ORIGINAL PAPER

Transcriptome profilings of female Schistosoma japonicum reveal significant differential expression of genes after pairing

Jun Sun · Su-Wen Wang · Chen Li · Wei Hu ·
Yi-Jiu Ren · Jin-Qiang Wang

Abstract Pairing of *Schistosoma japonicum* initiates female development, leads to female sexual maturation, and maintains this mature state. To understand the mechanism involved in these processes, we studied parasites isolated from single- and double-sex cercariae-infected mice using deep-sequencing analysis, Solexa, to uncover pair-regulated transcriptional profiles. In this study, we report the results of high-throughput tag-sequencing (Tag-seq) analysis of the transcriptome of female worms 18 and 23 days postsingle- and double-sex infections. We sequenced over 3 million tags, obtained a total of 14,034, 27,251, 22,755, and 22,555 distinct tags corresponding to 5,773, 9,794, 8,885, and 8,870 tag-mapped genes for 23-day-old female schistosomula from double-sex infections (23DSI), 23-day-old female schistosomula from single-sex infections (23SSI), 18-day-old female schistosomula from double-sex infections (18DSI), and 18-day-old female schistosomula from single-sex infections (18SSI), respectively. Analyses of differentially expressed genes revealed similarities in the gene expression profiles between 18SSI and 18DSI as well as rational differential gene expression between 18SSI and 23SSI. However, fewer upregulated genes were found in 23DSI compared with 18DSI. Of the 3,446 differentially expressed genes between 23DSI and 23SSI, 2,913 genes were upregulated in 23SSI, whereas only 533 genes were upregulated in 23DSI. In these upregulated genes in 23DSI, phosphoglycerate mutase, superoxide dismutase, egg antigen, ribosomal proteins, ferritin-1 heavy chain, and eukaryotic translation initiation factor 2 were detected. Detection of these genes suggests that gene expression in 23DSI is specialized for functions such as promotion and maintenance of female sexual maturation and egg production. Quantitative real-time (RT)-PCR analysis confirmed the Solexa results, thereby supporting the reliability of the system. Our results offer new insights into the biological significance of pairing, which directs the expression of genes specific for sexual maturation and egg production.

J. Sun (✉) · S.-W. Wang · C. Li · Y.-J. Ren · J.-Q. Wang
Institute for Infectious Diseases and Vaccine Development, Tongji University School of Medicine, Shanghai 200092, People's Republic of China
e-mail: swksj@tongji.edu.cn

S.-W. Wang
e-mail: wswtjzlx2012@hotmail.com

C. Li
e-mail: 133449096@qq.com

Y.-J. Ren
e-mail: 804745919@qq.com

J.-Q. Wang
e-mail: 867031496@qq.com

W. Hu
Department of Microbiology and Microbial Engineering, School of Life Science, Fudan University, Shanghai 200433, People's Republic of China
e-mail: hubeixf@yahoo.com

Introduction

Schistosomiasis is a chronic and debilitating parasitic disease caused by blood flukes of the genus *Schistosoma*. Over 200 million people are infected in more than 70 countries and close to 800 million are at risk (Steinmann et al. 2006). In P.R. China, approximately 65 million individuals are at risk of infection, and 320,000 cases of schistosomiasis japonica were estimated to occur in 2010 (Lei et al. 2010). The disease is transmitted when parasitic eggs in the feces of the host that reach fresh water and hatch into miracidia, which then infect an appropriate snail species and transform into sporocysts. The sporocysts asexually reproduce, thereby generating a

large number of cercaria. These offspring are shed by the snail and swim in fresh water until they find and infect appropriate vertebrate hosts. The worms then undergo differentiation and migrate to the bloodstream until they reach the mesenteric veins. In these veins, male and female worms pair eventually reach sexual maturity. These developmental phases are accompanied by remarkable morphological, biochemical, and molecular changes throughout the different life cycle stages of the flukes (Peng et al. 2003; Vermeire et al. 2006; Fitzpatrick and Hoffmann 2006; Sun et al. 2013; Williams et al. 2007; Pereira et al. 2013). Pairing is essential for triggering female sexual maturation. Previous research has proposed that the male contact is necessary for vitelline gland development and complete maturation of the ovary (Shaw 1987). Moreover, this interaction is ultimately linked to sexual maturation and maintenance of the mature state of females (Clough 1981; Popiel and Basch 1984; Shaw 1977). To date, the exact mechanisms by which males influence female maturation are unknown. Different factors have been suggested to be involved in schistosome sexual maturation, including physical or tactile contact (Armstrong 1965; Popiel and Basch 1984), nutrition (Basch 1990; Gupta and Basch 1987), and chemical stimuli (Haseeb 1998; Shaw et al. 1977). However, the mechanisms by which these factors facilitate female development have yet to be investigated.

Gender differences in gene expression of schistosomes, as well as stage-specific, strain-specific, maturation-specific, and species-specific differences, have been reported (Dillon et al. 2006; Fitzpatrick and Hoffmann 2006; Fitzpatrick et al. 2004; Gobert et al. 2006; Hoffmann et al. 2002; Moertel et al. 2006; Vermeire et al. 2006). These studies have provided important information concerning female maturation. Undoubtedly, female sexual maturation depends on continuous male interaction, which intimately affects transcriptional regulation in females. While a series of related studies has been completed using *S. mansoni*- and *Schistosoma japonicum*-specific DNA microarrays to identify gender-enriched gene transcripts associated with sexually mature stages of these trematodes (Fitzpatrick and Hoffmann 2006; Fitzpatrick et al. 2004; Hoffmann et al. 2002) and transcriptional information related to the dioecious state has been obtained, the expression profiles of females developed after pairing and the associated regulation mechanisms require further investigation.

The gene expression characteristics in organisms can provide insights into their development and associated regulation mechanisms. Microarray technology is an important method of choice for providing high throughput data. However, a microarray can only detect the characteristics of known sequences. In contrast, next-generation sequencing is capable of acquiring novel information without prior knowledge of the target gene sequence (Teng and Xiao 2009). Next-generation sequencing technologies, such as the Solexa/Illumina genome analyzer and ABI/SOLiD gene sequencer, have made deep

impacts on genomic research (Morozova and Marra 2008; Wang et al. 2009). Digital gene expression (DGE) based on next-generation sequencing has three important advantages, including production of digital signals, strong detection capabilities, and ability to discover unknown transcripts (t'Hoen et al. 2008). Therefore, DGE has a wide range of applications, including whole-genome expression profile mapping, differential expression gene profile analysis, and gene functional research.

During the development of *S. japonicum*, males and females begin to pair about 18 days postinfection, and the female begins to lay eggs about 24 days postinfection (He and Yang 1980). In this experiment, the Illumina genome analyzer platform was used to perform DGE analysis of the transcriptomes of RNA samples extracted from 18- and 23-day-old female *S. japonicum* from double- and single-sex infections to investigate differential gene expression. Whereas previous studies were performed using microarray and, thus, yielded limited results, the DGE analysis applied in the present study provides a wider coverage of the entire transcriptome. By mapping our DGE tag data to a reference genomic sequence, novel transcriptional information was obtained that had not been detected by microarray analysis before.

Methods

Ethics statement

This study was carried out in strict accordance with the recommendations in the Regulations for the Administration of Affairs Concerning Experimental Animals of the State Science and Technology Commission. The protocol was approved by the Internal Review Board of Tongji University School of Medicine.

