

# UCSF

## UC San Francisco Previously Published Works

### Title

A Comparative Chemogenomics Strategy to Predict Potential Drug Targets in the Metazoan Pathogen, *Schistosoma mansoni*

### Permalink

<https://escholarship.org/uc/item/9xq672ck>

### Journal

PLOS ONE, 4(2)

### ISSN

1932-6203

### Authors

Caffrey, Conor R  
Rohwer, Andreas  
Oellien, Frank  
[et al.](#)

### Publication Date

2009

### DOI

10.1371/journal.pone.0004413

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

# A Comparative Chemogenomics Strategy to Predict Potential Drug Targets in the Metazoan Pathogen, *Schistosoma mansoni*

Conor R. Caffrey<sup>1</sup>, Andreas Rohwer<sup>2</sup>, Frank Oellien<sup>2</sup>, Richard J. Marhöfer<sup>2</sup>, Simon Braschi<sup>1</sup>, Guilherme Oliveira<sup>3</sup>, James H. McKerrow<sup>1</sup>, Paul M. Selzer<sup>2\*</sup>

**1** Sandler Center for Basic Research in Parasitic Diseases, California Institute for Quantitative Biosciences, University of California San Francisco, San Francisco, California, United States of America, **2** Intervet Innovation GmbH, BioChemInformatics, Schwabenheim, Germany, **3** Laboratory of Cellular and Molecular Parasitology, Centro de Pesquisas René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Brazil

## Abstract

Schistosomiasis is a prevalent and chronic helminthic disease in tropical regions. Treatment and control relies on chemotherapy with just one drug, praziquantel and this reliance is of concern should clinically relevant drug resistance emerge and spread. Therefore, to identify potential target proteins for new avenues of drug discovery we have taken a comparative chemogenomics approach utilizing the putative proteome of *Schistosoma mansoni* compared to the proteomes of two model organisms, the nematode, *Caenorhabditis elegans* and the fruitfly, *Drosophila melanogaster*. Using the genome comparison software Genlight, two separate *in silico* workflows were implemented to derive a set of parasite proteins for which gene disruption of the orthologs in both the model organisms yielded deleterious phenotypes (e.g., lethal, impairment of motility), *i.e.*, are essential genes/proteins. Of the 67 and 68 sequences generated for each workflow, 63 were identical in both sets, leading to a final set of 72 parasite proteins. All but one of these were expressed in the relevant developmental stages of the parasite infecting humans. Subsequent in depth manual curation of the combined workflow output revealed 57 candidate proteins. Scrutiny of these for 'druggable' protein homologs in the literature identified 35 *S. mansoni* sequences, 18 of which were homologous to proteins with 3D structures including co-crystallized ligands that will allow further structure-based drug design studies. The comparative chemogenomics strategy presented generates a tractable set of *S. mansoni* proteins for experimental validation as drug targets against this insidious human pathogen.

**Citation:** Caffrey CR, Rohwer A, Oellien F, Marhöfer RJ, Braschi S, et al. (2009) A Comparative Chemogenomics Strategy to Predict Potential Drug Targets in the Metazoan Pathogen, *Schistosoma mansoni*. PLoS ONE 4(2): e4413. doi:10.1371/journal.pone.0004413

**Editor:** Jennifer Keiser, Swiss Tropical Institute, Switzerland

**Received:** October 29, 2008; **Accepted:** December 15, 2008; **Published:** February 9, 2009

**Copyright:** © 2009 Caffrey et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** Supported by the Sandler Foundation. GO receives funding from the United States NIH (TW007012) and FAPEMG (5323-4.01/07). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: Paul.Selzer@sp.intervet.com

## Introduction

Schistosomiasis is a parasitic disease infecting over 200 million people [1]. Considered a 'neglected tropical disease (NTD)' [2] for which, traditionally, there has been little in the way of a concerted drug discovery program, three major species of the flatworm parasite are responsible for disease in sub-Saharan Africa (*Schistosoma mansoni*, *S. haematobium*), South America (*S. mansoni*) and parts of China and South-East Asia (*S. japonicum*) [1]. Pathology associated with schistosomiasis mansoni and japonica results primarily from the accumulation of parasite eggs over the course of years and even decades giving rise initially to hepatomegaly that may be superseded by extensive liver fibrosis and possibly sequelae such as occlusion of the hepatic portal vein, portal hypertension, and gastrointestinal varices [3]. Furthermore, chronic schistosomiasis haematobia is a risk factor for squamous cell carcinoma of the bladder [4]. The disease is also known for its more subtle, and indeed underestimated [5] morbid effects, particularly in school-aged children. These include physical and cognitive under-performance, anemia and abdominal discomfort.

Despite ongoing attempts to produce a molecular vaccine [6,7], present treatment and control of schistosomiasis relies on chemotherapy [8]. Just one drug, praziquantel (PZQ), for which the detailed mode of action is still unclear [9,10], is widely available. Since its introduction in the late 1970's, PZQ has become the sole, WHO-recommended treatment, being safe, effective and affordable [11]. PZQ's success as a drug has contributed to a lack of urgency and investment in identifying new therapies, either in terms of chemical entities or molecular targets. This over-reliance on a single therapy to treat large populations is a serious concern regarding the potential for drug resistance [12,13]. Resistance to PZQ has been bred on more than one occasion in the laboratory [14] and foci of transient drug resistance have been reported in the literature [15]. Thus, with the WHO-backed goal to more widely disseminate PZQ (the Schistosomiasis Control Initiative [16]), it may just be a matter of time before clinically relevant drug resistance emerges [8].

