Journal of proteome research

Subscriber access provided by Access provided by University of Liverpool Library

Article

Quantitative Proteomics of Enriched Esophageal and Gut Tissues from the Human Blood Fluke *Schistosoma mansoni* Pinpoints Secreted Proteins for Vaccine Development

Leandro X. Neves, R. Alan Wilson, Philip Brownridge, Victoria M. Harman, Stephen W. Holman, Robert J. Beynon, Claire E. Eyers, Ricardo DeMarco, and William Castro-Borges

J. Proteome Res., **Just Accepted Manuscript** • DOI: 10.1021/acs.jproteome.9b00531 • Publication Date (Web): 15 Nov 2019

Quantitative Proteomics of Enriched Esophageal and Gut Tissues from the Human Blood Fluke *Schistosoma mansoni* Pinpoints Secreted Proteins for Vaccine Development

Leandro X. Neves [1], R. Alan Wilson [2], Philip Brownridge [3], Victoria M. Harman [3], Stephen W. Holman [3], Robert J. Beynon [3], Claire E. Eyers [3], Ricardo DeMarco [4] and William Castro-Borges [1*]

[1] Departamento de Ciências Biológicas, Universidade Federal de Ouro Preto, Campus Morro do Cruzeiro, Ouro Preto, Minas Gerais, Brasil.

[2] Centre for Immunology and Infection, Department of Biology, University of York, Heslington, York, United Kingdom.

[3] Centre for Proteome Research, Department of Biochemistry, Institute of Integrative Biology, University of Liverpool, Liverpool, L69 7ZB, United Kingdom.

[4] Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, Brasil.

*Correspondence should be addressed to:

William de Castro Borges, Departamento de Ciências Biológicas, Instituto de Ciências Exatas e Biológicas, sala 58, Laboratório de Enzimologia e Proteômica. Universidade Federal de Ouro Preto, Campus Morro do Cruzeiro. Ouro Preto - Minas Gerais - Brasil. CEP: 35.400-000. Phone: +55 31 3559 1705

FAX: +55 31 3559 1680;

e-mail: wborges@ufop.edu.br

ABSTRACT

Schistosomes are blood-dwelling helminth parasites that cause schistosomiasis, a debilitating disease resulting in inflammation and, in extreme cases, multiple organ damage. Major challenges to control the transmission persist and the discovery of protective antigens remains of critical importance for vaccine development. Rhesus macaques can self-cure following schistosome infection, generating antibodies that target proteins from the tegument, gut and esophagus, the last of which is the least investigated. We developed a dissection technique that permitted increased sensitivity in a comparative proteomics profiling of schistosome esophagus and gut. Proteome analysis of the male schistosomes esophagus identified 13 proteins encoded by microexon genes (MEG), eleven of which were uniquely located in the esophageal glands. Based on this and transcriptome information, a QconCAT was designed for absolute quantification of selected targets. MEGs 12, 4.2, 4.1 and Venom allergen-like protein 7 were the mostly abundant, spanning over 245-6 million copies per cell, while aspartyl protease, palmitoyl thioesterase and a galactosyl transferase were present at <1 million copies. Antigenic variation by alternative splicing of MEG proteins was confirmed together with a specialised machinery for protein glycosylation/secretion in the esophagus. Moreover, some gastrodermal secretions were highly enriched in the gut, while others were more uniformly distributed throughout the parasite, potentially indicating lysosomal activity. Collectively, our findings provide a more rational, better-oriented selection of schistosome vaccine candidates in the context of a proven model of protective immunity.

Keywords:

Schistosoma mansoni; esophageal gland; gastrodermis; microexon genes; quantitative proteomics; QconCAT

Introduction

Schistosomes are blood-dwelling helminth parasites responsible for a chronic and debilitating disease in individuals living in tropical and subtropical regions [1]. At least 165 million people are estimated to be infected in Africa alone, largely based on detection of eggs in urine or faeces. However, the recent use of more sensitive diagnostics in areas of low transmission (<50% prevalence) indicates that cases of schistosomiasis may be underestimated by a factor of 1.5 to 6-fold when circulating cathodic antigen tests are compared to classical Kato-Katz technique [2]. A single drug, Praziquantel, is currently available to treat infected individuals but has limitations, notably its lack of activity against juvenile worms [3]. Chemotherapy invariably is started when the inflammatory responses against the eggs lodged in the host liver, intestines and bladder have already been established. In addition, data for 2016 suggest that annually the drug is only reaching ~36% of people needing treatment. In this situation, a protective vaccine to prevent infection remains of critical importance in the battle to combat schistosomiasis [4].

The rhesus macaque (*Macaca mulatta*) is unusual among the various vertebrate hosts that schistosomes can infect, in that it can eliminate adult worms. A mature *Schistosoma mansoni* population establishes with oviposition peaking at week 9 before a gradual decline between 12 and 18 weeks [5]. Recovery of surviving worms by perfusion of the hepatic portal system reveals that they have ceased feeding and are starving to death. Classical immunoproteomics (2D-PAGE and western blotting) indicated potential antibody targets in both surface tegument and secreted gut protein preparations [5]. A more detailed investigation of *S. japonicum* infection in the Rhesus macaque has further illuminated the possible mechanisms that prevent blood feeding. The esophageal lumen, surrounded by the anterior and posterior esophageal glands, was identified as the primary site for interaction with a potent humoral immune response [6]. Differential transcriptomics indicates that the gland secretions comprise a

cocktail of enzymes and a variety of proteins encoded by microexon genes (MEGs) [7,8]. We have documented that initial processing of ingested blood occurs in the esophageal lumen by mixing with the gland secretions [9]. Erythrocytes and platelets are lysed, while leucocytes are inactivated before they are propelled to the gut lumen where further digestion and absorption of nutrients take place [10]. Notably, the rhesus macaque seems able to block this initial blood processing by disrupting the release of gland secretions, ultimately leading the parasites to starvation [6]. To date, the function of these proteins has been inferred only from bioinformatic analyses of the encoding transcripts and their tissue pattern of expression [6–8]. An important step in the characterisation of proteins released into the esophageal lumen as potential immune targets is to determine their profile and relative abundance.

There are multiple challenges associated with the proteomic characterisation of esophageal gland secretions from schistosomes. Perhaps, the most important is that both anterior and posterior esophageal glands represent a tiny percentage of the whole worm body [9]. Consequently, constituents derived primarily from muscle tissues prevent identification of the unique set of gland products from whole worm extracts. Thus our recent investigation of the soluble protein composition of a *S. mansoni* preparation failed to detect any predicted gland constituents, attesting to their low abundance [11]. In addition, the prediction of multiple splice variants for MEG proteins [12,13] indicates the need for a more refined experimental design for sample preparation, combined with the utilization of modern, fast and sensitive mass spectrometric platforms for protein identification and quantification.

Here we tackle these challenges using a dissection technique to specifically target and enrich the esophageal gland for proteomic characterisation. Absolute quantification of selected protein targets, using QconCAT technology [14,15], provided information on the abundance of gland constituents involved in blood processing. In addition, shotgun proteomics analysis confirmed that microexon genes expressed in the esophageal glands may be variably spliced.

Finally, a comparative analysis of the worm esophagus and posterior body (back end) served not only to highlight genuine gland products but also permitted the identification of proteins involved in nutrient acquisition in the lower alimentary tract. Overall, we believe our findings will provide the basis for a more rational and better-oriented selection of candidates of a vaccine against schistosome infection in the context of a proven model of protective immunity.

Experimental section

Ethical Statement

All procedures for maintenance of the *S. mansoni* life cycle and worm recovering were approved by the Ethics Committee on Animal Experimentation (Comissão de Ética no Uso de Animais – CEUA), Universidade Federal de Ouro Preto, protocol nº. 2011/55.