Unisexual and paired infections

Oncomelania hupensis snails were obtained from the Jiangsu Institute of Schistosome Diseases, Jiangsu Province, China. To obtain single-sex female worms, the snails were exposed to a single miracidium which generated from eggs acquired from the liver of infected rabbits or mice. Approximately 100 to 150 freshly shed cercariae were used to percutaneously infect each mouse. Schistosomula were recovered by perfusion within 18 and 23 days postinfection. The worms were washed in cold saline solution and checked by microscopy for possible undesirable mixed-sex infections. We separated single-sex female worms and froze them at −80 °C until further processing of the samples.

To obtain double-sex female worms, multiple cercariae freshly shed by snails were used to percutaneously infect mice

with 100 to 150 mixed-sex cercariae each. The mice were sacrificed 18 and 23 days postinfection. Females were recovered by washing with cold saline solution and then carefully separated from the paired worms under a microscope. All samples were frozen until further processing.

RNA extraction and amplification

Total RNA was extracted using TRIzol reagent (Invitrogen Life Technologies) according to the manufacturer's instructions. RNA concentration and purity were evaluated spectrophotometrically at 260 and 280 nm, respectively, using a NanoDrop ND1000 spectrophotometer and an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). RNA samples were stored at −80 °C.

Tag-seq library construction

After extraction, 6 mg of total RNA was used for RNA capture with magnetic oligo (dT) beads. First- and second-strand cDNA were synthesized. The 5′ ends of tags were generated by two types of endonuclease, namely, *Nla* III or *Dpn* II. The bead-bound cDNA was digested with *Nla* III, which cut off CATG sites. The fragments apart from the 3′ cDNA fragments connected to oligo(dT) beads were washed away. The Illumina adaptor 1 was ligated to the sticky 5′ end of digested bead-bound cDNA fragments. The junction of the Illumina adaptor 1 and the CATG site was the recognition site of *Mme* I, which cut 17 bp downstream of the CATG site to produce tags with adaptor 1. After removing 3′ fragments by magnetic bead precipitation, the Illumina adaptor 2 was ligated to the 3′ ends of the tags, yielding tags with different adaptors on both ends to form a tag library. After 15 cycles of linear PCR amplification, 105-bp fragments were purified by 6 % TBE PAGE gel electrophoresis. After denaturation, the single-chain molecules were fixed onto an Illumina Sequencing Chip (flow cell). Each molecule grew into a single-molecule cluster sequencing template through in situ amplification. Four types of nucleotides labeled by four colors were added to the template, and sequencing according to the sequencing by synthesis (SBS) method was performed. Each tunnel generated millions of raw reads with a sequencing length of 49 bp.

Data processing and identification of differentially expressed genes and pathways

Raw reads were filtered to obtain high quality data in the Tag-seq libraries by removing potentially erroneous tags. Briefly, the clean tags were obtained by trimming the 3′ adaptor sequence, filtering low-quality tags containing N, and removing small tags and one copy tag. All of the clean tags were mapped to the *S. japonicum* genome (http://www.chgc.sh.cn/japonicum/Resources.html) (predicted coding genes) and the *S. mansoni* genome (ftp://ftp.sanger.ac.uk/pub/pathogens/Schistosoma/mansoni/genome/gene_predictions/GeneDB_Smansoni_Genes.v4.0h.gz). Clean tags mapped to reference sequences from multiple genes were filtered, and the remaining clean tags were designed and annotated as unambiguous clean tags. The initial counts of the clean tags of each gene were normalized (transcripts per million) to obtain the normalized gene expression (Morrissy et al. 2009; t'Hoen et al. 2008). According to previous reports, a rigorous algorithm was adopted to screen genes that were differentially expressed between samples. The false discovery rate (FDR) determines the P value (corresponding to the P value in differential gene expression detection). The value of FDR is required to be less than 0.05 (Benjamini et al. 2001). In this study, a gene is considered differentially expressed when there is at least a onefold difference in expression between the two samples and the FDR is less than 0.001. During analysis of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway terms, we required P value to be less than 0.05. Differential pathways were ranked according to $-\mathrm{Lg}\, P$.

Quantitative RT-PCR analysis

Several differentially expressed genes were selected for validation using quantitative real-time (RT)-PCR. To ensure maximum specificity for PCR amplification of cDNAs under a standard set of reaction conditions, a stringent set of criteria was used for primer design, including predicted melting temperatures of 59 ± 2 °C, limited self-complementarity, a primer length ranging from 18 to 22 nt, and PCR amplicon lengths ranging from 150 to 250 bp. The first cDNA strand was synthesized from 500 ng total RNA using PrimeScript® RT reagent Kit (Perfect Real Time) (TaKaRa Code: DRR037A). Products were amplified using SYBR® Premix Ex Taq™ II (Perfect Real Time) (TaKaRa Code: DRR081A) in an ABI Prism 7300 sequence detection system (Applied Biosystems) at a final volume of 10 μL containing 1 μL of cDNA from the RT reaction, 5 μL of 2× SYBR Premix Ex Taq™ II (TaKaRa), and 200 nM forward and reverse primers with the following profile: 95 °C for 30 s and 40 cycles of 95 °C for 5 s and 60 °C for 31 s. Expression levels of *S. japonicum* PSMD4 (26S proteasome non-ATPase regulatory subunit 4) (accession no. FN320595) or trafficking protein particle complex subunit 2-like protein (accession no. FN319821) were used as endogenous controls within each sample (Liu et al. 2012). Relative levels of gene expression were calculated using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001). Each sample was analyzed for primer dimer, contamination, or mispriming by inspection of their dissociation curves.

Statistical analysis

Results are presented as mean ± standard deviation from at least three independent experiments. Statistical analyses were performed using one-way ANOVA and Student's *t* test. A value of $P < 0.05$ was considered statistically significant.

Results

Statistics of raw data

A summary of the number of DGE tags, as well as their mapping to the reference database, is shown in Table 1. In total, 5,643,424 (18-day-old female schistosomula from single-sex infections, 18SSI), 4,553,506 (23-day-old female schistosomula from single-sex infections, 23SSI), 5,640,472 (18-day-old female schistosomula from double-sex infections, 18DSI), and 4,512,651 (23-day-old female schistosomula from double-sex infections, 23DSI) clean tags were sequenced. The numbers of corresponding distinct tags were 88,762 (18SSI), 89,095 (23SSI), 93,455 (18DSI), and 50,139 (23DSI). In this study, the tag sequences of the four DGE libraries were mapped to the *S. japonicum* genome. For 18SSI, 22.63 % of the clean tags could be matched to the genes, and the number of tag-mapped genes reached 8,870. The corresponding values for 23SSI were 24.12 % and 9,794,

respectively; those for 18 DSI were 22.13 % and 8,885, respectively; and those for 23DSI were 30.67 % and 5,773, respectively. Unknown tags accounted for over 60 % of the total clean tags of the samples.

Sequencing saturation analysis

We performed sequencing saturation analysis to test whether or not the number of identified genes increased with increasing total tag number. As shown in Fig. 1a, when the total tag number of 18SSI reached 1 million, the increase in identified genes began to level out and stabilize when the number of tags reached 3 million. A similar trend can also be observed in 23SSI, 18DSI, and 23DSI (Fig. 1b, c, and d); here, identified genes began to level out when the total tag number reached 2 million and stabilized at 3 million. These findings suggest that distinct genes can no longer be identified when the total clean tag number reaches a certain value. Over 3 million clean tags were identified from the 18SSI, 18DSI, 23SSI, and 23DSI libraries, respectively, which indicates that the deep sequencing results were saturated and comprehensive.