The tenuousness of therapeutic options for schistosomiasis, together with better knowledge of the molecular and biochemical idiosyncrasies of the parasite, and improved genome sequence

information, are spurring increased investment in target discovery and validation [8,17]. Further, a completely annotated *Schistosoma mansoni* genome should, in the future, provide a rich source of information for both academia and non-profit interests to identify, prioritize and prosecute drug and vaccine targets. In advance of this milestone, sufficient characterization and annotation of the genome has already taken place [18] so that in the latest Version 4 of *Schistosoma mansoni* GeneDB [19] the prediction of genes, open reading frames and translation products has been accomplished.

Given the wealth of organized data to hand, therefore, we felt it timely to put this information to work in an *in silico* comparative genomics strategy to identify a subset of schistosome genes/proteins that have potential value as drug targets in order to jump-start focused discovery efforts. Our approach was to mine the proteomes of the model organisms *Drosophila melanogaster* and *Caenorhabditis elegans* for proteins with clear sequence similarities to those in the parasite in order to identify those experimentally proven as essential, i.e., targeted gene disruption produces deleterious phenotypes (e.g. lethal, paralyzed, impaired of motility) in both model organisms. Precedence has shown that even for parasite proteins that share significant sequence similarity with vertebrate proteins, anti-parasite drugs can, nevertheless, be developed (e.g.  $\beta$ -tubulin, the target protein of benzimidazoles) [20]. Accordingly, the 13,283 predicted gene products of *S. mansoni* were compared in a semi-automatic process to the proteomes and phenotypic databases of *D. melanogaster* and *C. elegans* using the software Genlight [21,22]. The output of 72 potential target proteins was manually curated leading to the identification of 35 *S. mansoni* proteins with druggable characteristics. Of these, 18 belong to protein families for which extensive 3D structural information is available, including bound small molecule ligands and drugs. Such structural data makes these proteins particularly suitable for prioritization of structure-based drug design strategies.

## Results

### Semi-automatic *in silico* workflows identify 72 candidate *S. mansoni* gene products

For the first *in silico* workflow, orthologs shared between the predicted proteome of *S. mansoni* [19], and the proteomes of *C. elegans* and *D. melanogaster* available at Wormbase [23] and Flybase [24], respectively, were determined (Figure 1A). For *S. mansoni* and *C. elegans*, and *S. mansoni* and *D. melanogaster*, 1778 and 1927 orthologs were identified, respectively. A subsequent comparison of both outputs demonstrated that 1258 sequences were identical. By reciprocal blastp [25,26], this set was then compared with the phenotype databases of *C. elegans* (Caltech server) [23] and *D. melanogaster* [24], and those orthologs displaying the appropriate, pre-set, mutant phenotypes identified. The *S. mansoni* sequences of both ortholog sets were pooled and only those proteins with a 100% sequence identity were considered as potential target proteins. Altogether, 68 *S. mansoni* sequences with orthologs in *C. elegans* and *D. melanogaster* were identified.

To evaluate the performance and results of the first *in silico* approach, a second workflow was generated (Figure 1B). First, orthologs shared between *C. elegans* and *D. melanogaster* were identified. The different numbers of orthologs (2933 in *C. elegans* compared to 3789 in *D. melanogaster*) are due to the high occurrence of identical gene copies with different identifiers in Flybase. Thus, certain *C. elegans* genes generate multiple hits in *D. melanogaster*. The orthologs from both model organisms were then separately compared with the *S. mansoni* putative proteome resulting in 1502 orthologs shared between *C. elegans* and *S. mansoni*, and 1520 orthologs between *D. melanogaster* and *S. mansoni*. A redundancy

check of these individual outputs established that 1302 sequences were identical. These 1302 sequences were then pooled with the extra non-redundant sequences ( $1502 - 1302 = 200$  and  $1520 - 1302 = 218$  from the *C. elegans* and *D. melanogaster* comparisons, respectively) to give a total of 1720 *S. mansoni* sequences. By reciprocal blastp and using the pre-set, mutant phenotype criteria, orthologs of these sequences in the phenotype databases of *C. elegans* and *D. melanogaster* were determined. Comparison of these outputs identified 67 potential *S. mansoni* target proteins that have orthologs in both model organisms. Finally, the sequence outputs from both *in silico* workflows (68 and 67 sequences, respectively) were compared in order to determine the extent to which they were identical (Figure 1C). Sixty-three proteins were shared, thus demonstrating the reliability of the software and the compatibility of the workflows. Overall, by adding the 63 shared proteins to the 5 and 4 extra sequences exclusive to the first and second workflows, respectively, 72 *S. mansoni* proteins were considered for further manual filtering in order to delineate potential drug targets.