Worm recovery and dissection

Adult *S. mansoni* worms were obtained by portal perfusion of 45 day-infected mice using 10 mM HEPES buffered RPMI-1640 medium, pH 7.4 (Sigma-Aldrich, USA), containing 4 IU/ml of heparin. After extensive washes in pre-warmed medium (37 °C), parasites were instantly fixed in RNA*later*® Solution (Invitrogen, UK) for protein stabilisation and stored at 4 °C, as previously described [7]. Briefly, adult male worms immersed in ice-cold RNAlater® were carefully held and dissected using fine curved tweezers (Ideal Tek, Chiasso, Switzerland) and Vannas scissors (John Weiss, Milton Keynes, UK), under a stereomicroscope at x35 magnification. First, the oral sucker was removed by an incision at its posterior, at the very beginning of the esophagus. Next, a second cut along the line of the transverse gut released the esophageal fragment (ESO). The back end (BE) comprising the posterior third of the male body was excised for comparative analysis. Dissected fragments

were preserved in RNA*later*® Solution at 4 °C until sample preparation, following the manufacturer's instruction. The detailed and illustrated dissection procedure is uploaded at protocols.io repository [16] (https://www.protocols.io/) and accessible via DOI dx.doi.org/10.17504/protocols.io.tq2emye. Biological replicates obtained from independent mouse infection, perfusion and worm dissection procedures provided material for individual shotgun and targeted proteomic analyses.

Sample preparation and in-solution digestion

ESO and BE fragments were washed twice in 1 mL of ice-cold PBS, pH 7.4, to remove RNA*later*® Solution, as recommended by the manufacturer. Samples were sonicated in ice bath for 5 cycles of 10 second pulses at 30% amplitude (Sonics Vibra Cell™; Jencons Scientific Ltd, UK), with 50 seconds intervals. Total protein concentration was determined using a Coomassie Plus (Bradford) assay kit (Thermo Scientific, UK) by interpolation from a BSA standard curve, according to the manufacturer's instructions.

Sample aliquots equivalent to 20 µg of protein were processed through in-solution digestion protocol, essentially as described elsewhere [17]. Briefly, the aliquots were transferred to LoBind tubes (Eppendorf, USA) and made up to 100 µL by the addition of 25 mM ammonium bicarbonate (AMBIC). Protein denaturation was induced by the addition of 20 µL of RapiGest SF 1% w/v (Waters, UK) and heating at 80 °C for 10 min. Next, sample volume was adjusted to 180 µL with AMBIC before reduction (addition of 10 µL 60 mM dithiothreitol, 60 °C for 10 min) and alkylation (addition of 10 µL 180 mM iodoacetamide and incubation at room temperature for 30 min). MS-grade trypsin (Promega) was added at an enzyme:substrate ratio of 1:50 and digestion allowed to occur at 37 °C for 16 h. Trypsinolysis was terminated by addition of 1.5 µL trifluoroacetic acid. Acidified samples (pH<2) were then incubated at 37 °C for 45 min to precipitate RapiGest SF, which was later

removed by centrifugation at 13,000*g* for 15 min at 7 °C. Supernatant was recovered for LC-MS analysis.

nLC-ESI-MS/MS analysis of ESO vs BE region

ESO and BE digests were analysed by a discovery-based proteomics approach using either a Q Exactive HF™ hybrid quadrupole-orbitrap (Thermo Scientific, Germany) or an Orbitrap Fusion™ Tribrid™ mass spectrometer (Thermo Scientific, Germany). Both instruments were coupled to Dionex™ UltiMate™ 3000 UHPLC system (Thermo Scientific, Germany) with the same column specifications. Briefly, 700 ng of digested samples were resolved over a 90- or 95-min gradient and analysed in positive mode using the Q Exactive HF and Orbitrap Fusion instruments, respectively. Spectral acquisition on the Q Exactive HF was performed using Top16 data-dependent acquisition (DDA) whereas Orbitrap Fusion operated in TopSpeed DDA mode with a 3-second cycle. In the latter platform, samples were analysed using CID and HCD activation capabilities of the instrument, and product ions were detected using the Ion Trap at rapid scan rate. Expanded LC-MS methods including a geLC-MS/MS analysis using a Q Exactive mass spectrometer (Thermo Scientific, Germany) for construction of a spectral library are available in the Supplementary Material S1.

Database search and label-free quantification

De novo sequencing-assisted database searching [18] was performed using PEAKS Studio v8.5 (Bioinformatics Solutions Inc, Canada). Firstly, spectral data were searched against *S. mansoni* sequences deposited at GeneDB.org (downloaded at 17/10/2017) supplemented with protein sequences of esophageal hydrolases detected in a previous RNAseq experiment [7] (10,875 sequences; 5,942,443 residues). Esophageal MEG proteins were individually searched using a database composed of the sequences deposited in GeneDB and NCBI plus

putative alternative spliced variants generated *in silico* by combinatorial exon permutation (2,269 sequences; 223,997 residues). Precursors and product ion mass error tolerances were set respectively to 15 ppm and 0.02 Da, for the Q Exactive HF, or 10 ppm and 0.35 Da, for the Fusion instrument. Enzyme specificity was set to trypsin (P1 = K/R, except if P1'= P) allowing up to one missed cleavage site. Cysteine carbamidomethylation and dynamic methionine oxidation were set as peptide modifications. The quality threshold for Peptide-spectral match (PSM) was adjusted to keep the false discovery rate (FDR) ≤0.01.

Relative label-free quantification for comparative ESO *vs* BE analysis was performed using Peaks Q enabling the exclusion of peptides with both modified and unmodified forms. Differentially abundant proteins with Benjamin–Hochberg adjusted *p*-values ≤0.01, fold difference ≥1.5 and at least two unique peptides comprised the subset of ESO- or BE-enriched components.

Gene ontology and functional categorisation

Differentially abundant proteins were subjected to gene ontology analysis using a PRO trial licenced Blast2GO v5 software [19]. Gene Ontology (GO) terms were mapped and the Fisher's Exact Test (FDR ≤0.05) assessed the enrichment of specific biological processes, molecular functions and cellular components amongst the proteins with higher relative abundances in the ESO compared to BE. Functional enrichment analysis on STRING v11 [20] provided KEGG pathways, keywords and GO terms over-represented in the list of differentially abundant proteins. Additional categorisation of cellular markers, namely the glycosyl transferases and gastrodermis secretion, was performed essentially as described elsewhere [7,21].

QconCAT design, expression and purification

Based on state of the art QconCAT technology [14,15] , an EsoCAT construct was designed by concatenating proteotypic peptides to obtain an artificial protein that served as the internal standard for simultaneous quantification of multiple analytes. The targets selected for absolute quantification comprised nine proteins with specific expression in the esophageal glands, previously validated by whole-mount *in situ* hybridization and confocal microscopy [7,9], and detected by mass spectrometry.

Fifteen proteotypic peptides were selected from our in-house spectral library. Two additional peptide sequences were chosen based on their putative detectability, as predicted by the CONSeQuence algorithm [22], and two N-terminal variant tryptic peptides were selected to investigate co-occurring MEG-12 isoforms (Supplementary Material S2). Uniqueness of proteotypic peptides was ascertained against the *Mus musculus* (UniProt UP000000589; 60,177 sequences and 27,265,929 residues) and *Escherichia coli* BL21-DE3 (UniProt UP000002032; 4,156 sequences and 1,298,178 residues) proteomes, and GPM collection of common contaminants cRAP (116 sequences and 38,459 residues).

A minimum of two proteotypic peptides was selected for each target protein, with the exception of MEG-4.2 for which only one tryptic peptide had been identified during our spectral library construction. Moreover, sequences consisting of three amino acids found naturally flanking the proteotypic peptides in the endogenous target proteins were preserved in the EsoCAT construct to provide conditions for both analyte and internal standard to be evenly digested, as described elsewhere [23]. The complete EsoCAT primary structure comprised a N-terminal initiator methionine followed by the glufibrinopeptide B (EGVNDNEEGFFSAR), and a C-terminal hexahistidine tag (Supplementary Material S2).

For heterologous expression of EsoCAT, the synthetic gene was ligated into a pET21a plasmid vector (Eurofins Genomics, Germany) and transformed into BL21 (λDE3) competent *E. coli* cells. EsoCAT expression was induced in the presence of 1 mM isopropyl β-D-1-

thiogalactopyranoside (IPTG) in minimal culture media containing $^{13}C_6$ labelled lysine and arginine, as described elsewhere [24]. After 3.5 hours of induction, *E. coli* cells were collected by centrifugation (1,450 *g*, 15 min, 4 °C), resuspended in 50 mM AMBIC, pH 8.0, containing 25 U/mL benzonase nuclease (Novagen), 1x Complete EDTA-free protease inhibitor tablet (Roche), and sonicated in an ice bath using 10 second pulses at 30% amplitude, with 50 second intervals, until 130 joules was reached. After centrifuging the homogenate at 6,000 *g*, 7 min, 7 °C, the EsoCAT protein present in the insoluble inclusion body pellets was solubilised in Buffer A (20 mM $NaPO_4$, 0.5 M NaCl, 10 mM imidazole, 6 M guanidine hydrochloride, pH 7.4) and filtered using a 1.2 µm syringe filter.