Genes differentially expressed between samples

In the clean tags database, the number of tag copies reflects the quantitative level of gene expression, and its statistical distribution may indicate whether or not the data was normal

**Table 1** Summary of DGE profiles and their mapping to the reference genes

Summary		18SSI	23SSI	18DSI	23DSI
Clean tag[a]	Total number	5,643,424	4,553,506	5,640,472	4,512,651
	Distinct tag number[b]	88,762	89,095	93,455	50,139
All Tag Mapping to Gene	Total number	1,276,877	1,098,491	1,248,379	1,384,237
	Total % of clean tag	22.63 %	24.12 %	22.13 %	30.67 %
All tag mapping to gene	Distinct tag number	22,555	27,251	22,755	14,034
	Distinct tag % of clean tag	25.41 %	30.59 %	24.35 %	27.99 %
Unambiguous tag mapping to gene[c]	Total number	1,108,858	957,699	1,081,933	1,200,128
	Total % of clean tag	19.65 %	21.03 %	19.18 %	26.59 %
Unambiguous tag mapping to gene	Distinct tag number	20,854	25,226	20,987	12,918
	Distinct tag % of clean tag	23.49 %	28.31 %	22.46 %	25.76 %
All tag-mapped genes	Number	8,870	9,794	8,885	5,773
	% of ref genes	34.30 %	37.88 %	34.36 %	22.33 %
Unambiguous tag-mapped genes	Number	7,275	7,982	7,284	4,658
	% of ref genes	28.13 %	30.87 %	28.17 %	18.01 %
Unknown tag	Total number	4,366,547	3,455,015	4,392,093	3,128,414
	Total % of clean tag	77.37 %	75.88 %	77.87 %	69.33 %
Unknown tag	Distinct tag number	66,207	61,844	70,700	36,105
	Distinct tag % of clean tag	74.59 %	69.41 %	75.65 %	72.01 %

[a] Clean tags are tags obtained after filtering of dirty tags (low quality tags) from raw data

[b] Distinct tags are different kinds of tags

[c] Unambiguous tags are the remaining clean tags after removal of tags mapped to reference sequences from multiple genes

Fig. 1 Assessment of the degree of saturation in DGE sequencing. **a**, **b**, **c**, and **d** present the relationships between the percentage of genes identified and total tag number in the libraries of 18SSI, 23SSI, 18DSI, and 23DSI, respectively
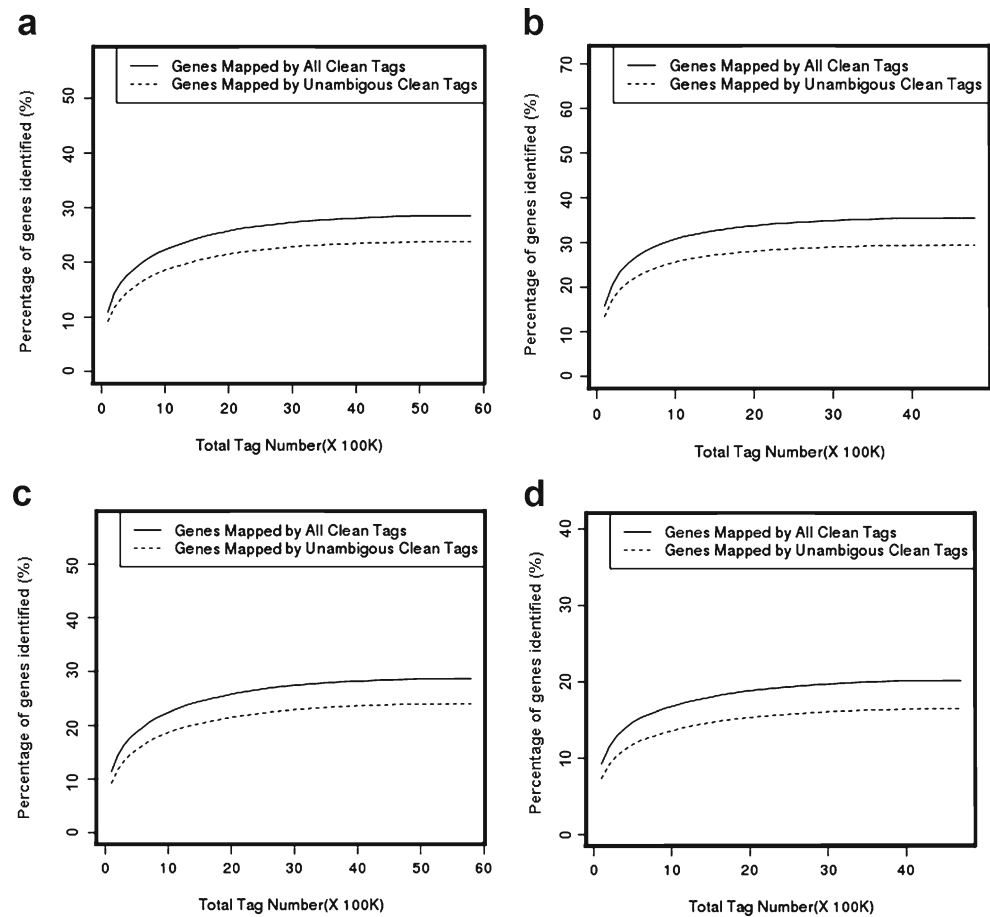

overall. We sequenced four DGE libraries, 18SSI, 23SSI, 18DSI, and 23DSI and analyzed the differential expression of genes between them. All annotated genes were analyzed for evidence of differential expression. As shown in Fig. 2a, a total of 7,005 genes were analyzed between 18DSI and 18SSI (Supplementary Table 1). Of these genes, approximately 318 genes were considered significant with an FDR of less than 0.001 and expression ratio (18DSI/18SSI) of over 1. From the 318 differentially expressed genes, 241 genes were upregulated in 18DSI, whereas only 77 genes were upregulated in 18SSI (Fig. 2a). Using the same method, 2,545 differentially expressed genes were detected between 23SSI and 18SSI. Of these, 2,193 genes were upregulated in 23SSI, whereas only 352 genes were upregulated in 18SSI (Fig. 2b). However, between 18DSI and 23DSI, the number of upregulated genes in 18DSI (1,258 genes) was higher than that in 23DSI (740 genes) (Fig. 2c), or close to the latter (not data shown). In particular, of the 3,446 differentially expressed genes between 23DSI and 23SSI, 2,913 genes were upregulated in 23SSI, whereas only 533 genes were upregulated in 23 DSI (Fig. 2d and Supplementary Table 2). Of the upregulated genes in 23 DSI, the top 30 genes with clear functions were ranked in Table 2. We found that phosphoglycerate mutase, super-oxide dismutase, major egg antigen, ribosomal protein, and

ferritin-1 heavy chain were upregulated in 23DSI (more details supplied in Supplementary Table 2). In contrast, of the upregulated genes in 23SSI, a large number of genes were involved in development, cell metabolism, glycometabolism, lipids metabolism, protein metabolism, material transport, and transduction compared with those in 23DSI (Fig. 3).