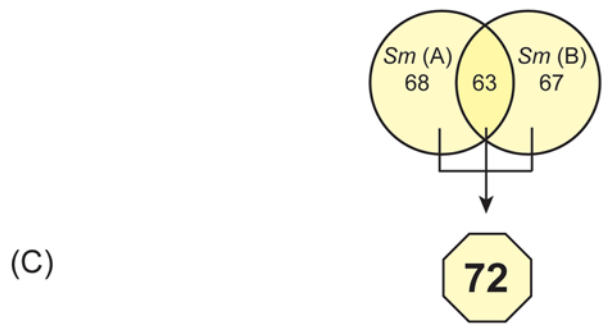
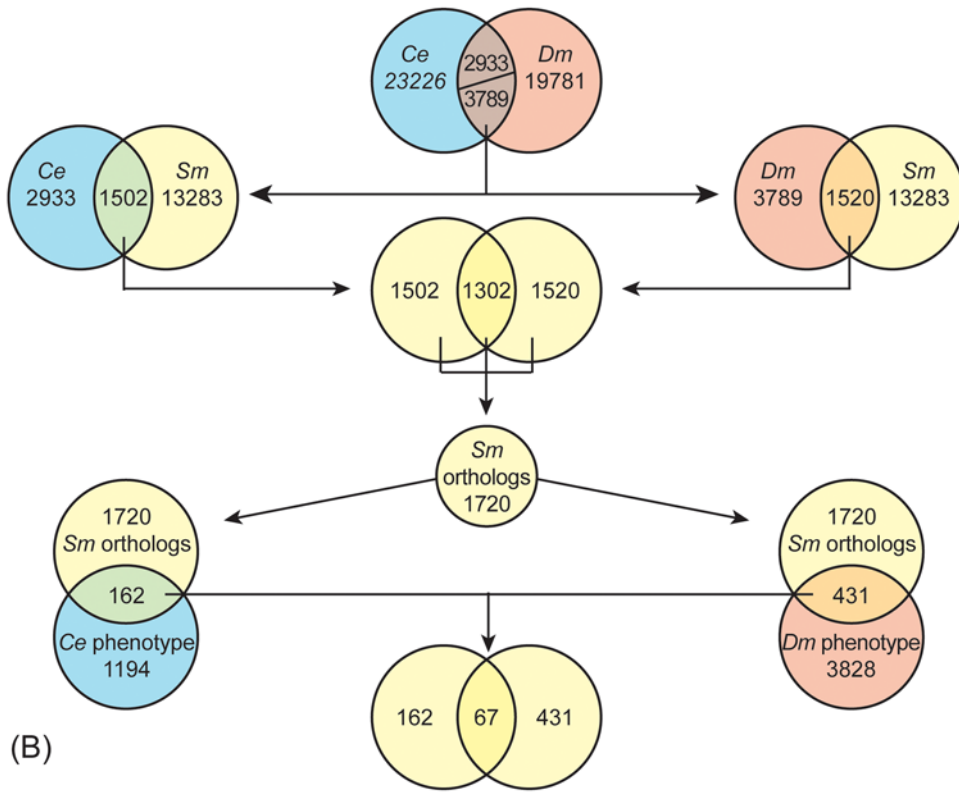
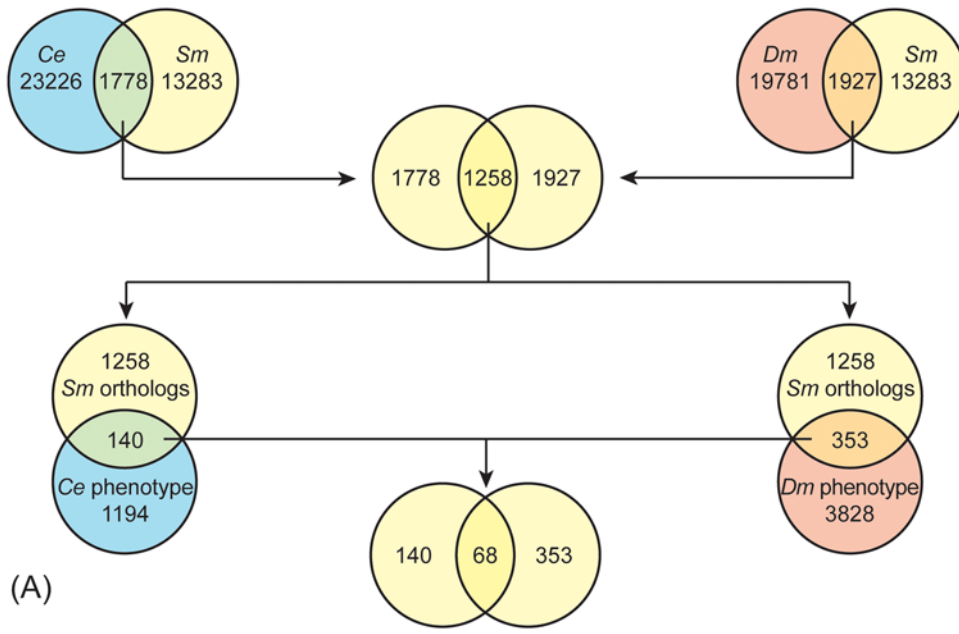
### Manual curation identifies 35 potential drug targets in *S. mansoni*

Manual scrutiny of the 72 sequences generated by the semi-automatic workflows was considered essential in order to remove possible redundant information and improve overall confidence in the results (Figure 1, Table 1). Each curation step is represented by a separate worksheet in Table S1.

First, duplicates of four sequence entries (Smp\_062300 ↔ Smp\_062300.2, Smp\_103470.3 ↔ Smp\_103470.4, Smp\_120700.1 ↔ Smp\_120700.2, and Smp\_138970.1 ↔ Smp\_138970.4) were removed to leave 68 sequences. Subsequently, three sequences (Smp\_028990.1, Smp\_124240, and Smp\_138970.1), not confirmed as the definitive schistosome orthologs after reciprocal blastp with both Wormbase and Flybase, were removed to produce 65 sequences. Next, seven of these sequences for which deleterious phenotypes (e.g. lethal, paralyzed, movement abnormal, etc.) could not be confirmed for the respective orthologs or for which the relevant orthologous allele or phenotype information was simply unavailable were removed. Upon scrutiny of the EST evidence (see footnote to Table S1), one additional sequence (Smp\_154270) was also removed because it is only expressed in the miracidium – a developmental stage not found within the human host. The remaining 57 sequences are expressed in the relevant parasite life-stages that persist in humans, namely, the immature schistosomulum form, either prepared *in vitro* or removed from a mammalian host (*in vivo*), adult (male and/or female) and egg.

