this first test were then subjected to a series of selection criteria. Using the EMBOSS program ‘dan’, regions were identified that were within the required GC content of 30–50%. The program ‘palindrome’ was used to identify regions with no gross secondary structure (defined as having seven or more bases capable of forming a perfect hairpin structure). The program ‘break’ was used to identify regions that had no ambiguous nucleotide codes. Additionally, using the program ‘RepeatMasker’ (http://www.repeatmasker.org), regions were identified that contained no low complexity sections, simple repeats or sequences matching the S. mansoni SR2 retrotransposon sequence [18]. Using BLASTs, each potential oligonucleotide source region was then compared against all 7638 S. mansoni sequences and areas with no matches of 12 or more bases were identified. If this BLAST criterion proved too restrictive to allow regions in a sequence to be used as oligonucleotides, then it was progressively relaxed to ≥20 and then ≥40 bases, respectively.

Starting from the 5′ end of each sequence, a maximum of four non-overlapping regions of sequence with a length of 50 bases or more that passed the selections criteria were identified. The Tm of these possible oligonucleotides was determined using the EMBOSS program ‘dan’, and the one from each sequence with a Tm closest to 72 °C was selected for oligonucleotide design. Originating from the initial 7638 DNA sequences, 7214 passed all of these tests and were used to design 50-mer oligonucleotides. The remaining 424 DNA sequences that failed were again subjected to the same tests, but with the GC content relaxed to 25–50%. Out of these 424 DNA sequences, 121 could generate a 50-mer oligonucleotide. The remaining 303 sequences failed these tests due to length (<50 bases), high ambiguity or a GC content outside of the required ranges. All sequences were analyzed using the InterProScan package [19] and Gene Ontology (GO) [20] terms were extracted from the output of this analysis using a perl script. A total of 7335 50-mer S. mansoni oligonucleotides were designed and, along with control oligonucleotides (Bacillus subtilis-specific and Arabidopsis thaliana-specific DNA elements), represented the oligonucleotide DNA microarray probes used in this study (Table 1). Oligonucleotide DNA microarrays were printed on CodeLink Activated Slides (Amine-Binding Slides)
denatured at 95 °C for 5 min then at 50 °C for a further 5 min. Microarray hybridization was performed in a humidified chamber at 45 °C for 16–18 h. Three successive, post-hybridization stringency washes were performed (2× SSC, 0.1× SSC/0.1% SDS and 0.1× SSC) for 3-min/wash solution at room temperature with agitation. Image acquisition (16-bit tiff) for the DNA microarray was performed using a GenePix® 4100 A (Axon Instruments Inc.) dual channel laser scanner at 10 μm resolution, 100% laser power and PMT levels ranging from 580 to 730. Image analysis used the Microarray Suite extension of IP Lab for Macintosh (Sc-analytics) software.

2.5. DNA microarray data filtering

Data analysis was performed as previously described [7]. Briefly, all data was intensity-dependent normalized using the program Vector Xpression 3.1 (Invitrogen® Life Technologies). Expression ratio (Ex. ratio) was thus defined as the normalized Log2(Cy5/Cy3) ratio. Poor quality spots and low intensity data were filtered and removed by a succession of applied statistical criteria. Initially, the arithmetic mean was calculated for all non-S. mansoni spots contained on the microarray (3650 duplicated negative controls, includ- ing B. subtilis, A. thaliana and buffer spots). The mean signal intensity for each hybridized S. mansoni element was required to be greater than one standard deviation above the mean of negative controls in at least one channel (Cy5 or Cy3). All data below this value were removed from further analysis. Oligonucleotides passing these filtering criteria had to display Log2 normalized expression ratios outside of the 90% confidence interval in at least three out of five replicate DNA microarray hybridizations to be included in the final set of differentially expressed, sex-associated transcripts. Correlation coefficients (R) for biological replicate hybridizations were derived from a goodness of fit measure of a linear model where values approaching one indicate a high level of agreement. Hierarchical cluster analysis was applied using single linkage and Euclidian distance correlation matrices. Care was taken to ensure all information is MIAME [21] compliant and all data was submitted to MIAMEExpress at European Bioinformatics Institute, Hinxton, UK (http://www.ebi.ac.uk/miameexpress/)[22].