KEGG pathway analysis

KEGG indicates the main biochemical and metabolic process involved in differentially expressed genes. Differentially expressed genes between 18SSI and 18DSI were involved in 146 KEGG pathways. Five differential pathways were mainly involved in ribosome biogenesis, base excision repair, purine metabolism, and so on (Fig. 4a). Between 18SSI and 23SSI, 216 KEGG pathways were involved; here, 19 differential pathways included most signaling pathways, such as the Notch signaling pathway, the mitogen-activated protein kinases (MAPK) signaling pathway, the ErbB signaling pathway, the insulin signaling pathway, the Wnt signaling pathway, and the adipocytokine signaling pathway. In addition, several morphogenesis-associated pathways, such as axon guidance, dorsoventral axis formation, and regulation of actin cytoskeleton, participated in the 18SSI and 23DSI 23SSI developmental
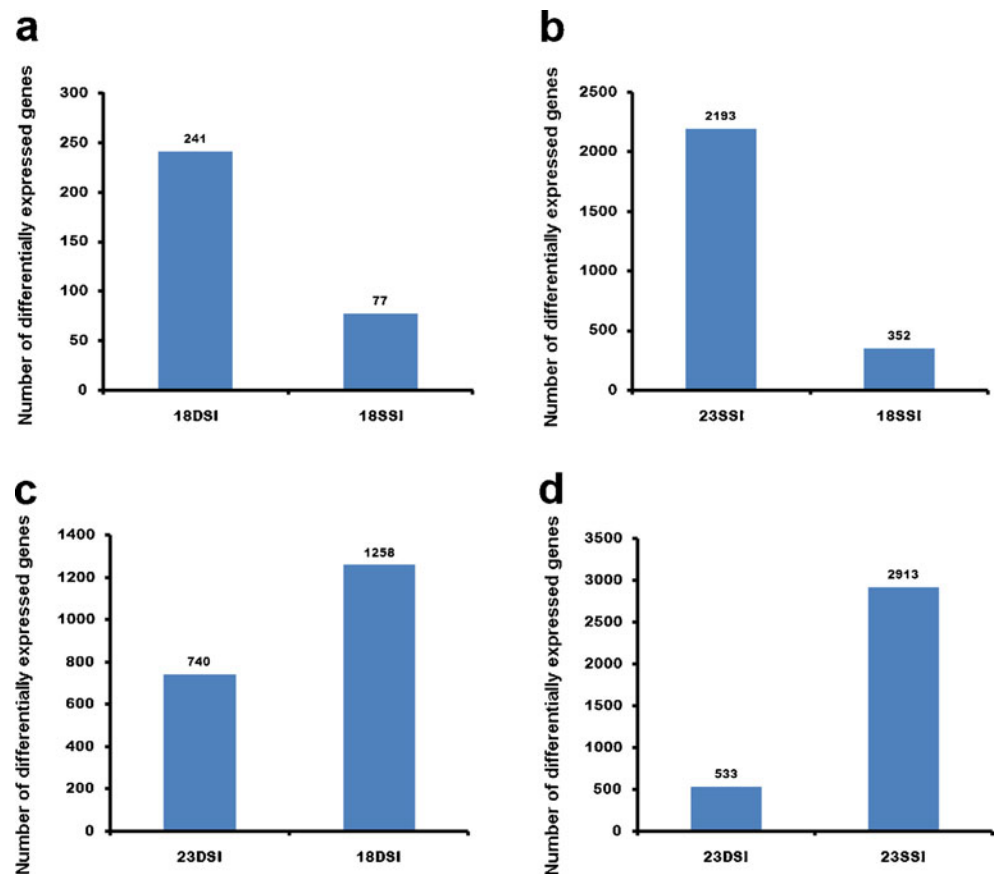
processes (Fig. 4b). Similarly, genes between 18DSI and 23DSI were involved in 210 pathways. However, of 12 significantly differential pathways, few signaling pathways were involved, except Wnt signaling pathway (Fig. 4c). Differences in pathways between 23SSI and 23DSI were similar to those between 23SSI and 18SSI, including the same pathways, such as cell adhesion molecules (CAMs), ribosome biogenesis in eukaryotes, the Notch signaling pathway, ECM–receptor interaction, regulation of actin cytoskeleton, endocytosis, glycosylphosphatidylinositol anchor biosynthesis, the MAPK signaling pathway, the ErbB signaling pathway, progesterone-mediated oocyte maturation, folate biosynthesis, inositol phosphate metabolism, focal adhesion, and axon guidance (Fig. 4d). Most genes of 23SSI in these pathways were upregulated compared with those of 18SSI, and most genes of 23DSI in these pathways were downregulated. Only a few genes of 23DSI were significantly upregulated in a few pathways, such as ribosome biogenesis, progesterone-mediated oocyte maturation, and folate biosynthesis, compared with those of 23SSI.

Confirmation of differentially expressed genes by quantitative RT-PCR analysis

To confirm the differentially expressed genes especially in 23DSI, seven 23DSI upregulated genes were selected at random for quantitative RT-PCR analysis. Expression of six

differentially expressed genes major egg antigen (p40) (CAX78215.1), dihydrofolate reductase (XP_002580557.1), ferritin-1 heavy chain (CAX77379.1), oocyte maturation-associated gene (AAW27816.1), and fatty acid-binding protein (AAG50052.1) fitted in well with the pattern of Solexa analysis (Fig. 5). The differential expression of thioredoxin peroxidase (CAX75860.1) in 18DSI, 23DSI, 18SSI, and 23SSI determined by quantitative RT-PCR analysis was not as typical as that by Solexa analysis. A few genes, such as ATP synthase (CAX76793.1), did not show consistent expression between qRT-PCR and Solexa analysis. Results showed mostly consistent expression between real-time PCR and Solexa analyses.

Discussion

A global analysis of transcriptomes will help facilitate identification of systemic gene expression and regulatory mechanisms for pairing. In this study, transcriptome profiling was performed to identify genes differentially expressed in 18SSI, 23SSI, 18DSI, and 23DSI and understand the mechanism by which pairing promotes the development of females. This analysis was completed using Solexa sequencing technology. The present study represents the most comprehensive analyses of transcriptomes postpairing to date. Three million clean tags

Table 2 Top 30 upregulated annotated genes between the 23DSI and 23SSI libraries based on log2 ratio (23DSI/23SSI)

Gene ID	log2 ratio (23DSI/23SSI)	Functional annotation
CAX76329.1	14.75295552	Phosphoglycerate mutase
CAX71782.1	12.44708323	Extracellular superoxide dismutase [Cu–Zn] precursor
CAX70600.1	10.2772874	Asparagine-rich antigen Pfa35-2
XP_002569863.1	8.717676423	Methyltransferase related
AAX30103.1	8.471675214	Dynein light chain 2
XP_002581192.1	8.276124405	Amine GPCR
CAX78232.1	7.930737338	Major egg antigen (p40)
CAX77261.1	7.930737338	Carbonyl reductase 1
CAX69701.1	4.554588852	Quinoid dihyropteridine reductase
CAX76793.1	4.132487462	ATP synthase
CAX72276.1	3.980745443	Stress protein DDR48 (DNA damage-responsive protein 48)
CAX76919.1	3.95924452	Ribosomal protein S25
CAX72121.1	3.791291346	Ribosomal protein S14
CAX77379.1	3.683946431	Ferritin-1 heavy chain
XP_002578635.1	3.619519723	Histone H3
CAX76497.1	3.612786454	Small nuclear ribonucleoprotein Sm D1
CAX78924.1	3.584822689	Ribosomal protein LP1
XP_002577967.1	3.471305719	Helicase
XP_002572241.1	3.45711868	Polyadenylate binding protein
CAX76083.1	3.423554582	Eukaryotic translation initiation factor 2
CAX76993.1	3.398007987	Lesswright
AAP06111.1	3.320740529	Similar to GenBank accession number AB049644 mitochondrial ribosomal protein L20 (L20mt) in *Homo sapiens*
CAX77896.1	3.306792386	Dolichyl-diphosphooligosaccharide–protein glycosyltransferase subunit DAD1
CAX83693.1	3.273666533	Gap-Pol polyprotein
CAX75173.1	3.265842677	Putative H/ACA ribonucleoprotein complex subunit 1-like protein
CAX71543.1	3.166321117	Ribosomal protein L27
XP_002573711.1	3.098989095	Thyrotropin-releasing hormone receptor
CAX71749.1	3.084875506	Small subunit ribosomal protein S30e
CAX77982.1	3.076786688	Ribosomal protein L24
XP_002575364.1	3.035852383	Centromere protein A (cenp-A) (centromere autoantigen A)