Next, each of the 57 sequences was assessed for druggability that is defined as the likelihood of being able to modulate a target's activity with a small-molecule drug [27,28]. We mined a number of biological and literature databases to document whether orthologous proteins or proteins of the same family have been reported to be manipulated by ligands, inhibitors or even targeted by known drugs. For 35 of the *S. mansoni* proteins we found unambiguous evidence that homologous proteins are druggable (Table 1, Table S1). Accordingly, these might be prioritized as high value drug targets for new treatments of schistosomiasis.

Finally, we searched among the 35 *S. mansoni* proteins for homologous proteins that have a 3D structure and/or a 3D structure complexed with a ligand, inhibitor or drug. Such structural information would enhance the druggability value by facilitating a structure-based drug design strategy, including homology modeling, docking, virtual screening or pharmacophore-based screening [29,30]. Eighteen of the 35 *S. mansoni* proteins fulfilled these conditions (Tables 1 and 2) and another 8 had at least partial 3D structure information available (Table S1).



**Figure 1. In silico workflows to identify putative drug target proteins in *S. mansoni* based on sequence and phenotype comparisons.** A and B, representations of two independent workflows leading to a similar number of potential targets. C, the combination of workflows A and B generating a final number of 72 sequences (octagon) of which 63 were identical. Numbers of sequences used in each step are indicated within the respective circles. Depending on the intersection, the numbers within represent either sequence orthologs or *S. mansoni* proteins for which a deleterious phenotype is recorded in either Wormbase or Flybase. Blue, red and yellow circles display sequences from *C. elegans* (*Ce*), *D. melanogaster* (*Dm*), and *S. mansoni* (*Sm*), respectively. Details of the workflows are described in the text. doi:10.1371/journal.pone.0004413.g001

## Discussion

The comparative chemogenomics strategy described herein provides a prioritized and testable list of potential target proteins for *S. mansoni*, a metazoan pathogen causing chronic and debilitating disease in humans. The intent was first to mine the *S. mansoni* genome and identify putative essential genes based on similarity to experimentally-determined essential genes/proteins in two model metazoans and then define a subset of potential drug targets for which structural information of known target proteins, including bound ligands, exist. Both the strategy and outputs are in keeping with the recent establishment of a TDR Drug Targets Prioritization Database [31] by the World Health Organization's Special Programme for Research and Training in Tropical Diseases that facilitates the identification and prioritization of target genes in a number of pathogenic organisms, including those responsible for NTDs.

A number of factors were considered during the development and execution of the workflows in order to provide both confidence in the data generated and a solid platform from which to predict the druggability of individual *S. mansoni* proteins. First, rather than comparison with one metazoan proteome we incorporated two into the analysis, particularly as both *C. elegans* and *D. melanogaster* are phylogenetically remote from *Schistosoma*. Secondly, we only considered those orthologous genes/proteins for which phenotypes were generated via targeted gene mutagenesis. We discounted orthologs with phenotypes arising from RNAi (RNA interference) due to the potential for non-specific, off-target effects and false positive results including with *Drosophila* cells [32,33,34]. Nevertheless, the present protocol can be adapted to include RNAi phenotypes should one wish to cast the net wider. As a final stricture in the analysis, we selected for severely deleterious phenotypes such as death or those involving motility disorders with the aim of producing both a short list of testable targets and enriching for phenotypes that should be obvious when targeting the respective schistosome genes by chemical and/or genetic means (see below). We would state that our strategy focuses

on potential targets based on loss-of-function and that gain-of-function targets such as those involved in drug agonism e.g., ion channels, will be missed. As many current anthelmintics, including PZQ, are agonists, it would be worthwhile developing systems for gain-of-function mutants in model organisms (for example using targeted overexpression banks).

The present strategy differs from previous comparative genomic screens for infectious diseases and for which the goal was to identify potential drug targets unique to the pathogen based on a user-prescribed similarity cut-off value in order to decrease the potential for toxicity of any new chemotherapeutic [35,36]. Though this 'exclusion' approach is sound, nevertheless, examples abound of pathogen proteins that possess considerable similarity to human proteins and yet are valid drug targets e.g., parasite cysteine proteases [37,38,39], and the remarkable example of the highly conserved  $\beta$ -tubulin protein, the target of benzimidazoles [20], even though they might have failed a user-prescribed similarity cut-off as part of an exclusion genomics screen [35]. In addition, it is often the case that drug selectivity for a pathogen arises due to idiosyncrasies in pathogen physiology such as slower turnover of the target protein allowing for more pronounced drug action, e.g., ornithine decarboxylase in *Trypanosoma brucei*, the causative agent of African Sleeping Sickness [40], differences in protein regulation, e.g., dihydrofolate reductase in *Plasmodium falciparum* [41] or a lack of functional redundancy compared to the host [42]. Thus, we consider that the present similarity-based strategy to mine the *S. mansoni* genome represents a useful contribution to the consideration of new anti-schistosomal therapeutic targets.