EsoCAT purification was carried out using an AKTA start chromatography system (GE Healthcare) equipped with a 1 mL HisTrap™ HP column (GE Healthcare). System conditioning and washing were performed using Buffer A before bound proteins were eluted over a 20 min gradient of 0-100% Buffer B (20 mM $NaPO_4$, 0.5 M NaCl, 500 mM imidazole, 6 M guanidine hydrochloride, pH 7.4), at a flow rate of 1 mL·min^{-1}. Eluted fractions containing EsoCAT were pooled and dialysed against 50 mM AMBIC containing 1 mM dithiothreitol. Purified EsoCAT preparation was supplemented with 0.1% (w/v) RapiGest SF (Waters) and aliquoted for storage at -20 °C in LoBind tubes (Eppendorf). Immediately before use, EsoCAT was thawed then heated to 60 °C in the presence of 3 mM dithiothreitol for protein reduction, thus preventing aggregate formation and precipitation.

In-solution co-digestion with EsoCAT internal standard

Absolute quantification required the co-digestion of the proteome of interest and the QconCAT internal standard. Four replicates of a master mixture were prepared by spiking an estimated concentration of purified EsoCAT protein into 20 µg of the ESO proteome. The master mixtures were subjected to our standard in-solution digestion protocol in the presence

of 100 fmol glufibrinopeptide B (Glufib B >98% purity; Severn Biotech) enabling the accurate quantification of the EsoCAT protein. In parallel, 20 µg digests containing only the ESO proteome were produced and used for serial dilution of the master mix by factors of 1:10 and 1:100. This ensured the EsoCAT internal standard spanned a concentration range of two orders of magnitude (~0.15-150 fmol/µg of ESO proteome). LoBind tubes were used in all steps of this experiment.

Absolute quantification by SRM-MS

Absolute quantification was performed on a nanoACQUITY UPLC™ system coupled to a Xevo TQS triple quadrupole mass spectrometer (Waters) set to Selected Reaction Monitoring (SRM) acquisition mode, with Q1 and Q3 operating at unit mass resolution. Up to one microgram of digested samples was loaded onto a Symmetry C_{18} trap column (5 µm, 180 µm x 20 mm; Waters) and washed for 3 min, at a flow rate of 5 $\mu L \cdot min^{-1}$, with 0.1% (v/v) formic acid in water. Peptides were resolved on the analytical column HSS T3 nanoACQUITY C18 (1.8 µm, 75 µm x 150 mm; Waters) over a 90 min linear gradient of 3-40% (v/v) MeCN in 0.1% (v/v) formic acid, at a constant flow of 300 $nl \cdot min^{-1}$, at 35 °C. SRM acquisition was scheduled with 4 min windows and dwell times automatically adjusted to achieve ≥12 sampling points per peak. Three to four product ions with highest relative intensity in experimental MS/MS spectra and *m/z* > precursor ion were selected to compose the list of transitions.

Data analysis of QconCAT assay

Quantification of endogenous and surrogate peptides was performed using Skyline v4.2 [25] by integrating extracted ion chromatograms for transitions monitored in light and heavy channels, respectively (Supplementary Material S2). An equivalent number of transitions

with random mass shifts (decoys) was also monitored allowing peak validation and FDR calculation by mProphet [26]. The concentration of labelled peptides was firstly assessed based on the reference Glufib B peptide. Analyte quantification was finally achieved using the light to heavy ratios observed in the isotopic dilution that fulfilled the following criteria: (i) closest 1:1 ratio (never exceeding 10x fold) between the internal standard and endogenous analyte that exhibited a (ii) mProphet's FDR ≤ 0.05 score in at least three out of four replicates.

To classify the performance of proteotypic peptides monitored by the SRM quantitative assay, we adopted the terminology described by Brownridge *et al* (2011) [27]. Accordingly, peptides were classified as Type 'A' when EsoCAT standard and endogenous analyte delivered high quality quantification data ('S^+/A^+' or 'standard positive, analyte positive'), Type 'B' when solely EsoCAT internal standard ('S^+/A^-') was observed, or Type 'C' when peptides could not be observed in either heavy or light channels ('S^-/A^-'). We refined this classification system by incorporating mProphet scores into the results. This divided Type 'A' category into two further subgroups: Type 'A1' peptides successfully passed mProphet's 5% FDR threshold and were used for protein quantification, whereas Type 'A2' peptides failed to give a clearer signal compared to decoy transitions (mProphet's FDR >0.05) and were removed.

Results

Mass spectrometry analyses of ESO detect predicted esophageal gland proteins and sequence variation

For the first time we were able to identify predicted esophageal gland proteins using a mass spectrometry-based approach on the ESO-enriched preparation. Detection of the group of MEG-encoded proteins, known to be specifically expressed in the gland cells [7,28,29], confirmed the successful combination of shotgun proteomics and microdissection procedures.

Database search against deposited MEG-encoded sequences and alternative spliced variants generated by *in silico* exon permutation allowed the identification and study of MEG isoforms expressed in the esophageal gland. In parallel, database interrogation against *S. mansoni* protein sequences deposited in GeneDB resulted in a list of 4,024 proteins identified at ≤1% FDR. This was later reduced to 3,019 identities confidently assigned to at least two unique peptides.

Compositional analyses revealed that ESO and BE preparations share over 90% of the proteins confidently detected, whilst 141 and 54 proteins were exclusive to ESO and BE samples, respectively (Figure 1A). Among the proteins exclusively detected in the ESO sample, ten MEG proteins (MEG-3.4, MEG-4.1, MEG-4.2, MEG-5, MEG-8.1, MEG-8.2, MEG-8.4, MEG-14, MEG-15 and MEG-24) were confidently identified in this proteomic analysis, with ≥2 two PSM per protein at a false discovery rate ≤1%. Additionally, MEG-26.3, MEG-16 and MEG-12 were detected using less stringent filter criteria (1 PSM at ≤1% FDR). Median peptide intensity values of unique peptides indicated that eleven of these MEG proteins are uniquely present in the worms' esophagus (Figure 1B; Supplementary Material S3). The remaining two namely MEG-5, a protein assigned to the tegument [12], and MEG-24 were detected at similar levels in the ESO and BE preparations normalised by their total protein content.

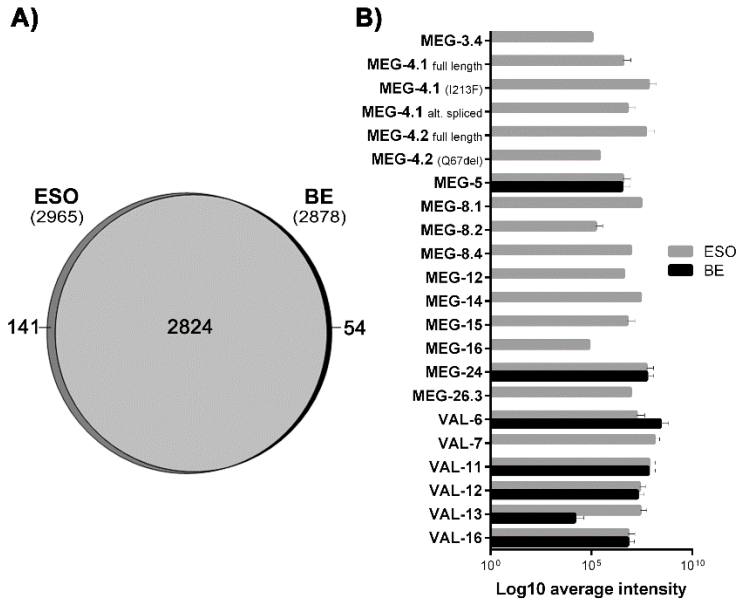


Figure 1. Composite analyses of ESO and BE samples and differential abundance of esophageal signatures. **(A)** Venn diagram indicating over 90% commonalities between the proteins confidently identified (≥2 unique peptides; FDR ≤0.01 FDR) in the samples. A fraction of 141 proteins were detected only in the esophageal region (ESO) whilst 54 were present uniquely in the back end (BE) of adult male worms. **(B)** Abundance of unique peptides indicated that 11 out of 13 MEG are among the exclusively expressed in the schistosome esophagus, similarly VAL-7.