2.6. RT-PCR analysis

Reverse transcription PCR was carried out using 1 μg of parasite total RNA as described [23]. All RNA was treated with DNase I (Ambion Inc.) prior to reverse transcription to remove any potential genomic DNA contamination. The oligonucleotide primers for PCR are shown in Table 2 (see supplementary information). All amplifications were separated on a 2% agarose gel, detected using ethidium bromide and images captured using digital photography.

3. Results

3.1. S. mansoni sequence analysis and fabrication of a novel oligonucleotide microarray to profile schistosome gene expression

Clustering of the parasite EST sequences contained in NCBI protein nr databases as of June 2002 (using the CAP3 program) and inclusion of full-length genes as of April 2003, led to a collection of 7638 sequences available for oligonucleotide design. Selective criteria for design of opti- mal oligonucleotides enabled 7335 S. mansoni oligonu- cleotides (duplicated on microarray to 14,670) to be syn- thesized and arrayed (Table 1). BLASTx manual annota- tion of each contig/EST revealed 4621 sequences displaying no significant similarity using BLASTx p-value (E) crite- ria of ≤10−10, representing ~63% of the total S. mansoni sequences. Therefore, from the initial BLASTx annotation, 2714 sequences displayed significant similarity to known genes in NCBI (~37% of total S. mansoni sequences). The 4621 sequences without similarity were then submitted to BLASTx-analysis at the University of Sao Paulo/ONCA. S. mansoni EST database [9,24] and hits with a bit score value of ≥200 were merged with the sequences to create longer con- tigs. These new contigs were then re-submitted to the NCBI nr protein database using BLASTx in an attempt to annotate them. This left 2591 sequences still with no significant similarity to entries deposited in NCBI (using same BLASTx criteria as previously described). In contrast, 2030 sequences gained some annotation through this second database search, 1156 of which having significant similarity to hypothetical proteins (874 of these containing some functional annota- tion). Therefore, the total number of S. mansoni sequences used for oligonucleotide design displaying some annotation is approximately 65% (49% without hypothetical proteins). Additionally, 1140 sequences were potentially presented in the databases in the anti-sense direction and therefore the fi- nal oligonucleotide was designed from the sense strand (see Table 1).

All sequences were further analyzed for putative func- tion using the InterProScan package [19] and GO terms were assigned to those sequences with recognized potential func- tional domains. Briefly, 3605 GO terms were assigned to 1242 S. mansoni sequences (Table 1). The GO terms assigned (476 unique) included 249 molecular function, 161 biological pro- cess and 66 cell component terms. This reveals the extensive range of potential biological activities and processes repre- sented by the S. mansoni sequences deposited on this DNA microarray (see supplementary information).