An important aspect of developing drugs to NTDs such as schistosomiasis and for which profits are nil or marginal at best, is the need to keep costs of drug development to a minimum. Experience has shown that drug development programs for pathogen-specific targets of often unknown function, although a valid scientifically, are, in many cases, more expensive. Our strategy, therefore, to identify homologous rather than pathogen-specific potential targets is intended to reduce costs by leveraging the biochemical, chemical and structural data and tools already available for therapeutic targets of proven value in other clinical contexts, i.e., 'piggy-back' drug discovery [43,44,45]. The finding that 35 of the *in silico*-identified 72 sequences are indeed homologous to known druggable targets supports our strategy. Moreover, 18 of these belong to protein families for which 3D structural information including ligands or even drugs is available (e.g., N-(4-Methoxybenzyl)-N'-(5-Nitro-1,3-Thiazol-2-Yl)Urea and Carbidopa). This opens the route not only for classical biochemical studies but also for structure-based drug discovery approaches [29,46,47]. For instance, among the 18 candidate proteins, small molecule scaffolds exist for methionine aminopeptidase targeting cancer [48] and malaria [49], N-myristoyl transferase against fungal infections [50] and Rac-GTPase against cancer [51]. The availability of specific small molecules adds incentive to prioritizing the experimental validation of the orthologous schistosome proteins.

Unlike for both model organisms, standardized functional genomic tools, such as targeted gene disruption or gene 'knock-

**Table 1. Automatic and manual filtering for potential target proteins**

Filter	Number of <i>S. mansoni</i> proteins
Comparative Genomics	72
Manual Curation*	57
Druggable Targets based on surveys of literature and biological databases	35
3D Structure with co-crystallized ligand	18

\*Removal of sequences that were redundant, not confirmed as the ortholog by reciprocal blastp, for which a phenotype was not confirmed or not expressed in the relevant life stages that persist in the human host. Numbers correspond to those on the tabs in the Microsoft Excel-worksheets provided in Table S1. doi:10.1371/journal.pone.0004413.t001

**Table 2.** Druggable target proteins belonging to protein families with 3D structural information, including co-crystallized ligands

<i>S. mansoni</i> Putative Protein	Molecular Function GO Annotation at <i>Schistosoma mansoni</i> GeneDB [19]	GeneDB Accession Number
GTP-binding protein	GTP binding, signal transducer activity	Smp_005790
Glycogen synthase kinase 3 related	ATP binding, protein kinase activity	Smp_008260.1
Methionine amino peptidase	Metallo exopeptidase activity	Smp_011120
Calmodulin dependent protein kinase II	Protein kinase activity	Smp_011660.2
Protein phosphatase-2a	Hydrolase activity	Smp_030710
Nuclear transport factor	Transporter activity	Smp_037700
Vesicular-fusion protein nsf	ATP-binding, nucleoside triphosphatase activity, nucleotide binding	Smp_057320
Rac GTPase	GTP-binding	Smp_062300
Elongation factor tu	GTP-binding, translation elongation factor activity	Smp_073500.1
Neuroendocrine convertase	Subtilase activity	Smp_077980
Myosin heavy chain	Motor activity	Smp_085540.6
Nucleoside diphosphate kinase	ATP-binding, nucleoside diphosphate kinase activity	Smp_092750
Rab GDB-dissociation inhibitor	Rab GDP-dissociation inhibitor activity	Smp_094420
Heat shock protein 70	ATP-binding	Smp_106130.2
N-myristoyl transferase	Glycylpeptide N-tetradecanoyltransferase activity	Smp_121420
Choline o-acyltransferase	Acyltransferase activity	Smp_146910
Rab 6	GTP-binding	Smp_163580
Amino acid decarboxylase	Carboxy lyase activity	Smp_171580

Additional information such as key references reviewing druggability are available in Supplementary Table S1.  
doi:10.1371/journal.pone.0004413.t002

in' technology, are not yet established for schistosomes. This represents a stumbling block for interrogating gene function with confidence. However, transient RNAi with double stranded RNAi or small interfering (si)RNA has gained a foothold as a useful technique for gene knock down (if not knock out), including in those developmental stages relevant to infection and pathology in humans (somules, adults and eggs) [52,53]. We would assume, therefore, that transient RNAi will be an important tool to experimentally establish gene essentiality until more rigorous reverse genetic techniques become established. Where available, a complementary strategy to RNAi would be to employ a chemical genetics approach using protein-selective small molecules (see examples above). Such an approach has already been validated regarding schistosome cysteine proteases [37] and enzymes involved in redox metabolism [54].

We have mined and compared the predicted proteome of the metazoan pathogen, *S. mansoni*, with those of two well-studied model organisms in order to identify potential drug targets. The chemogenomics strategy has produced a tractable list of prioritized genes for further investigation, one or more of which might contribute to badly-needed drug discovery programs for this prevalent human disease.

## Methods

### Datasets

All protein datasets are available on Flybase (version FB2006\_01) [24], Wormbase release 195 [23], and *Schistosoma mansoni* GeneDB v4.0 [19]. The *D. melanogaster* phenotype database was generated using the Flybase QueryBuilder. All proteins with the mutant phenotypes “lethal” and/or “neurophysiology defective” were downloaded. Similarly, the *C. elegans* phenotype database was generated using the Caltech server based on Wormbase release 177 [23]. All proteins with the mutant

phenotypes “lethal”, “paralyzed”, “movement abnormal” and/or “muscle system physiology abnormal” were downloaded. However, only data from knock-out mutants (alleles) were used; RNAi phenotypes were excluded due to the potential for non-specific effects [32,33,34].