A second protein family investigated comprised the Venom Allergen-Like (VAL) proteins since VAL-7 was the second gland-specific constituent discovered [28]. Our data confirmed that VALs 7 and 13 were highly enriched in the ESO fraction compared to BE (Figure 1B), whereas VALs 11, 12 and 16 had similar intensities in the two preparations, and VAL-6 found at higher levels in the BE.

Notably, our analysis produced not only the first MS-based detection of esophageal MEG proteins but also delivered evidence of co-occurring isoforms of these gland secretions. Sequence variations detected from the pool of worms includes alternative spliced sequences, residue substitution and deletion. For instance, in comparison to the full-length of MEG-4.1 (GenBank accession AAA29855) we observed two amino acid substitutions, K57E and I213F (Figure 2A), that matched other deposited transcript sequences (*e.g.* GenBank accession AAP13803.1 and RefSeq XM_018793269.1). Moreover, we gathered evidence for the co-occurrence of MEG-4.1 isoforms resulting from alternative splicing. The peptide 'EYIEDKNVDIR' present in the full-length sequence of MEG-4.1 is formed over an exon-exon junction so that five N-terminal amino acids are encoded by exon 8, whereas the following residues are encoded by exon 10 (27 bp). Hence, skipping exon 10 results in

ligation between exons 8/11 and omission of the 9-mer sequence 'KNVDIRIIG' (K62_G70del), ultimately creating the new detectable tryptic product 'EYIEDNKK' (Figure 2A). It is noteworthy that exon 9 is rarely incorporated into MEG-4.1 transcripts and its asymmetrical structure of 89 bp would produce an incomplete protein by disrupting the open reading frame.

In a second case, the deletion of a glutamine residue (Q67) in MEG-4.2 (Smp_085840.1) also resulted in an alternative detectable proteotypic peptide for this esophageal protein. The peptide 'QLEEEQNPFHK' is formed over the junction of exons 6/7 so that the second glutamine of the sequence overlaps a splice site (Figure 2B). The alternative sequence, 'QLEEENPFHK', is obtained via an alternative splice site at the 3'-end of exon 6 that changes the exon length from 21 to 18 bp, excluding the glutamine codon. Finally, peptide identifications of MEG-14 matched alternative sequences shared by several known isoforms

Figure 2. Coverage of MEG-4.1 and 4.2 isoforms by shotgun proteomics. (A) Full length sequence of MEG-4.1 (AAA29855) highlighting the mechanism of sequence variation of the detected isoforms. Amino acids encoded by exon 10 (red area) are deleted and direct ligation of exons 8 / 11 occurs (green area). Additionally, two residue substitutions (K57E and I213F) are indicated in the boxes. Residues in orange are predicted sites of O-glycosylation. Versions of MEG-4.1 incorporating exon 9 are rare and incomplete. The hashed line above the protein sequence indicates the tandemly repeated region. **(B)** MEG-4.2 protein sequence (Smp_085840.1) with a residue encoded by an alternative splicing site (Q67) highlighted in red. The alternative sequence with Q67del is symbolised by the box with "X". Peptide matches (FDR ≤0.01) are indicated by blue (≥2 PSM) and yellow (1 PSM) bars below the sequences.

produced by alternative splicing [13]. However, none of the five peptides detected are self-excluding, hence a unique isoform cannot be identified unambiguously. Instead, we observed that the peptide sequences are sometimes present in two, or in up to seven different MEG-14 variants (Supplementary Material S3).

QconCAT proteomics allows the abundance of gland secreted proteins to be determined

We employed QconCAT-based quantification to capture the abundance of nine specific gland antigens in the ESO proteome. Seventeen out of 21 precursor ions monitored by SRM for absolute quantification were classified as Type 'A1' peptides (*i.e.* both analyte and surrogate peptides detected at mProphet FDR ≤ 0.05), thus being suitable for quantification (Supplementary Material S4). After removal of low scoring peptides (mProphet FDR >0.05), MEG-4.1, MEG-8.2, MEG-15, VAL-7, aspartyl protease and palmitoyl thioesterase could be quantified by two of their proteotyptic peptides and three other proteins (MEG-4.2, MEG-12 and β-1,3-GalTase) by a single peptide standard each.

Using a mass-to-mass unit (ng/µg) the nine quantified proteins amounted to only ~0.5% of the total protein found in the male worms' esophagus (0.5 ± 0.17 µg per ESO fragment; Supplementary Material S4). However, since only 750 and 1,000 cells comprise the anterior and posterior glands [9] respectively, it is reasonable to expect that the gland products would not dominate an esophageal homogenate. Therefore, we estimated the number of protein copies per cell (cpc) as a better representation of our quantitative data since it normalised protein abundance to the number of producing cells (Figure 3).

When the data are viewed in this way, MEG-12, the only selected target demonstrated as expressed in the anterior gland, showed extraordinary abundance with over 245 million cpc, likely the integral of production and externalisation. VAL-7, MEG-4.2 and MEG-4.1 were also highly abundant with between 12.2-4.5 million cpc (Supplementary Material S4). The

remaining predicted secreted proteins, namely MEG-15, MEG-8,2, palmitoyl thioesterase and aspartyl protease presented lower levels of 0.5-1 million cpc, whilst the endoplasmic protein β-1,3-GalTase showed the lowest cellular content of ~380 thousand cpc.

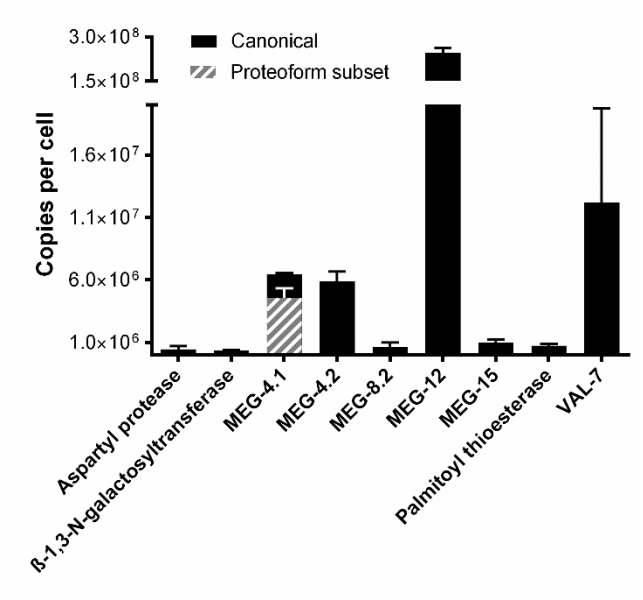


Figure 3. Absolute quantification of esophageal gland proteins. Protein copies normalised by the number of cells comprising the anterior and posterior esophageal glands in *S. mansoni* adult male worms, according to Li *et al* (2013). MEG-12, the only selected target demonstrated as being expressed in the anterior gland showed the highest abundance (~245 million cpc), followed by VAL-7 and MEGs-4.1 and 4.2. The EsoCAT construct allowed the quantification of MEG-4.1 via one peptide canonically expressed in the isoforms detected in this study and a second proteotypic peptide exclusive to an isoform subset (hashed bar).

Importantly, the knowledge about isoforms of MEG-4 family members must be considered in the QconCAT analysis. The two proteotypic peptides of MEG-4.1 included in the EsoCAT allowed the quantification of specific isoform subsets. Thus, MEG-4.1 variants containing an isoleucine rather than a phenylalanine residue at position 213 accounted for 71% of all copies detected using the second peptide, QINDGTSDKPK, conserved among known alternative sequences (Figure 3; MEG-4.1 solid and hashed bar). In the case of MEG-4.2, we only present the absolute quantification of the full length containing the QLEEEQNPFHK peptide - without the glutamine deletion. The occurrence of putative splicing variants of MEG-12 was not confirmed by the second peptide inserted in the EsoCAT construct. More specifically, we detected and quantified ENYEQQLQQPK (Smp_152630.1; both internal standard and endogenous analyte present) but not the FHIVFFCGENYEQQLQQPK (Smp_152630.2; only internal standard present), suggesting that the putative splicing variant Smp_152630.2 lies below the instrumental limit of detection (Figure 4).