indicated that the deep sequencing results were saturated and comprehensive. However, only 5,000 to 10,000 tag-mapped genes were identified for each library. A large number of unknown tags, approximating 69 to 78 % of the total percentage of clean tags, could not be matched with *S. japonicum* genomic sequences. These findings indicate that a large

number of novel genes require further identification and suggest that the present sequencing of the *S. japonicum* genome is still incomplete. The complete set of genomic sequences for *S. japonicum* may also be unavailable. As such, our present analysis presents several limitations.

Pairing of *S. japonicum* initiates a cascade of events in female worms that ultimately lead to female development and egg production. To identify genes associated with this important biological process, we studied parasites isolated from single- and mixed-sex cercariae-infected mice using Solexa analysis to uncover pair-regulated transcriptional profiles. According to the growth and development of *S. japonicum* in hosts, the male and female worms begin to pair about 15–18 days postinfection. The pairs then produce the first batch of eggs about 22–24 d postinfection (He and Yang 1980). Thus, the transcriptional profile of 18-day-old worms mainly showed the features of genes expression before pairing or during the early stage of pairing, whereas the profile of 23-day-old worms revealed the features of gene expression of newly developed worms after pairing. Comparing 18- and 23-day-old females, which just developed mature reproductive organs, was expected to obtain specific molecular information at the early stage of sexual maturation. As such, our research differs from previous studies that focus on the differential genes expression between adult male and adult female (Gobert et al. 2006; Hoffmann et al. 2002; Moertel et al. 2008).

We found that the transcriptional profile of 18SSI is similar to that of 18DSI. Few differentially expressed genes and pathways were observed between these groups, which is in accordance with the fact that both groups are at an early stage of development before pairing. By contrast, a large number of differentially expressed genes between 18- and 23-day-old females from both single-sex and double-sex infections were found. In particular, of over 2,500 differential genes between 18SSI and 23SSI, most were significantly upregulated in 23SSI and involved in 216 KEGG pathways. Many signaling pathways, such as the notch signaling pathway, the MAPK signaling pathway, and the insulin signaling pathway, the wnt signaling pathway and the chemokine signaling pathway were involved in these KEGG pathways. Furthermore, most of the genes participating in these pathways, such as ribosome biogenesis, axon guidance, CAMs, focal adhesion, basal transcription factors, progesterone-mediated oocyte maturation, steroid biosynthesis, endocytosis, regulation of actin cytoskeleton, fatty acid elongation in mitochondria, protein digestion and absorption, and terpenoid backbone biosynthesis, were significantly upregulated in 23SSI compared with genes in 18SSI. The result indicated that gene expression of 23-day-old females from single-sex infection seemed to not be confined, so that they upregulated almost all genes (2,193 of 2,545 differentially expressed genes between 23SSI and 18SSI) to support its growth and development.
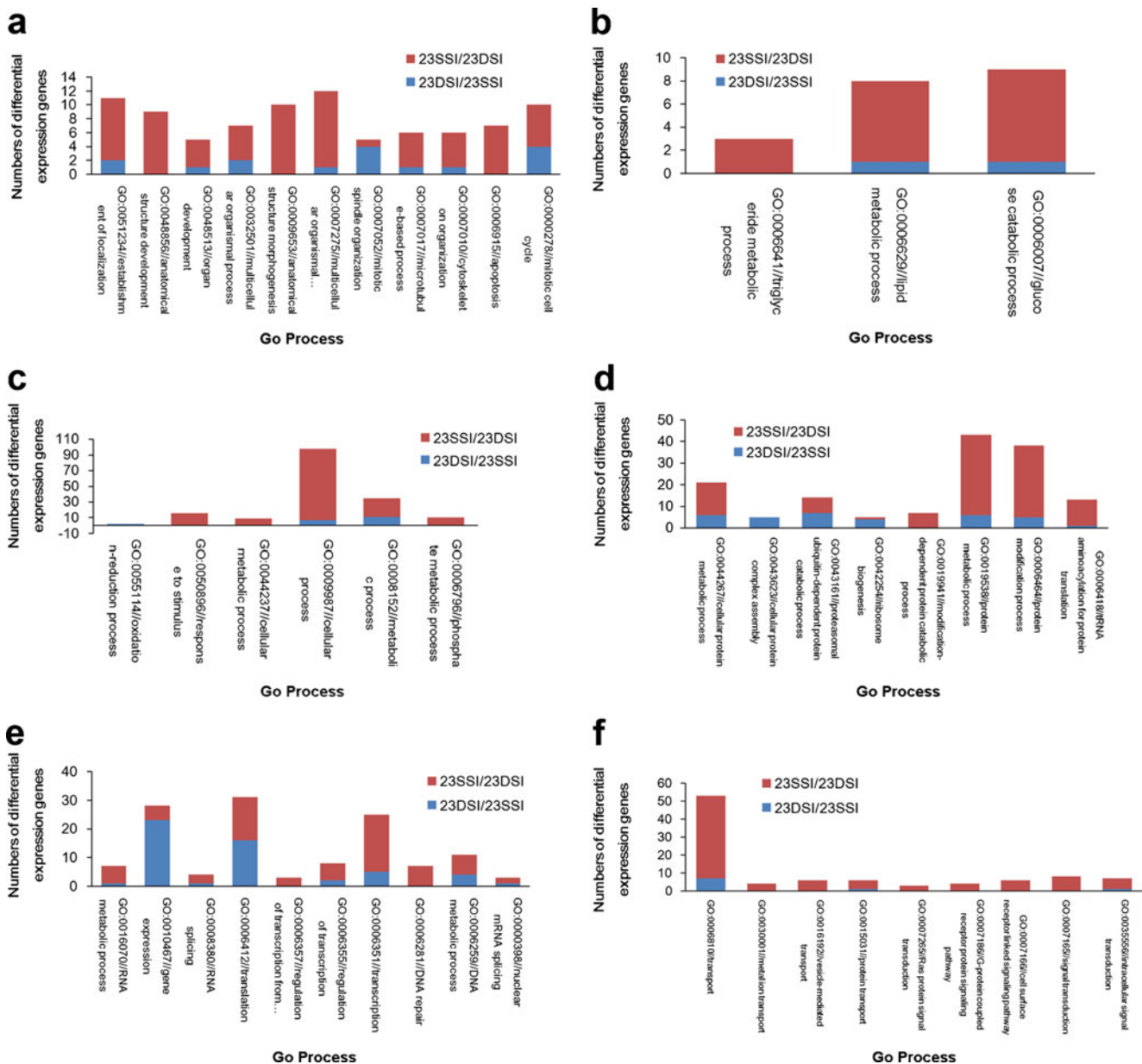
Fig. 3 Gene ontology analysis of the biological processes of differentially expressed genes between 23SSI and 23DSI showed that 23SSI dominates in morphogenesis, glucose, and lipid metabolic process, cellular metabolic process, protein metabolic process, regulation of DNA and RNA, transport and signal transduction. **a** morphogenesis, **b** glucose and lipid metabolic process, **c** cellular metabolic process, **d** protein metabolic process, **e** regulation of DNA and RNA, **f** transport and signal transduction. *23SSI/23DSI* the number of upregulated genes in 23SSI compared with 23DSI, *23DSI/23SSI* the number of upregulated genes in 23DSI compared with 23SSI