### Comparative Genomics using Genlight

All genome comparisons performed were based on translated genomes using the software Genlight [21], for which a public WWW-server is available at the University of Bielefeld, Germany [22]. Genlight is a client/server based program suite developed for large scale sequence analyses and comparative genomics calculations. A key functionality of Genlight is the determination of orthologs via reversed or reciprocal blast searches. Sequences from organism A are compared with those from organism B and *vice versa* using the respective blast program [25,26]. Orthologous sequences are then defined as those best hit sequences that find each other in such bidirectional blast searches. The sequence alignment overlap and the E-value cut off can be preset in accordance with the goal of the experiment [21,29]. Within Genlight, we defined orthologs as best reciprocal blastp hits with a minimum of 70% sequence alignment overlap and an E-value of 0.01 or smaller.

A useful feature of the Genlight software is its user friendliness in the post-processing of results including the employment of predefined filters and operations. For example, result sets can be re-used directly to exclude duplicates in order to generate non-redundant datasets. This feature has been employed as described in the two *in silico* workflows (Figure 1) and resulted in the identification of 72 potential target proteins in the *S. mansoni* predicted proteome (Table S1) for which deleterious phenotypes were identified for the respective orthologs in both *C. elegans* and *D. melanogaster*.

## Manual curation and filtering for potential target proteins

After executing two separate electronic workflows (see Figure 1), we manually curated the schistosome protein sequence output for consistency with our goal of identifying potential drug targets. First, redundant sequences were removed from the dataset. Further curation included removal of sequences (i) not confirmed as orthologs by reciprocal blastp searches with both Wormbase and Flybase, (ii) for which targeted gene disruption of the appropriate ortholog does not yield a deleterious phenotype, and (iii) which, based on the available EST evidence, are not expressed in the relevant parasite developmental stages infecting humans (i.e., schistosomulum (immature worm) either prepared *in vitro* or removed from a mammalian host (*in vivo*), adult (male and/or female) and egg). Consideration was also given to the potential druggability of the parasite protein by manually mining biological and literature databases on the internet including DrugBank [55,56], UniProt [57,58], Prosite [59,60], OMIM [61,62], InterPro [63,64] and Pfam [65,66]. Druggability was defined as the likelihood of being able to modulate the activity of the protein target with a small-molecule drug [27,28]. Finally, the PDB database [67] was searched for 3D structural information of proteins homologous to each parasite protein, including co-crystallized ligands. The 72 sequences, together with the subsequent manual curation steps, are presented in Table S1.

## Supporting Information

**Table S1** This table shows the output and manual curation of potential *S. mansoni* drug targets generated by comparative genomics with *Caenorhabditis elegans* and *Drosophila melanogaster*. The table comprises 7 worksheets each with a different level of information for the respective sequence lists. Leftmost in the table, the first worksheet contains the 72 sequences generated from the two separate electronic workflows utilizing the genome comparison software, Genlight. Each worksheet thereafter represents a subsequent manual curation step involving the removal of sequences (i) that are redundant (remaining 68 non-redundant entries), (ii) that are not confirmed as orthologs by reciprocal blastp searches with both Wormbase and Flybase (remaining 65 confirmed orthologs), (iii) for which targeted gene disruption of

the appropriate ortholog does not yield a deleterious phenotype as declared in both Wormbase and Flybase (remaining 58 confirmed phenotypes), (iv) that are not expressed in the relevant developmental stages of the parasite infecting humans (remaining 57 expressed in the relevant life stage; for an explanation of how this was performed see below), (v) that are not druggable as indicated by manual mining of the biological and literature databases available on the internet (remaining 35 druggable targets), and (vi) for which homologous proteins with 3D structural information (including co-crystallized ligands) was not found (remaining 18 proteins with both a 3D structure and co-crystallized ligand). To determine in which developmental stage a putative protein is expressed, the following procedure was performed. In the *S. mansoni* genome browser [68], the sequence name (e.g., Smp\_000040) is input into the text field, “landmark or region”. On the returned search page, the graphic produced by the latest GeneDB working model and displaying the organization of the gene is clicked to reveal a variety of information, including exon/intron junctions, physicochemical characteristics of the putative protein and gene ontology. Activating the hyperlink “DNA” reveals the unspliced DNA, spliced DNA and amino acid sequences. Next, a blastn search at NCBI of the spliced sequence is performed via the hyperlink “Send to BLAST at NCBI”. The analysis is constrained using the search set ‘non-human, non-mouse ESTs (EST\_others)’, the organism ID number of 6183 for *S. mansoni* (taxid:6183) and the program selection set to ‘somewhat similar sequences’. On the EST list returned, each accession is scrutinized for its life-stage origin by activating the relevant link. The lowest maximum score accepted for consideration was 250. Lower scores tended to be too short to be reliably ascribed to the gene under study.

Found at: doi:10.1371/journal.pone.0004413.s001 (0.26 MB XLS)

## Author Contributions

Conceived and designed the experiments: CC PMS. Performed the experiments: AR FO RJM PMS. Analyzed the data: CC AR FO RJM SB GO PMS. Contributed reagents/materials/analysis tools: SB GO. Wrote the paper: CC AR JHM PMS.