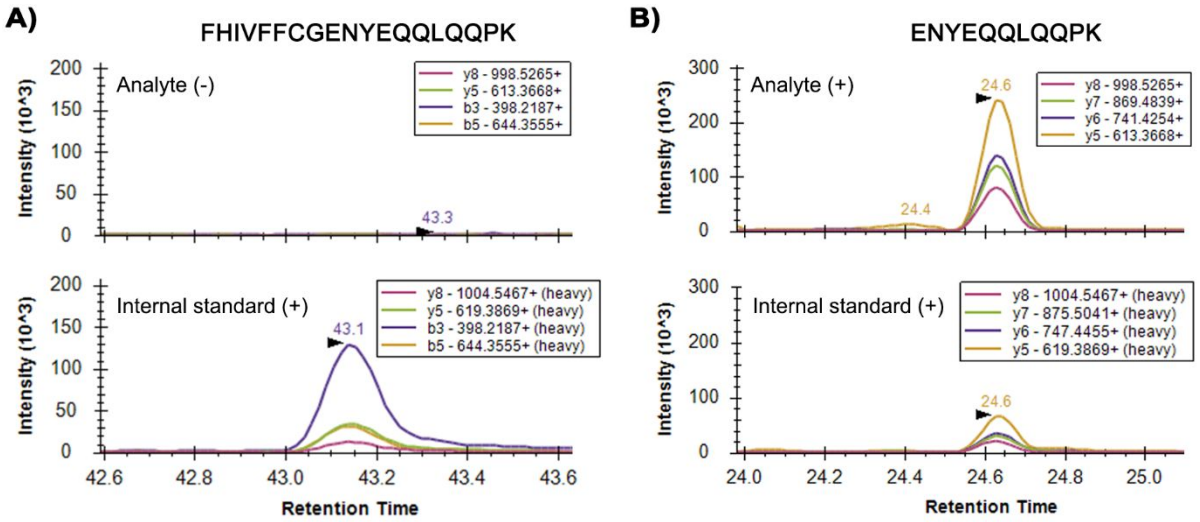


Figure 4. Proteotypic peptides monitored for MEG-12 quantification. (A) Extracted ion chromatogram of transitions monitored for FHIVFFCGENYEQQLQQPK peptide indicated that the Smp_152630.2 isoform is below the limit of detection. **(B)** Conversely, at the same spiking levels of ENYEQQLQQPK internal standard the endogenous analyte derived from Smp_152630.1 isoform is detected.

GO analyses suggest distinctive molecular functions in the worm esophagus

Next, we performed an ESO *vs* BE comparative analysis using label-free relative quantification. A high similarity between the preparations was observed in the Orbitrap

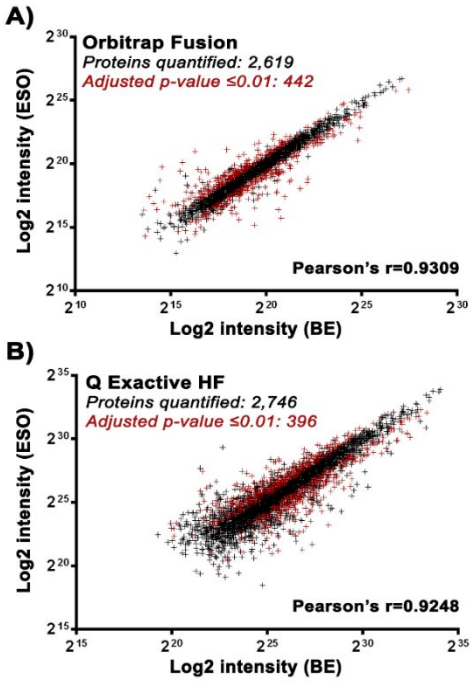


Figure 5. Shotgun quantitative analysis of ESO and BE samples. Compositional differences unveiled by label free relative quantification of the data acquired using Orbitrap Fusion **(A)** and Q Exactive HF **(B)** indicating, respectively, 442 and 396 differentially abundant proteins (adjusted p-value ≤0.01, ≥1.5-fold difference, ≥2 unique peptides). Pearson's correlation indicates high similarity between the two preparations (R^2 >0.92).

Fusion and Q Exactive HF analysis (Pearson' r > 0.9; Figures 5A-B). Nevertheless, a total of 669 proteins were found differentially abundant in the two platforms together (adjusted p-values ≤0.01, fold difference ≥1.5, ≥2 unique peptides; Supplementary Table S1A-B).

Functional analysis using STRING database indicated the enrichment of a cluster of esophageal proteins involved in glycerophospholipid and sphingolipid metabolism, and lysosomal lecithin:cholesterol acyltransferases among the differentially abundant proteins (Figure 6; right). On other hand, a remarkable enrichment of proteases in the BE was indicated by the overrepresentation of cysteine/papain-like peptidases metalloproteases and C1A peptidases (Figure 6; left). The lysosomal pathway and a cluster of proteases/C1A peptidases displayed a heterogenous composition with molecules differentially abundant in BE and ESO (Figure 6, hashed markers) indicating site-specific activities.

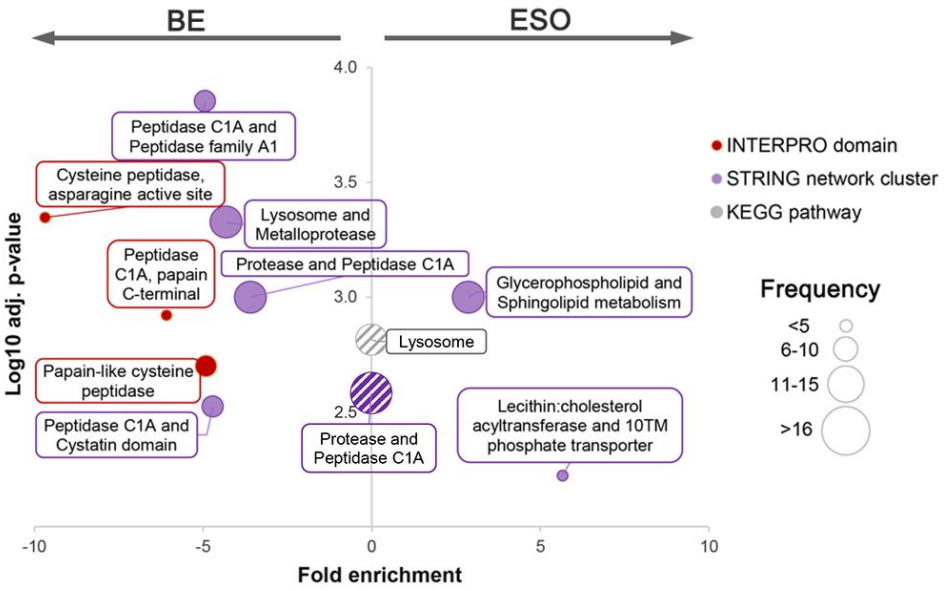


Figure 6. STRING functional analysis of differentially abundant proteins. INTERPRO domains associated with proteolytic activity are significantly overrepresented among proteins enriched in the BE (red circles, left). Two clusters of proteins associated with the glycerophospholipid/sphingolipid and lecithin:cholesterol metabolism were detected in the ESO component (purple; right). Constituents of lysosome pathway and a set of proteases/peptidases C1A were enriched in both samples suggesting site-specific activities (hashed circles; centre).

Next, a GO enrichment analysis was performed by direct comparison of the 359 proteins enriched in the male worm's ESO versus the 330 more abundant in the BE. Twenty-three over-represented biological processes (FDR ≤ 0.05) could be abridged to seven GO terms, representing protein biosynthesis and glycosylation, vesicle-mediated transport - including Golgi vesicles, oxidation-reduction processes and drug metabolism. Molecular functions and cellular components supported the presence of hexosyltransferases, membrane complexes, including those in endoplasmic reticulum/Golgi and mitochondrion, as well the oxidoreductase complex (Figure 7).