3.2. Adult gender-associated transcripts are reproducibly detected by fabricated S. mansoni oligonucleotide microarrays

As multi-gene families are represented throughout the S. mansoni genome (cathepsins, superoxide dismutase, dynein...
light chains etc.) it was important to ensure that the 50-mer representation of each DNA element was capable of distinguishing among individual members. Results presented here suggest that multiple *S. mansoni* oligonucleotides representing various members of closely related gene families are able to distinguish among gene family members (Fig. 1) Expression values show that the oligonucleotides were designed with optimal physical properties to allow sensitive and reproducible hybridization, e.g., consistent *Tm*, minimal secondary structure, minimal sequence similarity to other oligonucleotides within the array (not representing the same gene), not designed over repeat regions or other regions of low complexity. Utilizing two previously reported and well-characterized superoxide dismutase (SOD) genes as a model ([25,26]), expression values were reported for multiple oligonucleotides representing these genes. Each independent oligonucleotide representing different regions of the same gene, gave comparable expression values as illustrated in Fig. 1A and B. Five oligonucleotides representing exon regions of extracellular superoxide dismutase (M27529) all demonstrated female-associated gene expression values (in agreement with results obtained by Fitzpatrick et al. [7] using a *S. japonicum* cDNA microarray). Normalized Log2 expression values ranged from −3.4 to −6.1 as means of five separate experiments. This finding was independently confirmed by gene-specific RT-PCR in comparison to the constitutively expressed (in adult parasites [27]) α- tubulin gene. In contrast, four oligonucleotides representing cytosolic superoxide dismutase (MW7298) (Fig. 1B) showed equal expression between the two sexes (again in agreement with Fitzpatrick et al. [7]). Furthermore, oligonucleotide (a) (Fig. 1B) displays zero expression values (no signal in either channel above background—NS), being representative of an intron region of the cytosolic SOD gene. Additional confirmation by RT-PCR demonstrated the equivalent expression of cytosolic SOD between male and female *S. mansoni* parasites.

*S. mansoni* oligonucleotide microarrays performed with a high degree of precision and were reproducible in generating similar expression results from five independent biologically replicated experiments. Each individual microarray experiment represented a biological replication of an adult male and adult female comparison (i.e., different batches of worms removed from mice infected with different independent batches of cercariae). Analyzing the reproducibility between these individual and discrete experiments suggest that the microarrays are capable of generating information which is directly comparable between multiple and distinct biological samples of parasites. Scatter plots comparing the expression values from one biological experiment to a different biological experiment demonstrated a high correlation coefficient between experiments (Fig. 1C), these representations displaying *R* = 0.848 (experiments 3 and 5, *n* = 4524) and *R* = 0.816 (experiments 2 and 5, *n* = 4645), respectively. The comparisons were performed using all oligonucleotides that passed the initial filtering criteria (expression values above background levels in the experiments compared, as calculated from negative controls, see Section 2).

The number of genes detected as expressed in the adult ranged from 4678 to 5169 (mean 4973). This figure equates to around 65% of the oligonucleotides deposited on the microarray. Taking into account some redundancy within the arrayed DNA elements, this percentage is comparable with that predicted by Verjoski-Almeida et al. [9] where it was estimated by two independent methods, that ~50% of the gene complement (~14,000 genes) of *S. mansoni* was expressed in the adult stage.

### 3.3. Expression profiling of sexually mature adult-stage
*S. mansoni* using the oligonucleotide DNA microarray reveals novel gender-associated transcripts

As quality control experiments illustrated the reproducible nature of this *S. mansoni* DNA oligonucleotide microarray, analyzing the differences in gene expression between adult female and adult male parasites initially validated this resource. Here, all transcripts identified as differentially expressed displayed Log2 normalized gene expression ratios outside of the 90% confidence interval in three out of five independent, biologically replicated experiments. A hierarchical clusterogram (single linkage analysis/Euclidian distance correlation) shown in Fig. 2 displays all transcripts identified as being highly differentially expressed in either male (fold: positive ratio) or female (green: negative ratio) parasites. These male/female bimodal comparisons revealed 141 unique genes highly expressed in the female when compared to the male, and 86 genes highly expressed in the male when compared to the female. This considerably increases the number of known sex-associated transcripts in *S. mansoni* adult parasites. Importantly, dye-swap experiments showed a high correlation with data passing these strict filtering criteria and therefore suggest minimal dye effect on the prediction of differentially expressed transcripts. Twenty-four oligonucleotides (16 classified as having no significantly sequence similarity) highly expressed in adult females did not reproduce as differentially expressed when compared to adult male in multiple pooled RNA dye-swap experiments. Moreover, only six oligonucleotides described as male associated did not repeat using the dye-swap experiments. Therefore, after considering the minimal dye effect observed in our experiments, 117 unique genes were highly expressed in the adult female and 80 unique genes were highly expressed in the adult male (see supplementary information for full list of oligonucleotides identified as differentially expressed between the adult genders).