The same pattern of gene expression was not observed between 18DSI and 23DSI. First, only 740 of the 1,998 differentially expressed genes were upregulated in 23DSI compared with those in 18SSI. Second, differentially expressed genes between 18DSI and 23DSI were involved in pathways different from those between 18SSI and 23SSI. In particular, few signaling pathways participated in these pathways, except for the wnt signaling pathway. The 740 differentially expressed genes in 23DSI were involved in a limited number of pathways, such as DNA replication, base excision

repair, porphyrin and chlorophyll metabolism, spliceosome, axon guidance, cell cycle, proteasome, oocyte meiosis, pyrimidine metabolism, N-glycan biosynthesis, purine metabolism, citrate cycle, antioxidant processes, pyruvate metabolism, ribosome, glycolysis/gluconeogenesis, and oxidative phosphorylation.

Compared with the developed and larger 23-day-old females from double-sex infection, far more upregulated genes (2,913 of 3,446 differentially expressed genes) were found in undeveloped and small 23-day-old females from single-sex
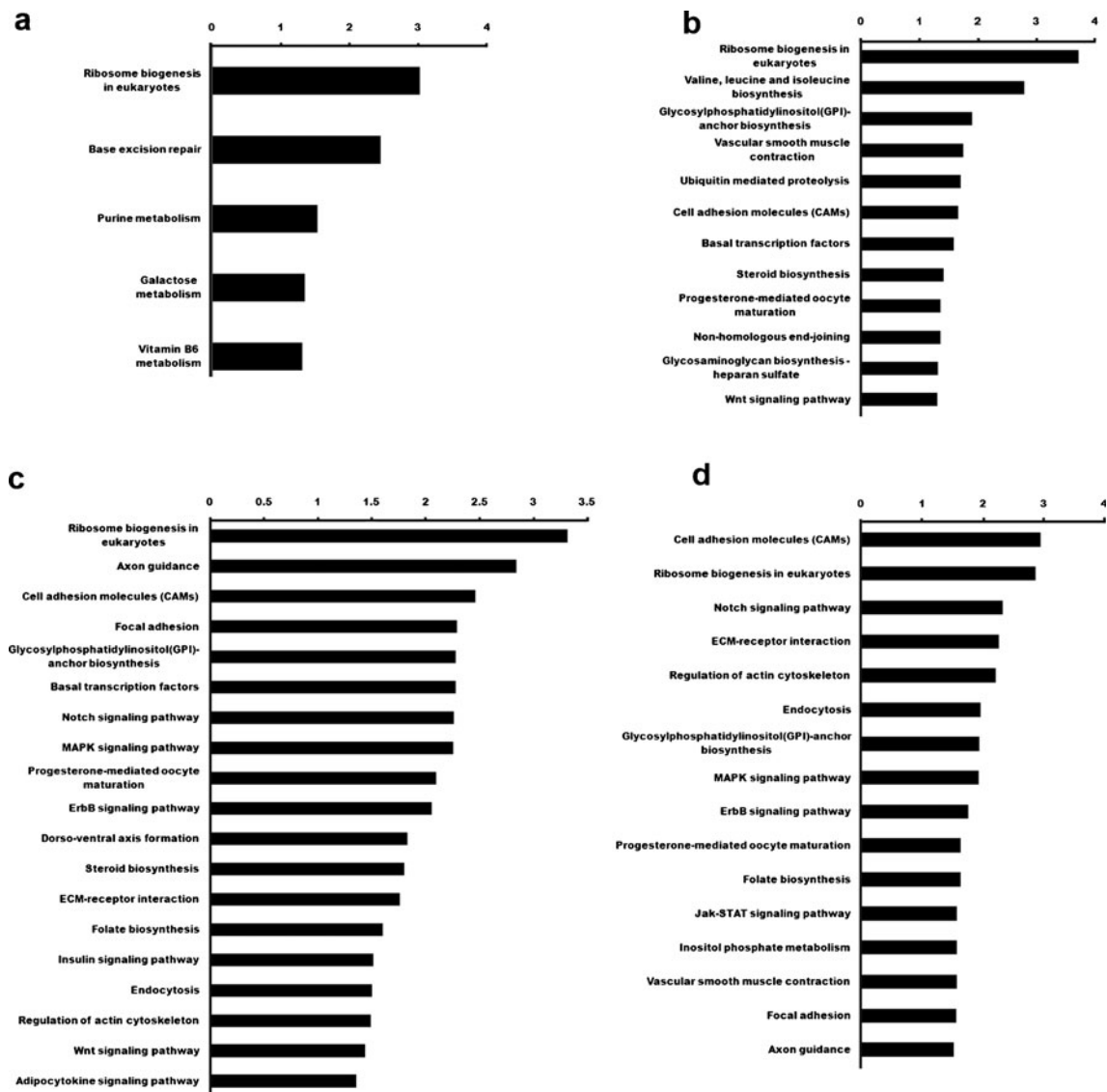
Fig. 4 Comparison of differential KEGG pathways between samples. **a** Differential KEGG pathways between 18SSI and 18DSI, **b** between 23SSI and 18SSI, **c** between 23DSI and 18DSI, **d** between 23SSI and 23DSI

infection. We found that upregulated genes in 23SSI dominate in anatomical structure morphogenesis, organ development, cytoskeleton organization, response to stimulus, lipid metabolic process, glucose catabolic process, protein metabolic process, protein modification process, RNA and DNA metabolic process, transcription, transport, and signal transduction. Although there were similar patterns of differential KEGG pathways between 23DSI and 23SSI compared to 23SSI and 18SSI, actually most differentially expressed genes between 23DSI and 23SSI in these pathways were upregulated in 23SSI, rather than in 23DSI. Why then did 23-day-old developed worms express fewer genes than 23-day-old undeveloped worms? It is possible that the developed worms may require the expression of specific genes, rather than universal upregulation of all gene expressions. Our results show that genes upregulated in 23SSI worms are involved in

morphogenesis, organ development, and cytoskeleton organization. These functions focus on growth rather than reproduction. By contrast, genes such as phosphoglycerate mutase, superoxide dismutase, egg antigen, ribosomal protein, ferritin-1 heavy chain, and eukaryotic translation initiation factor 2 were significantly upregulated in 23DSI. These genes function in glycolysis, antioxidant defense, protein biosynthesis, egg formation, iron transport and utilization, and translation regulation, which may be necessary for egg production. Previous research compared differential gene expressions between females and males and showed similar effects. Egg antigen (Chen et al. 1992; Fitzpatrick and Hoffmann 2006; Hoffmann et al. 2002), ferritin-1 (Fitzpatrick and Hoffmann 2006; Fitzpatrick et al. 2004; Grevelding et al. 1997; Hoffmann et al. 2002), ribosomal proteins (Fitzpatrick and Hoffmann 2006; Hoffmann et al.