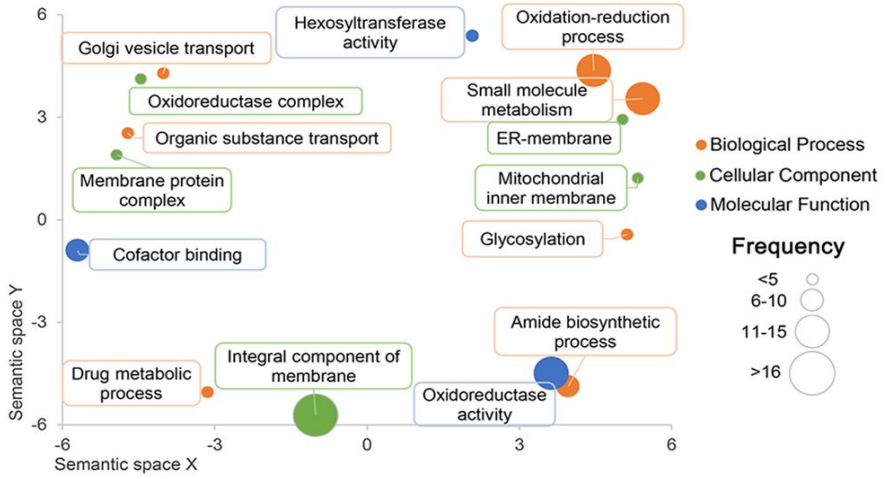


Figure 7. GO terms over-represented in the ESO compared to BE. Direct comparison of proteins differentially abundant in the ESO *vs* BE revealed that GO terms associated with bioenergetics (oxidative phosphorylation), biosynthetic processes (protein biosynthesis and glycosylation) and vesicle-mediated transport are over-represented in the esophagus (FDR≤0.05).

A closer look at the total of 25 quantified glycosyl transferases showed a greater abundance in the esophageal region, by a median factor of 1.8-fold (Figure 8; Supplementary Table S1C). Notably, 16 out of 18 differentially abundant glycosyl transferases were enriched in the ESO sample. Seven of these are dolichol lipid-linked endoplasmic reticulum transferases. Five enzymes with transferase activities spanning galactosyl (Smp_151220), GalNAc (Smp_057620, Smp_005500) and mannosyl (Smp_042790, Smp_177080, Smp_102430) glycans were also significantly enriched, as were two endoplasmic reticulum glucosidases (alpha-mannosidase, Smp_143430; mannosyl oligosaccharide glucosidase, Smp_024580).

Together, these findings indicate a distinct machinery for production of specialized glycans in the esophageal glands.

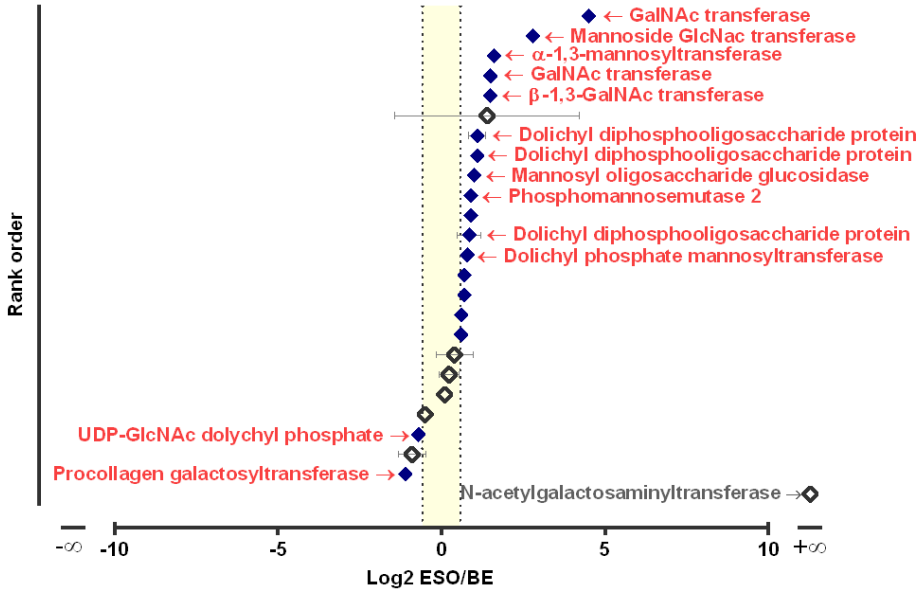


Figure 8. Differential abundance of glycosyl transferases in the ESO and BE samples. Diagram showing an enrichment of proteins related to glycosylation processes in the schistosome esophagus (right side). The shaded area indicates the equivalence range (<1.5-fold difference). Proteins are rank ordered according to their average Log2 fold ESO/BE. Differential expression supported by statistical analysis (adjusted p-value ≤0.01) are indicated by solid symbols.

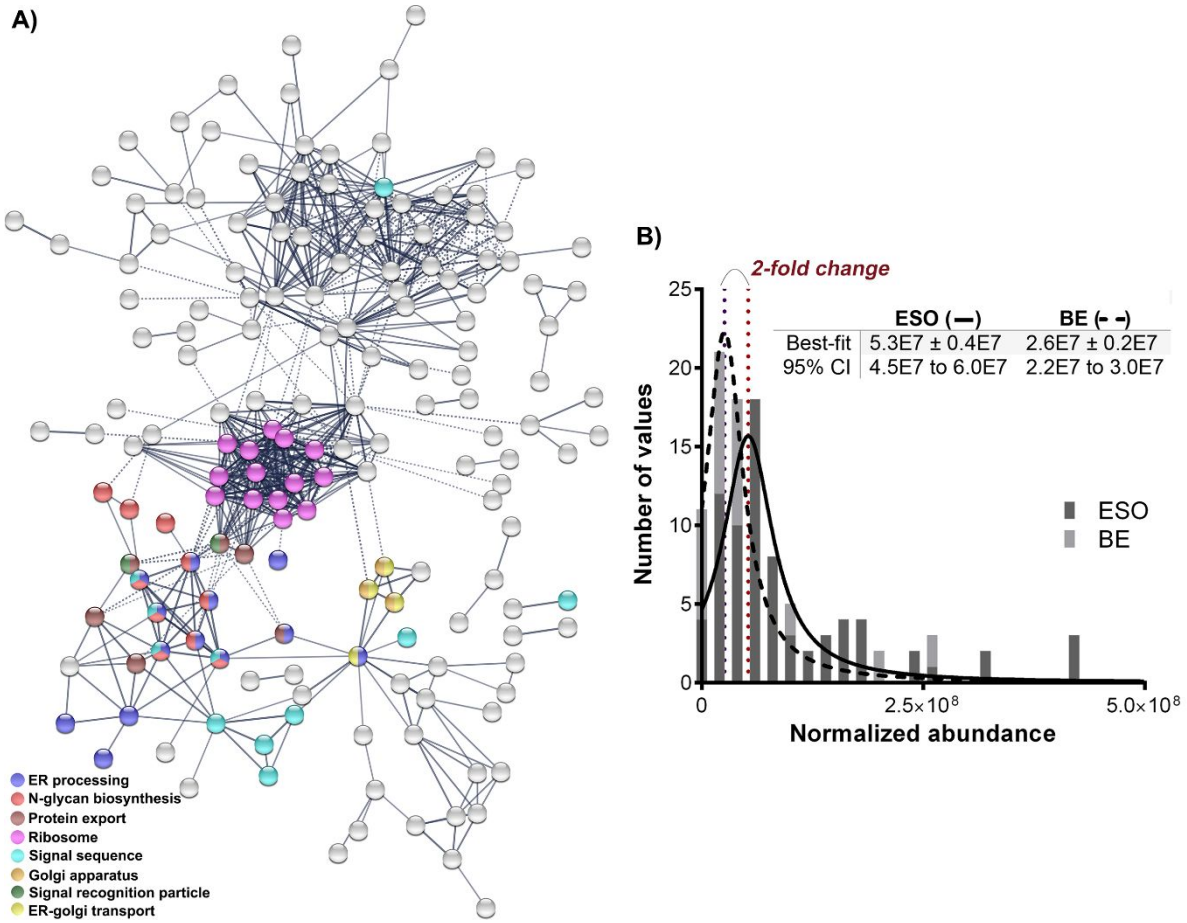


Figure 9. Network featuring protein biosynthesis, glycosylation and secretion in the schistosome esophagus. (A) Diagram highlighting the clustering of protein biosynthesis, glycosylation and secretory pathways (coloured circles) among proteins differentially abundant in the ESO. Blank circles are mostly related to bioenergetics and small molecules metabolism. **(B)** Normalised distribution shows an average 2-fold increase on the abundance of proteins associated to vesicle transport and protein secretion. Best-fit values and 95% confidence interval (CI) were calculated using Lorentzian function.