Gene expression results for all differentially expressed transcripts relate to the vast majority of previously appreciated gender-associated genes. Microarray expression values reveal these putative positive controls to be differentially expressed in the previously documented gender-associated
Fig. 1. *S. mansoni* oligonucleotide DNA microarrays are capable of generating reproducible expression results and can distinguish transcriptional differences among gene family members. (A) Hybridization performance of multiple oligonucleotides spanning different intragenic regions of two superoxide dismutase homologues. Five oligonucleotides spanning the three extracellular superoxide dismutase exons (Ex-SOD, M27529) reveal similar gene expression profiles (for all five biological replicates averaged) and strongly support the female-enriched expression of this SOD homolog (verified by gene-specific RT-PCR). (B) Cytosolic SOD (Cyt-SOD, M97298) did not display a statistically significant gender association as determined from hybridization information originating from four oligonucleotides spanning two exons and the 3′-UTR. No signal intensity (NS) was observed for a Cyt-SOD specific oligonucleotide designed over an intron. Boxes represent exons, solid lines represent non-coding DNA elements and dashed lines represent positions of designed oligonucleotides for the 5.1 kb Ex-SOD and the 8.5 kb Cyt-SOD genes. RT-PCR conditions and derivation of expression ratio (Ex. ratio) are described in Section 2. (C) Statistical analysis of biologically reproduced DNA microarray hybridization experiments. Scatterplots compare the Log2-calibrated ratio generated for each oligonucleotide probe (mean of two replicate spots) from one biological batch of adult worm material to the Log2-calibrated ratio generated for the same oligonucleotide probe (mean of two replicate spots) from a different biological batch. The correlation coefficient values from two representative comparisons, R = 0.848 and 0.816 indicate a high degree of agreement between biologically replicated experiments. Scatterplots display oligonucleotide probes that passed the initial filtering criteria (signal intensity greater than 1 S.D. above the mean of negative control elements) for quality data in each biological batch/hybridization comparison and include 4524 for experiments 3 and 5, and 4645 elements for experiments 1 and 5. Lines represent the line of regression (centre line) and the predicted 99% confidence intervals of the plotted data.
Fig. 2. *S. mansoni* oligonucleotide microarrays can reproducibly detect gender-enriched gene expression profiles in sexually mature *S. mansoni* male and female adult parasites. Single-linkage hierarchical clustering (using Euclidean distance correlation matrices) of all gender-enriched transcripts passing filtering criteria from five independent experimental replicates and dye-swap experiments. Representative gender-enriched transcripts verified by RT-PCR (Fig. 3) are listed next to the clustergram along with their unique identifier (contig ID, accession number or name). All other gender-enriched transcripts identified in this study can be found in accompanying supplementary information. Increasing shades of red depict greater expression in adult male parasites, increasing shades of green represent greater expression in adult female parasites and grey represents gene expression data absent for that particular replicate experiment.

Fig. 2. *S. mansoni* oligonucleotide microarrays can reproducibly detect gender-enriched gene expression profiles in sexually mature *S. mansoni* male and female adult parasites. Single-linkage hierarchical clustering (using Euclidean distance correlation matrices) of all gender-enriched transcripts passing filtering criteria from five independent experimental replicates and dye-swap experiments. Representative gender-enriched transcripts verified by RT-PCR (Fig. 3) are listed next to the clustergram along with their unique identifier (contig ID, accession number or name). All other gender-enriched transcripts identified in this study can be found in accompanying supplementary information. Increasing shades of red depict greater expression in adult male parasites, increasing shades of green represent greater expression in adult female parasites and grey represents gene expression data absent for that particular replicate experiment.
4. Discussion

DNA microarray analysis is now a well-established functional genomics tool for the global analysis of gene expression [41,42]. The oligonucleotide microarray fabricated here represents ~50% of the estimated S. mansoni gene complement [9]. The sequences utilized for oligonucleotide design are proportionally representative of the entire S. mansoni genome, since the relative percentages of sequences displaying significant similarity to known genes in the databases is similar to that estimated by large-scale EST sequencing efforts [9]. Furthermore, the distribution of sequences between distinct biological functions and processes reflects the transcriptome of this parasite as a whole and suggests diverse roles in numerous biological functions for independent sequences. Since the sequences examined here represent a range of important processes and originate from diverse developmental stages, this established DNA microarray has major implications for probing different biological questions.