Fig. 5 Quantitative RT-PCR validation of genes recognized as upregulated genes by Solexa analysis in 23DSI. **a** Differential expression of major egg antigen (p40) (CAX78215.1) in 23DSI, 23SSI, 18DSI, and 18SSI was analyzed by quantitative RT-PCR and Solexa, **b** egg maturity-associated gene (AAW27816.1), **c** thioredoxin peroxidase (CAX75860.1), **d** ferritin-1 heavy chain (CAX77379.1), **e** fatty acid-binding protein (AAG50052.1), **f** dihydrofolate reductase (XP_002580557.1). *PCR* Quantitative RT-PCR method, *DGE* Solexa analysis method
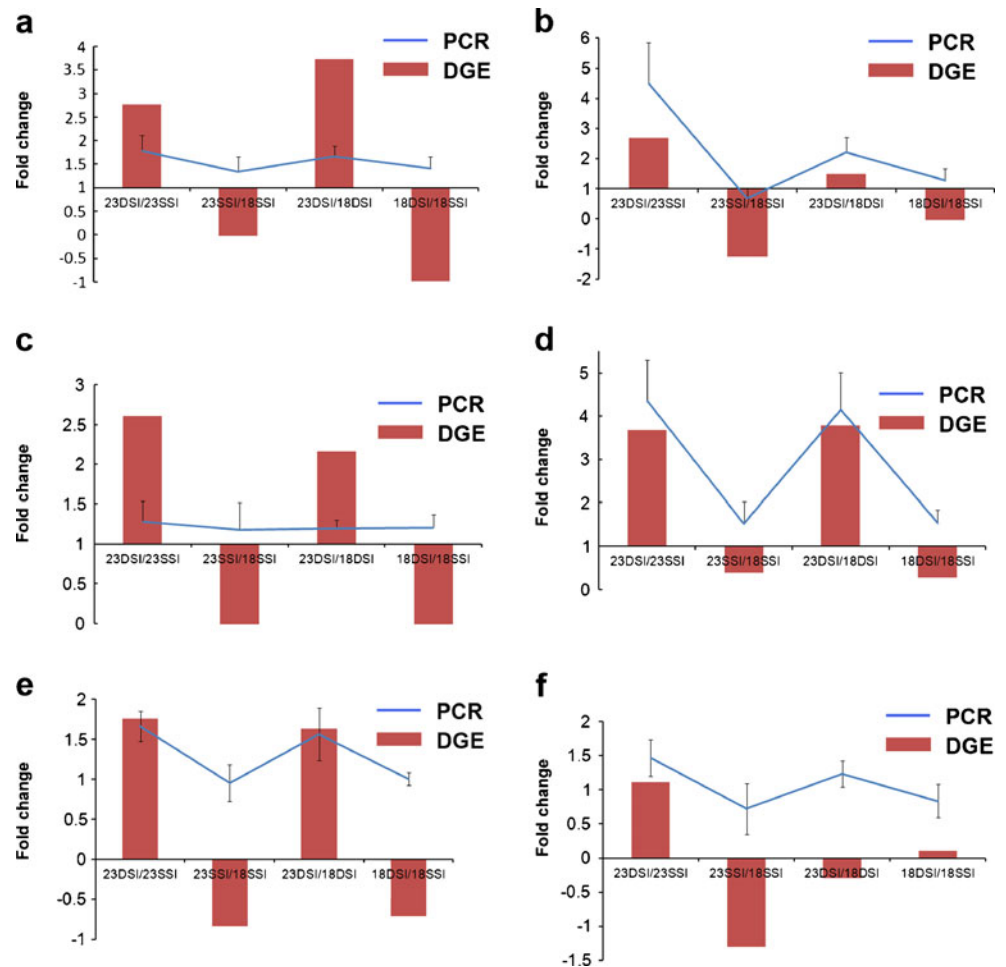
2002; Waisberg et al. 2007; Farias et al. 2011), ATPase (Moertel et al. 2008; Waisberg et al. 2007), cathepsin (Moertel et al. 2008; Farias et al. 2011), extracellular superoxide dismutase (Fitzpatrick and Hoffmann 2006; Fitzpatrick et al. 2004; Waisberg et al. 2007; Williams et al. 2007), cytochrome *C* oxidase (Fitzpatrick and Hoffmann 2006), tyrosinase (Fitzpatrick et al. 2004; Williams et al. 2007), mucin-like protein (Menrath et al. 1995), fs800 (Reis et al. 1989; Williams et al. 2007), and adenylosuccinate lyase (Fitzpatrick and Hoffmann 2006) were often detected in females. However, between 23DSI and 23SSI, not all the above-mentioned genes were upregulated in 23DSI. For instance, cathepsins were downregulated in 23DSI, compared with 23SSI (Supplementary Table 2). Differential expressions of tyrosinase and mucin-like protein were not detected between 23DSI and 23SSI (Supplementary Table 2). Not all genes associated with oviposition were expressed in 23DSI. This result suggests that although sexual maturation had been achieved, 23DSI worms still remained in a status before egg production. Previous studies compared the differential expression between immature females and mature egg-laying females (Waisberg et al. 2007; Fitzpatrick and Hoffmann

2006), providing evidences that: (1) more genes associated with carbohydrate and protein metabolism were expressed in unpaired females rather than paired females, suggesting that males help paired females to feed in order to allow females to focus on egg generation and (2) more genes related to egg generation and antioxidants were expressed in paired females. Actually, there are many genes known to be associated with the mature status of egg-laying females, but not all have been directly related to pairing. Besides, after pairing, it is likely that genes related to female maturation may not all be expressed at the same time. Although pairing governs gene expression associated with oviposition, there are additional tasks after pairing and before egg laying, which have to be performed. Our study provided new insights into the biological processes at this stage.

The limited number of upregulated genes in 23DSI suggests that pairing does not actually regulate all facets of transcription but is still focused on sexual development, as indicated in previous reports (Fitzpatrick and Hoffmann 2006; Waisberg et al. 2007). The differentially expressed genes and their functions observed in our research indicate the major function of egg production and establishment of relevant

conditions. Eggs, especially eggshells, have been reported to store large amounts of iron (Glanfield et al. 2007; Jones et al. 2007). Ferritin-1 is responsible for iron transport and is stored in the vitelline glands (Ford et al. 1984; Schussler et al. 1995). Our recent results revealed that only after worms have paired, they start to produce large amounts of hemozoin in the gut lumen, accumulate iron in the vitelline glands, and upregulate ferritin-1 (Sun et al. 2013). In addition, superoxide dismutase can protect the worm from heme and free radicals during hemoglobin decomposition to safely obtain enough iron for egg production. Upregulation of egg antigen proteins and ribosomal protein levels also facilitates egg production. Thus, after pairing, 23DSI worms downregulate the expression of a large number of irrelevant genes (2,913 of 3,446 differentially expressed genes between 23DSI and 23SSI), whereas they especially upregulate the expression of genes closely associated with egg production and other relevant biological processes, but not of all genes known to be expressed in mature females. These results strongly suggest that pairing promotes and maintains sexual maturation and oviposition by the targeted expression of a small number of specific genes.