## References

- 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: 411–425.
- Hotez PJ, Molyneux DH, Fenwick A, Kumaresan J, Sachs SE, et al. (2007) Control of neglected tropical diseases. *N Engl J Med* 357: 1018–1027.
- Gryseels B, Polman K, Clerinx J, Kestens L (2006) Human schistosomiasis. *Lancet* 368: 1106–1118.
- Herrera LA, Benitez-Bribiesca L, Mohar A, Ostrosky-Wegman P (2005) Role of infectious diseases in human carcinogenesis. *Environ Mol Mutagen* 45: 284–303.
- King CH, Dickman K, Tisch DJ (2005) Reassessment of the cost of chronic helminthic infection: a meta-analysis of disability-related outcomes in endemic schistosomiasis. *Lancet* 365: 1561–1569.
- Loukas A, Tran M, Pearson MS (2007) Schistosome membrane proteins as vaccines. *Int J Parasitol* 37: 257–263.
- Wilson RA, Coulson PS (2006) Schistosome vaccines: a critical appraisal. *Mem Inst Oswaldo Cruz* 101 Suppl 1: 13–20.
- Caffrey CR (2007) Chemotherapy of schistosomiasis: present and future. *Current Opinion in Chemical Biology* 11: 433–439.
- Angelucci F, Basso A, Bellelli A, Brunori M, Pica Mattoccia L, et al. (2007) The anti-schistosomal drug praziquantel is an adenosine antagonist. *Parasitology* 134: 1215–1221.
- Greenberg RM (2005) Ca<sup>2+</sup> signalling, voltage-gated Ca<sup>2+</sup> channels and praziquantel in flatworm neuromusculature. *Parasitology* 131 Suppl: S97–108.
- Doenhoff MJ, Pica-Mattoccia L (2006) Praziquantel for the treatment of schistosomiasis: its use for control in areas with endemic disease and prospects for drug resistance. *Expert Rev Anti Infect Ther* 4: 199–210.
- Cioli D, Pica-Mattoccia L (2003) Praziquantel. *Parasitol Res* 90 Suppl 1: S3–9.
- Utzinger J, Xiao SH, Tanner M, Keiser J (2007) Artemisinins for schistosomiasis and beyond. *Curr Opin Investig Drugs* 8: 105–116.
- Fallon PG, Doenhoff MJ (1994) Drug-resistant schistosomiasis: resistance to praziquantel and oxamniquine induced in *Schistosoma mansoni* in mice is drug specific. *Am J Trop Med Hyg* 51: 83–88.
- Botros S, Bennett JL (2007) Praziquantel resistance. *Expert Opinion in Drug Discovery* 2: S35–S40.
- http://www.schisto.org/ Schistosomiasis Control Initiative.
- Keiser J, Utzinger J (2007) Advances in the discovery and development of trematocidal drugs. *Expert Opinion in Drug Discovery* 2: S9–S23.
- Haas BJ, Berriman M, Hirai H, Cerqueira GG, Loverde PT, et al. (2007) *Schistosoma mansoni* genome: closing in on a final gene set. *Exp Parasitol* 117: 225–228.
- http://www.genedb.org/genedb/smansoni/ *Schistosoma mansoni* GeneDB.
- Robinson MW, McFerran N, Trudgett A, Hoey L, Fairweather I (2004) A possible model of benzimidazole binding to beta-tubulin disclosed by invoking an inter-domain movement. *J Mol Graph Model* 23: 275–284.
- Beckstette M, Mailänder JT, Marhöfer RJ, Sczyrba A, Ohlebusch E, Giegerich R, Selzer PM (2004) Genlight: interactive high-throughput sequence analysis and comparative genomics. *Journal of Integrative Bioinformatics* 1.
- http://piranha.techfak.uni-bielefeld.de/ Genlight.
- http://www.wormbase.org/ Wormbase.
- http://flybase.bio.indiana.edu/ Flybase: A Database of *Drosophila* Genes & Genomes.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410.

26. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402.
27. Cheng AC, Coleman RG, Smyth KT, Cao Q, Soulard P, et al. (2007) Structure-based maximal affinity model predicts small-molecule druggability. *Nat Biotechnol* 25: 71–75.
28. Keller TH, Pichota A, Yin Z (2006) A practical view of 'druggability'. *Curr Opin Chem Biol* 10: 357–361.
29. Krasky A, Rohwer A, Schroeder J, Selzer PM (2007) A combined bioinformatics and cheminformatics approach for the development of new antiparasitic drugs. *Genomics* 89: 36–43.
30. Oellien F, Cramer J, Beyer C, Ihlenfeldt WD, Selzer PM (2006) The impact of tautomer forms on pharmacophore-based virtual screening. *J Chem Inf Model* 46: 2342–2354.
31. <http://tdrtargets.org/> TDR Drug Targets Prioritization Database.
32. Echeverri CJ, Perrimon N (2006) High-throughput RNAi screening in cultured cells: a user's guide. *Nat Rev Genet* 7: 373–384.
33. Jackson AL, Bartz SR, Schelzer J, Kobayashi SV, Burchard J, et al. (2003) Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol* 21: 635–637.
34. Moffat J, Reiling JH, Sabatini DM (2007) Off-target effects associated with long dsRNAs in *Drosophila* RNAi screens. *Trends Pharmacol Sci* 28: 149–151.
35. Luscher A, de Koning HP, Maser P (2007) Chemotherapeutic strategies against *Trypanosoma brucei*: drug targets vs. drug targeting. *Curr Pharm Des* 13: 555–567.
36. Odds FC (2005) Genomics, molecular targets and the discovery of antifungal drugs. *Rev Iberoam Micol* 22: 229–237.
37. Abdulla MH, Lim KC, Sajid M, McKerrow JH, Caffrey CR (2007) Schistosomiasis mansoni: novel chemotherapy using a cysteine protease inhibitor. *PLoS Med* 4: e14.
38. Engel JC, Doyle PS, Hsieh I, McKerrow JH (1998) Cysteine protease inhibitors cure an experimental *Trypanosoma cruzi* infection. *J Exp Med* 188: 725–734.
39. Selzer PM, Pingel S, Hsieh I, Ugele B, Chan VJ, et al. (1999) Cysteine protease inhibitors as chemotherapy: lessons from a parasite target. *Proc Natl Acad Sci U S A* 96: 11015–11022.
40. Iten M, Mett H, Evans A, Enyaru JC, Brun R, et al. (1997) Alterations in ornithine decarboxylase characteristics account for tolerance of *Trypanosoma brucei* rhodesiense to D,L-alpha-difluoromethylornithine. *Antimicrob Agents Chemother* 41: 1922–1925.
41. Zhang K, Rathod PK (2002) Divergent regulation of dihydrofolate reductase between malaria parasite and human host. *Science* 296: 545–547.
42. Renslo AR, McKerrow JH (2006) Drug discovery and development for neglected parasitic diseases. *Nat Chem Biol* 2: 701–710.
43. Gelb MH, Van Voorhis WC, Buckner FS, Yokoyama K, Eastman R, et al. (2003) Protein farnesyl and N-myristoyl transferases: piggy-back medicinal chemistry targets for the development of antitrypanosomatid and antimalarial therapeutics. *Mol Biochem Parasitol* 126: 155–163.
44. Nwaka S, Hudson A (2006) Innovative lead discovery strategies for tropical diseases. *Nat Rev Drug Discov* 5: 941–955.
45. Overington JP, Al-Lazikani B, Hopkins AL (2006) How many drug targets are there? *Nat Rev Drug Discov* 5: 993–996.
46. Keil M, Marhöfer RJ, Rohwer A, Selzer PM, Brickmann J, et al. (2008) Molecular visualization in the rational drug design process *Frontiers in Bioscience* in press.
47. Selzer PM, Chen X, Chan VJ, Cheng M, Kenyon GL, et al. (1997) *Leishmania major*: molecular modeling of cysteine proteases and prediction of new nonpeptide inhibitors. *Exp Parasitol* 87: 212–221.
48. Zhong H, Bowen JP (2006) Antiangiogenesis drug design: multiple pathways targeting tumor vasculature. *Curr Med Chem* 13: 849–862.
49. Chen X, Chong CR, Shi L, Yoshimoto T, Sullivan DJ Jr, et al. (2006) Inhibitors of *Plasmodium falciparum* methionine aminopeptidase 1b possess antimalarial activity. *Proc Natl Acad Sci U S A* 103: 14548–14553.
50. Lodge JK, Jackson-Machelski E, Higgins M, McWherter CA, Sikorski JA, et al. (1998) Genetic and biochemical studies establish that the fungicidal effect of a fully depeptidized inhibitor of *Cryptococcus neoformans* myristoyl-CoA:protein N-myristoyltransferase (Nmt) is Nmt-dependent. *J Biol Chem* 273: 12482–12491.
51. Nassar N, Cancelas J, Zheng J, Williams DA, Zheng Y (2006) Structure-function based design of small molecule inhibitors targeting Rho family GTPases. *Curr Top Med Chem* 6: 1109–1116.
52. Delcroix M, Sajid M, Caffrey CR, Lim KC, Dvorak J, et al. (2006) A multienzyme network functions in intestinal protein digestion by a platyhelminth parasite. *J Biol Chem* 281: 39316–39329.
53. Skelly PJ, Da'dara A, Harn DA (2003) Suppression of cathepsin B expression in *Schistosoma mansoni* by RNA interference. *Int J Parasitol* 33: 363–369.
54. Sayed AA, Simeonov A, Thomas CJ, Inglese J, Austin CP, et al. (2008) Identification of oxadiazoles as new drug leads for the control of schistosomiasis. *Nat Med* 14: 407–412.
55. <http://redpoll.pharmacy.ualberta.ca/drugbank/> DrugBank.
56. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, et al. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* 34: D668–672.
57. <http://www.pir.uniprot.org/> Uniprot: the universal protein resource.
58. Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, et al. (2006) The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res* 34: D187–191.
59. <http://www.expasy.org/prosite/> Prosite: database of protein domains, families, and functional sites.
60. Hulo N, Bairoch A, Bulliard V, Cerutti L, De Castro E, et al. (2006) The PROSITE database. *Nucleic Acids Res* 34: D227–230.
61. <http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim> OMIM: Online Mendelian Inheritance in Man.
62. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, et al. (2007) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* .
63. <http://www.ebi.ac.uk/interpro/> InterPro.
64. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, et al. (2007) New developments in the InterPro database. *Nucleic Acids Res* 35: D224–228.
65. Finn RD, Tate J, Mistry J, Coghill PC, Sammut SJ, et al. (2008) The Pfam protein families database. *Nucleic Acids Res* 36: D281–288.
66. <http://pfam.sanger.ac.uk/> Pfam.
67. <http://www.rcsb.org/pdb/home/home.do> PDB: Protein Data Bank.
68. <http://www.genedb.org/perl-gb/gbrowse/S.mansoni/> Schistosoma mansoni genome browser.