Although DNA microarray technology is relatively widely used, this is the first time long DNA oligonucleotides have been employed for large-scale profiling of gene expression in S. mansoni. Consequently, in order to characterize, optimize and assess this post-genomics tool, a set of experiments profiling the transcriptome of mature adult schistosomes was performed, similar to previous studies [5,7]. While sexually mature, adult male and female parasites transcribe the majority of genes to a similar degree, it is anticipated that numerous genes will be differentially expressed, as the sexes are morphologically, functionally and chromosomally distinct. Having evolved from hermaphroditic ancestors, it is likely that differential gene expression has driven sexual dimorphism and labour division within the genus and ultimately led to cooperative conjugal biology as a highly successful means of maximizing parasite transmission. Although highly effective for continuing the lifecycle, intravascular conjugal biology resulting in egg production by adult female parasites also contributes to host-mediated, inflammatory, cirnouval immune responses. If uncontrolled, these immune responses can precipitate a series of pathohological complications resulting in severe morbidity and mortality in infected individuals [43].

Since the transcriptional basis of adult sexual maturation and egg production remains relatively unknown, any information gained will be crucial to the elucidation of the specific processes involved. To be classified here as differentially expressed between the sexes, genes had to pass independent statistical criteria: (1) expression values had to be significantly above background levels in order to remain in the dataset (as calculated from negative control elements); and (2) genes had to be outside of the 90% confidence interval of the entire range of expressed data in three of five independent biological replicates. In these criteria enabled a total of 117 female and 80 male genes to be reproducibly and confidently identified as differentially expressed.
Fig. 3. RT-PCR analysis confirmed gender-enriched DNA microarray data for all transcripts tested, illustrating a high correlation between global gene expression profiling and individual gene transcript abundance. One microgram of pooled (five experimental replicates) *S. mansoni* total RNA (DNase I treated (Ambion Inc.)) was used in a RT reaction to prime cDNA synthesis as described in Section 2. Unique identifier (Unique ID) and BLASTx NCBI nr protein database annotation of amplified transcript, sequence of each PCR primer pair, cycle number, expected product size and annealing temperatures for each verified cDNA are listed in Table 2, supplementary information. (A) RT-PCR confirmation of randomly selected female-enriched transcripts identified by DNA microarray analysis. (B) RT-PCR confirmation of randomly selected male-enriched transcripts identified by DNA microarray analysis. Amplification of constitutive α-tubulin served as an internal control for all PCR reactions. (*) indicates female-enriched expression of a glutathione peroxidase ortholog determined solely by RT-PCR as oligonucleotides for SCMGPX1 deposited on DNA microarray were designed over intronic regions. (‡) denotes two transcripts that did not pass filtering criteria for differential expression in the dye-swap experiments. BLASTx E-values for amplified gene products are given to indicate sequence similarity.
pressed, revealing a massive amount of new molecular data.

Many positive controls confirmed the validity of these results, including the female-associated eggshell proteins, as previously demonstrated in *S. mansoni* [5] and *S. japonicum* cDNA [7] microarray experiments.

Investigating the putative function of many of these gender-associated transcripts has revealed a number of interesting and previously unexplored patterns of expression.

It is now clear that the vast majority of male-associated transcripts are intimately involved in the structural organization of the parasite (supplementary information) which contrasts directly with those transcripts highly expressed in the adult female, some having a presumed role in reproductive and egg-laying mechanisms. Structural elements expressed differentially by the male include components of the muscular system, tegumental proteins and the underlying cytoskeleton. This apparent ‘division of labour’ between the genders [44, 45] now has an extensive transcriptional basis. The male ensures the survival of the egg-laying female by providing physical support and musculature to aid feeding [46], physiological transportation within the vasculature [47] and potential extracellular maturation factors [48, 49]. The female is thus allowed to concentrate energy expenditure on egg-production, as reflected by the large number of transcripts involved in reproduction and development.