STRING analysis of proteins differentially abundant in the ESO revealed a network connecting protein biosynthesis, glycosylation and secretory pathways (Figure 9A). Proteins involved in a variety of vesicle-trafficking processes featured prominently in the parasite esophagus. Those included a tetraspanin CD63 receptor (Smp_041460; 4.6x-fold difference), a non-tegument annexin (Smp_155580; 37x-fold), charged multivesicular body protein (Smp_079000; 1.5x-fold) and proteins involved in sphingolipid biosynthesis (*e.g.* longevity assurance gene 1, Smp_147460, 2.4x-fold; ceramidase, Smp_122050, 26x-fold; fatty acid desaturase 2, Smp_132740, 8.7x-fold; ormdl protein, Smp_210370, 1.6x-fold). The enrichment of signal recognition particles (*e.g.* SRP-14, 19, 68, 72; average 1.7x-fold

difference) and signal peptidases (Smp_031730, Smp_024390.2 and Smp_024390.3; average 1.7x-fold) reinforced that the classical secretory pathway is operational in the esophagus. In addition, a number of differentially abundant proteins related to SEC-translocation channels (*e.g.* SECs23, 63, 24a, 16a; average 2.1x-fold difference), Coatomer subunits for COPII- and COPI-vesicle coat (Smp_015090, Smp_031860, Smp_245450 and Smp_124430; average 1.8x-fold difference), translocon-associated proteins (Smp_075870.1, Smp_071600, Smp_074700; average 2.1x-fold difference) and ARF GTPases (Smp_086900, Smp_179610; average 1.8x-fold difference) provides evidence for formation of endoplasmic reticulum/Golgi secretory vesicles. Althogether, the abundance of all those proteins was skewed by a factor of 2-fold towards the parasite esophagus (Figure 9B). Of note, similarities between the esophageal gland and tegument secretion processes led us to inspected our data for 36 proposed tegument signatures [7]. Their average ESO/BE fold difference was 0.95 indicating an even distribution over the worm surface thus reinforcing that the enrichment we observed in the ESO preparation is a true reflection of differences in protein composition.

Gastrodermal secretions are also pinpointed by the comparative analysis

The results of our ESO *vs* BE label-free quantification provided the expression pattern of proteins related to digestive processes over the extent of the esophagus and lower gastrodermis. Classical gut-associated proteins previously identified in the worm's vomitus [21] were confirmed as highly enriched in the BE sample (Figure 10; Supplementary Table S1D), these included cathepsin B1 (Sm31 - Smp_103610; Smp_067060), cathepsin S (Smp_139240), asparaginyl endopeptidase (Sm32 - Smp_075800), lysosome-associated membrane glycoprotein (Smp_162770) and saposin B (Smp_014570).

On the other hand, we unexpectedly observed that a set of proteins thus far attributed to the gastrodermis on the basis of vomitus analysis were equally detected in the ESO and BE

samples. For instance, Acyl-CoA thioesterase, β-D-xylosidase, DJ-1/PARK7-like protease, Nieman Pick C2 (NPC2), vesicle associated membrane protein, lysosomal Pro X carboxypeptidases (Smp_002600 and Smp_071610) and two saposins (Smp_130100 and Smp_016490) did not show statistical differences in relative quantification between the ESO and BE samples. Furthermore, some vomitus proteins were differentially expressed in the BE sample but with only a modest enrichment (≤ 2-fold) compared to the ESO; these included ferritin, apoferritin, α2-macroglobulin, cathepsins C/D, ester hydrolase, serpin and a saposin (Smp_194910). It is noteworthy to mention that inaccurate dissection could lead to contamination of ESO fragments with highly abundant gastrodermis components impairing the identification of sample-specific proteins. However, we did not observe any bias suggesting that the MS-signal intensity was directly linked to the presence or absence of differential expression between ESO and BE (Supplementary Material S5).

In addition to the gut-secreted proteins described by Hall *et al*. (2011), we detected other molecules potentially related to gastrodermal functions, such as a serine carboxypeptidase (Smp_172590) 22x-fold more abundant in the BE sample. An additional saposin B

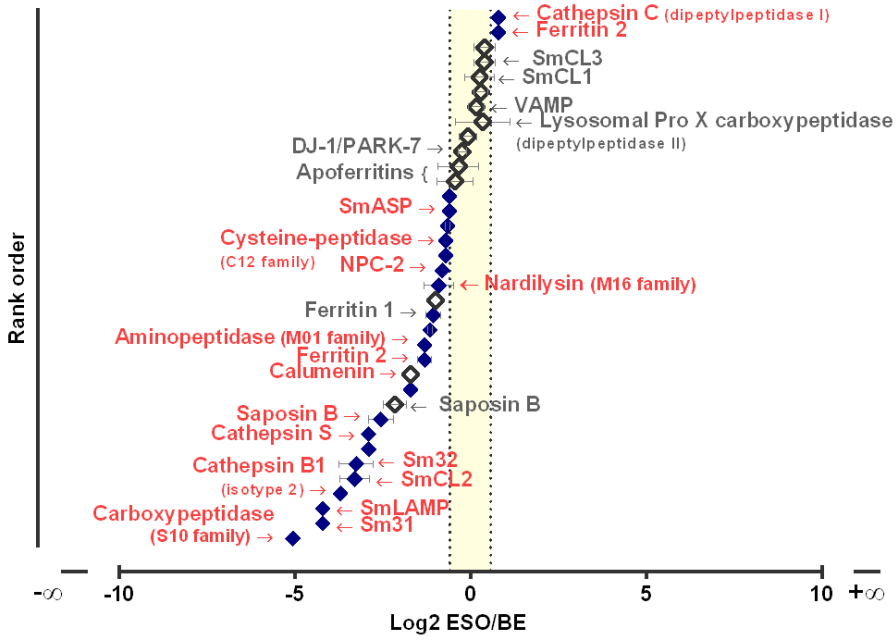


Figure 10. Differential abundance of putative gut secretions in the ESO and BE samples. Diagram showing a biased expression (median difference of 1.7x-fold) of gastrodermal markers towards the schistosome back end (left side). Intriguingly, a group of proteins so far attributed to the gastrodermis exhibited an even distribution between the ESO and BE samples (shaded area). Proteins are rank ordered according to their average Log2 fold ESO/BE. Differential expression supported by statistical analysis (adjusted p-value ≤0.01) are indicated by solid symbols.

(Smp_105420; ~5x-fold difference) was also detected significantly enriched in the BE supported by only one unique peptide (Supplementary Material S5). Furthermore, the expression pattern of cathepsins F/L (Smp_034410, Smp_139160 and Smp_210500), also known as SmCL1, SmCL2, SmCL3, revealed that only the second confidently exhibits higher abundance in the BE sample.

Discussion

In this work we used quantitative proteomics to assess the composition of the *S. mansoni* alimentary tract. This investigation allowed us to designate proteins functionally associated with the anterior and lower regions of the parasite's alimentary tract, namely the esophagus and gastrodermis, represented by the ESO and BE preparations, respectively. The combination of a microdissection procedure and relative quantification of these preparations has permitted the correct assignment of signature proteins to their specific sites of expression. Particularly, proteins encoded by microexon genes known to be expressed in the anterior and posterior esophageal glands were detected by proteomics for the first time, expanding our knowledge of these tissues, thus far limited to transcriptomic analysis [7,8]. This enabled us to acquire further information on gland constituents by absolute quantification, using QconCAT methodology. Lastly, ESO *vs* BE comparisons also provided an informative update of the repertoire of *S. mansoni* gastrodermal markers.