References

Armstrong JC (1965) Mating behavior and development of schistosomes in the mouse. J Parasitol 51:605–616

Basch PF (1990) Why do schistosomes have separate sexes? Parasitol Today 6:160–163

Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I (2001) Controlling the false discovery rate in behavior genetics research. Behav Brain Res 125:279–284

Chen LL, Rekosh DM, LoVerde PT (1992) Schistosoma mansoni p48 eggshell protein gene: characterization, developmentally regulated expression and comparison to the p14 eggshell protein gene. Mol Biochem Parasit 52:39–52

Clough ER (1981) Morphology and reproductive organs and oogenesis in bisexual and unisexual transplants of mature Schistosoma mansoni females. J Parasitol 67:535–539

Dillon GP, Feltwell T, Skelton JP, Ashton PD, Coulson PS, Quail MA, Nikolaidou-Katsaridou N, Wilson RA, Ivens AC (2006) Microarray analysis identifies genes preferentially expressed in the lung schistosomulum of Schistosoma mansoni. Int J Parasitol 36:1–8

Farias LP, Tararam CA, Miyasato PA, Nishiyama MY Jr, Oliveira KC, Kawano T, Verjovski-Almeida S, Leite LC (2011) Screening the Schistosoma mansoni transcriptome for genes differentially expressed in the schistosomulum stage in search for vaccine candidates. Parasitol Res 108:123–135

Fitzpatrick JM, Hoffmann KF (2006) Dioecious Schistosoma mansoni express divergent gene repertoires regulated by pairing. Int J Parasitol 36:1081–1089

Fitzpatrick JM, Johansen MV, Johnston DA, Dunne DW, Hoffmann KF (2004) Gender-associated gene expression in two related strains of Schistosoma japonicum. Mol Biochem Parasit 136:191–209

Ford GC, Harrison PM, Rice DW, Smith JM, Treffry A, White JL, Yariv J (1984) Ferritin: design and formation of an iron-storage molecule. Philos Trans Roy Soc ser B Biol Sci 304:551–565

Glanfield A, McManus DP, Anderson GJ, Jones MK (2007) Pumping iron: a potential target for novel therapeutics against schistosomes. Trends Parasitol 23:583–588

Gobert GN, McInnes R, Moertel L, Nelson C, Jones MK, Hu W, McManus DP (2006) Transcriptomics tool for the human Schistosoma blood flukes using microarray gene expression profiling. Exp Parasitol 114:160–172

Grevelding CG, Sommer G, Kunz W (1997) Female-specific gene expression in Schistosoma mansoni is regulated by pairing. Parasitology 115:635–640

Gupta BC, Basch PF (1987) The role of Schistosoma mansoni males in feeding and development of female worms. J Parasitol 73:481–486

Haseeb MA (1998) Schistosoma mansoni: females enhance [14C]-tyrosine incorporation in males maintained in vitro. J Helminthol 72:123–126

He YX, Yang HZ (1980) Physiological studies on the post-cercarial development of Schistosoma japonicum. ACTA Zoologica Sinica 26:32–41

Hoffmann KF, Johnston DA, Dunne DW (2002) Identification of Schistosoma mansoni gender-associated gene transcripts by cDNA microarray profiling. Genome Biol 3:research0041.1–research0041.12

Jones MK, McManus DP, Sivadorai P, Glanfield A, Moertel L, Belli SI, Gobert GN (2007) Tracking the fate of iron in early development of human blood flukes. Int J Biochem Cell B 39:1646–1658

Lei ZL, Zheng H, Zhang LJ, Zhu R, Guo JG, Li SZ, Wang LY, Chen Z, Zhou XN (2010) Schistosomiasis status in People's Republic of China in 2010. Chin J Schisto Control 23:599–604

Liu S, Cai P, Hou N, Piao X, Wang H, Hung T, Chen Q (2012) Genome-wide identification and characterization of a panel of house-keeping genes in Schistosoma japonicum. Mol Biochem Parasit 182:75–82

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(−delta delta C(T)) method. Methods 25:402–408

Menrath M, Michel A, Kunz W (1995) A female-specific cDNA sequence of Schistosoma mansoni encoding a mucin-like protein that is expressed in the epithelial cells of the reproductive duct. Parasitology 111:477–483

Moertel L, McManus DP, Piva TJ, Young L, McInnes RL, Gobert GN (2006) Oligonucleotide microarray analysis of strain- and gender-associated gene expression in the human blood fluke, Schistosoma japonicum. Mol Cell Probe 20:280–289

Moertel L, Gobert GN, McManus DP (2008) Comparative real-time PCR and enzyme analysis of selected gender-associated molecules in Schistosoma japonicum. Parasitology 135:575–583

Morozova O, Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. Genomics 92:255–264

Morrissy AS, Morin RD, Delaney A, Zeng T, McDonald H, Jones S, Zhao Y, Hirst M, Marra MA (2009) Next-generation tag sequencing for cancer gene expression profiling. Genome Res 19:1825–1835

Peng HJ, Chen XG, Wang XZ, Lun ZR (2003) Analysis of the gene expression profile of Schistosoma japonicum cercariae by a strategy based on expressed sequence tags. Parasitol Res 90:287–93

Pereira AS, Cavalcanti NL, Nascimento GA, Nascimento-Silva JL, Padilha RJ, Viegas LF, Alves LC, Lima-Filho JL, Chaves ME (2013) Morphological and morphometric study of cercariae and adult worms of Schistosoma mansoni (SLM strain) isolated from infected mice. Parasitol Res 112:1087–96

Popiel I, Basch PF (1984) Reproductive development of female Schistosoma mansoni (Digenea: Schistosomatidae) following bisexual pairing of worms and worm segments. J Exp Zool 232:141–150

Reis MG, Kuhns J, Blanton R, Davis AH (1989) Localization and pattern of expression of a female specific mRNA in Schistosoma mansoni. Mol Biochem Parasit 32:113–119

Schussler P, Potters E, Winnen R, Bottke W, Kunz W (1995) An isoform of ferritin as a component of protein yolk platelets in Schistosoma mansoni. Mol Reprod Dev 41:325–330

Shaw JR (1977) Schistosoma mansoni: pairing in vitro and development of females from single-sex infections. Exp Parasitol 41:54–65

Shaw MK (1987) Schistosoma mansoni: vitelline gland development in females from single sex infections. J Helminthol 61:253–9

Shaw JR, Marshall I, Erasmus DA (1977) Schistosoma mansoni: in vitro stimulation of vitelline cell development by extracts of male worms. Exp Parasitol 42:14–20

Steinmann P, Keiser J, Bos R, Tanner M, Utzinger J (2006) Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. Lancet Infect Dis 6:411–25

Sun J, Hu W, Li C (2013) Beyond heme detoxification: a role for hemozoin in iron transport in S. japonicum. Parasitol Res 112:2983–90

t'Hoen PA, Ariyurek Y, Thygesen HH, Vreugdenhil E, Vossen RH, de Menezes RX, Boer JM, van Ommen GJ, den Dunnen JT (2008) Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. Nucleic Acids Res 36:e141

Teng X, Xiao H (2009) Perspectives of DNA microarray and next-generation DNA sequencing technologies. Sci China C Life Sci 52:7–16

Vermeire JJ, Taft AS, Hoffmann KF, Fitzpatrick JM, Yoshino TP (2006) Schistosoma mansoni: DNA microarray gene expression profiling during the miracidium-to-mother sporocyst transformation. Mol Biochem Parasit 147:39–47

Waisberg M, Lobo FP, Cerqueira GC, Passos LK, Carvalho OS, Franco GR, El-Sayed NM (2007) Microarray analysis of gene expression induced by sexual contact in Schistosoma mansoni. BMC genomics 8:181

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63

Williams DL, Sayed AA, Bernier J, Birkeland SR, Cipriano MJ, Papa AR, McArthur AG, Taft A, Vermeire JJ, Yoshino TP (2007) Profiling Schistosoma mansoni development using serial analysis of gene expression (SAGE). Exp Parasitol 117:246–58