Differentially expressed transcripts of special interest were revealed in both genders (a number of which were subsequently confirmed by RT-PCR analysis, Fig. 3). Evidence exists to the importance of regulated intracellular calcium levels to parasitism by *S. mansoni*; Ca²⁺ signalling is absolutely required for muscular contraction (important more so in the male given their extensive muscular scaffold and infrastructure). Multiple voltage-gated Ca²⁺ channels have been identified within schistosomes, including both α- and β-subunits of *S. mansoni* (and *S. japonicum*) [50, 51]. Here, microarray analysis showed the expression of SmCaVβ₁ to be significantly greater in the male than the female (in direct contrast to SmCaVβ₂, which appears to be similarly expressed between the two genders). Although in one dye combination, SmCaVβ₁ did not demonstrate a significantly strong male association, RT-PCR analysis of this transcript (Fig. 3) confirmed the original microarray results and suggested that dye-swap experiments can be influenced at multiple levels and should be interpreted with caution. Praziqanthel (PZQ), the current drug of choice for the treatment of schistosomiasis, is believed to instigate disruption of Ca²⁺ homeostasis within the parasite, although the exact mechanism is unknown [52, 53]. It has become clear in recent years that female parasites are much less sensitive to the action of PZQ than males [54], the SmCaVβ₁ subunit being intimately involved in the conference of sensitivity [55]. Moreover, these DNA microarray results also revealed a male-associated expression of desmoyokin (AHNAK protein) [55]. Although AHNAK may possess multiple functions within a cell, it has been shown to interact specifically with the β-subunits (1, 2 and 2a) of the L-type Ca²⁺ channel and with F-actin of human heart muscle, mediating directly Ca²⁺ signal transduction [56]. It is possible then that high levels of SmCaVα₁ and AHNAK in the male may be instrumental in their increased sensitivity to PZQ observed when compared to females. Numerous other important Ca²⁺-binding proteins (CaBP) were additionally identified as being male-associated, including calpain (calcium-activated neutral protease (CANP)) [57], calmodulin [58], Sm20 [59] and a ryanodine receptor [60].

From the list of differentially expressed transcripts, females express a higher proportion of enzymes than do males. Included here as specifically female-associated is the expression of two endoplasmic reticulum neutral lipid synthesis enzymes acyl-CoA:diacylglycerol acyltransferase 2-like (DGAT2-like) and acyl-CoA cholesterol-acyltransferase-1 (ACAT1). ACAT covalently joins cholesterol and fatty acyl-CoA molecules to form cholesterol esters [61–63] and DGAT catalyzes a similar reaction to generate triglycerides, using diacylglycerol as the acyl group acceptor [64, 65]. Subsequent sequence analysis and database queries identified a further two potential DGAT1 and DGAT2-like molecules in *S. mansoni*. RT-PCR analysis (in comparison to a-tubulin) revealed a female-associated expression bias for each of these four genes (data not shown). *S. mansoni* is known to be unable to synthesize fatty acids and sterols de novo [66]. The specific role of these enzymes in the female at this time is still unclear, although the generation of triglyceride stores within the parasite is likely to represent a major function. Since schistosomes do not generate energy/ATP through the β-oxidation of fatty acids [67] (despite much of the enzymatic capacity for this being encoded within the genome, data not shown), the parasites may use neutral lipids as stores to guard against high intracellular (toxic) levels of free fatty acids and cholesterol. In addition, the adult female may utilize neutral lipids during egg production.

Concomitant with the egg-laying process, the female consumes a large excess of red blood cells in comparison to the male [68]. The female therefore requires the by-products of haemoglobin digestion to be metabolized, eliminated and stored to reduce direct toxicity and lipid peroxidation. The high expression of the enzymes superoxide dismutase and aspartate aminotransferase, together with the previously known female-associated cathepsins and aspartic proteases [69] and ferritin-1 (stores Fe³⁺ in non-toxic form) [35] in the female, as shown by both DNA microarray analysis and RT-PCR, suggests these enzymes may share a functional role in this process.

The DNA microarray analysis described in this study and combined with previous studies [5, 7] revealed a far greater number of transcripts displaying female-associated expression than male. This may be directly indicative of the enhanced metabolic requirements of extensive egg-production, but in addition, may also reflect the presence of actual egg transcripts within the female sample pool. Current investigations utilizing this oligonucleotide microarray to longitudinally analyze the development of the parasite within the host will identify such transcripts and enable a more detailed
analysis of the transition between female and egg as well as subsequent maturation throughout the lifecycle. It is important and timely to note here that numerous transcripts were additionally analyzed at the protein/enzyme activity level. All enzyme activities tested, including tyrosinase and aspartate aminotransferase correlated directly with the predicted pattern of mRNA transcript expression (data not shown).

This study suggests that the design and use of an oligonucleotide microarray to profile the S. mansoni transcriptome, as well as a feasible and an efficient strategy to examine transcriptional differences between parasite life-stages, time-points, drug challenges and physiological conditions, for example. This oligonucleotide DNA microarray is presently being employed by other laboratories interested in different aspects of schistosome biology and will facilitate joint and collaborative efforts towards a better understanding of this important pathogen. Future investigations into individual functions, specifics and potential interactions of these newly-defined and potential fundamentally-important genes will further define their differential role(s) in the conjugal biology of schistosomes (several of these functional studies are currently ongoing). Highlighting the potential functional properties of unknown genes, by virtue of their expression profiles, has provided priorities and avenues for further research. Whole genome sequence annotation, in conjunction with further DNA microarray analyses, along with other techniques, such as signal-sequence trap [70], RNAi [71,72], proteomics [73] and investigations of large-scale protein–protein interactions [74] provides prospects to dramatically improve the current understanding of the molecular basis of pathogenicity and pathology by this complex and important major parasite of humans.

Acknowledgements

This work was supported by means of a Wellcome Trust Career Development grant awarded to KFH and a Wellcome Trust Career Development grant awarded to DAJ. We thank Maureen Laidlaw and Frances Jones for excellent technical help with and Andrew McArdle for computer support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.molbiopara.2005.01.007.

References


[5] Hoffmann KF, Johnston DA, Danne DW. Identification of Schisto-


[8] Gardner MI, Hall N, Fung E, et al. Genome sequence of the hu-


[10] Hu W, Yan Q, Shen DK, et al. Evolutionary and biomedical im-


[22] MIAPEexpress. (http://www.ebi.ac.uk/miapeexpress/).


[37] Kohn AB, Roberts-Misterly JM, Anderson PA, Khan N, Greenberg RM. Specific sites in the beta interaction domain of a schistosome Ca(2+)-channel subunit are key to its role in sensitivity to the anti-schistosomal drug praziquantel. Parasitology 2001;127(Part 4):349–56.

[38] Kohn AB, Roberts-Misterly JM, Anderson PA, Khan N, Greenberg RM. Specific sites in the beta interaction domain of a schistosome Ca(2+)-Ca(2+)-channel subunit are key to its role in sensitivity to the anti-schistosomal drug praziquantel. Parasitology 2001;127(Part 4):349–56.


[70] Smyth D, McManus DP, Smout MJ, et al. Isolation of cDNA NAs encoding secreted and transmembrane proteins from Schisto-

[71] Skelly PJ, Da'dara A, Harn DA. Suppression of cathepsin B expres-

[72] Boyle JP, Wu XJ, Sheemaker CB, Yoshino TP. Using RNA inter-
ference to manipulate endogenous gene expression in Schisto-
15.


32.