The efficiency of our dissection method for enrichment was firstly confirmed by the ability to detected specific esophageal MEG proteins (*e.g.* MEG-4.1) and VAL-7 in the ESO sample, since evidences gathered from independent molecular analysis such as whole-mount *in situ* hybridisation [28,29], confocal microscopy and RNA-seq [7] verified their specific expression and location in the esophageal gland tissues. The successful detection of the putative esophageal gland secretions can be attributed to substantial enrichment achieved through dissection,

followed by sensitive high-resolution mass spectrometry. Our previous attempts using 2D-PAGE of the whole male worm heads followed by mass spectrometry, failed to detect a single genuine gland secretion [9]. Instead, house-keeping proteins and those massively abundant in the major schistosome tissues (*i.e.* sucker and body wall muscles, and parenchyma) dominated the preparation. Similarly, shotgun proteomics of soluble parasite components using modern LC-MS instrumentation was also inefficient for the detection of such low abundance antigens [11].

Although the exact functions of MEG proteins have yet to be demonstrated, most display peculiar structural elements and low homology outside the *Schistosoma* genus, suggesting they may possess unique roles in worm physiology. The unusual structure of these genes, mainly composed of short (<36bp) and symmetric exons (multiple of three bases), allows the generation of several variants by alternative splicing events without disrupting the reading frame [30], thus preserving the main protein backbone. In fact, the expression of multiple MEG isoforms was confirmed in our data. It has been hypothesised that these variation events could be linked to evasion mechanisms deployed under a selective pressure created by the host immune system [12]. More specifically, the amino acid substitution observed for MEGs 4.1 and 4.2 corroborate the previously reported high synonymous and nonsynonymous substitution rates (dN/dS) in schistosome microexon genes [31]. It is noteworthy that high dN/dS rates have been reported for antigens constantly exposed at host-parasite interfaces (*e.g.* gut secretions and tegument exposed proteins), including the vaccine candidates Sm29 and Tetraspanin-2 (Sm-TSP2), and VAL proteins [31]. This could in fact anticipate that esophageal MEGs are targeted by the host immune response and might constitute a set of new vaccine candidates. However, we must bear in mind that in those cases, sequence variation has to be contemplated during vaccine design since the accumulation of antigenic diversity might impair the outcome of the protective immunization [32].

Faced with a complex repertoire of putative gland secreted proteins, it may be beneficial to design new vaccines against the most abundant candidate molecules. Since the MEG proteins display short sequences, lack conserved protein domains and are likely to possess intrinsically disordered structures [33], it is unlikely these molecules exhibit any catalytic activity. This led us to hypothesise that their abundance in the parasite esophagus may reflect the rate at which these proteins are consumed by interaction with incoming blood. Absolute quantification revealed MEG-12 as the most abundant target protein per cell, among the nine putative gland secretions evaluated in the QconCAT assay. Specific expression of MEG-12 in the anterior esophageal gland of male and female adult worms was confirmed by whole mount *in-situ* hybridization [7], and together these data may suggest it could play an important role in the first contact with incoming blood. So far, MEG-12 has been proposed to function by interacting with erythrocytes and promoting hemolysis through a predicted amphipatic helix [7]. VAL-7 and MEG 4 family proteins, all secreted by the posterior gland, also exhibited a high number of copies per cell and is reinforced by high expression of the cognate transcripts, as already revealed by microarray and quantitative RNAseq [7,29,34].

The esophageal glands and the syncytial gastrodermis are the major secretory structures in the upper and lower alimentary tract, respectively [35]. Our results showed that GO terms related to secretion via vesicle transport are significantly overrepresented among proteins differentially expressed in the ESO proteome when compared with the BE. The remarkable expression of molecules tightly related to vesicle-trafficking such as a tetraspanin CD63 receptor and a non-tegument annexin (Smp_155580), as well as proteins related to sphingolipid biosynthesis, might constitute a route for a therapeutic intervention. In particular, annexins have been proposed as potential drug and vaccine targets because of their important role in secretory processes and exposure to the host immune system [36]. Notably,

disrupting vesicle release in the esophageal lumen has been implicated in the self-cure response of the Rhesus macaque [6].

The highlighted activity of secretory pathways is probably linked to the repertoire of glycosyltransferases differentially abundant in the schistosome esophagus. The enrichment of proteins linked to N- and O-glycosylation in the ESO sample matches the requirement of a specialised machinery for decorating gland products with specific glycan moieties. Notably, the enrichment of beta-1,3-galactosyltransferase and GalNAc transferases in the esophageal region indicates a favourable environment for production of glycoproteins modified with galactosyl (β-1,3) N-acetylgalactosamine, which may explain the strong reactivity of the lectin PNA (peanut agglutinin) towards the esophageal glands [37]. In previous experiments we demonstrated MEG-4.1 as one of the PNA-reactive antigens, based on bidimensional western blotting [9]. Remarkably, MEG-4.1 reactivity was shown split into two spots, indicating either the presence of glycoforms or alternatively spliced sequences. We acknowledge that several O-glycosylation sites predicted for MEG-4.1 are located within a tandemly repeated sequence. Detection of non-decorated peptides derived from this region suggests that glycosylation can be differential. However, the extent of the glycosylation over the repetitive sequence cannot be inferred from bottom-up proteomics. Nevertheless, unceasing production of fully glycosylated forms would require constant availability of glycan precursors. This condition might not be fulfilled since esophageal MEG proteins are the most rapidly synthesised, during *in vitro* turnover experiments (manuscript in preparation). In future investigations it would be beneficial to characterise the extent and type of glycans linked to the protein antigens of interest.

Finally, we interrogated our data for proteins known to be present in the worm vomitus [21]. Although gut secretions dominate these preparations, additional components originated from the esophagus, including MEG proteins (manuscript in preparation) and molecules released

from the membranocalyx were among the secreted molecules. In this context, our quantitative ESO *vs* BE analysis provided additional information regarding the preferential site of expression of many proteins so far attributed to the gastrodermis. Although regurgitation of gut content and imprecise dissection might contribute unwanted contamination by gastrodermal secretions in the esophageal lumen we didn't observe any bias suggesting that our ESO preparation had significant contamination with abundant gut secretions. Classical gut-associated antigens such as Sm31 and Sm32, were highly expressed in the BE, enzymes and transporters such as Acyl-CoA thioesterase, beta-D-xylosidase, NPC2, and some saposins were shown to be equally abundant in both ESO and BE fragments. This might suggest that some hydrolases thought to be exclusive to the gut compartment do in fact function in unrelated worm tissues not resolved by the ESO *vs* BE approach. For instance, this might include lipid and cholesterol binding and transport – by saposins and NPC2 – and iron binding by ferritin, by lysosomes in other tissue locations.

Moreover, we highlight that the ESO *vs* BE comparison provided new information on a group of cathepsins. SmCL1, CL2 and CL3 have been reported as specific gastrodermal constituents [38,39]. However, only SmCL2 was enriched in the BE sample. This might indicate that both SmCL1 and SmCL3 exhibit a more promiscuous biological role in other tissues, besides blood digestion in the parasite intestine. Differences in the pH optima of these enzymes should be considered. Whilst SmCL2 activity is restricted to acidic environments, such as that found in the parasite gut lumen, SmCL1 exhibits a broad pH profile (80% at pH 5-7.2, optimum pH 6.5) [40], similar to SmCL3 [39]. Therefore, our data suggest SmCL2 as the only member among SmCL peptidases, that is solely active in the schistosome gut.

Conclusion

The schistosome feeding process is multistep and takes place within three chambers, clearly demarcated along the alimentary tract, namely the anterior and posterior esophageal lumen and the gut proper [35]. Unlike the gut, proteomic analysis of esophageal secretions produced by the esophageal glands has not previously been attempted since isolation of this microproteome is challenging. In this study we developed a dissection technique enabling the generation of esophageal fragments for detailed characterisation of the *S. mansoni* esophagus. Our results revealed a complex protein composition in the esophageal region with evidence of antigenic variation of MEG-encoded proteins that might have serious implications in immune evasion. Nevertheless, the remarkable abundance of esophageal MEG proteins suggests a high demand for those constituents in the incessant blood processing cascade. In addition, dissection of the esophagus versus the back end provided an overview of differential biological processes enriched in the two body fragments, as well as updating the repertoire of proteins involved in the later steps of blood digestion and nutrient acquisition. Together, our data provide the basis for a better-oriented selection of alimentary tract candidates for vaccine development.

Competing interests

The authors declare that they have no competing interests.

SUPPORTING INFORMATION

The following supporting information is available free of charge at ACS website http://pubs.acs.org

Supplementary Material S1_ Expanded methodology of shotgun proteomic analyses

Supplementary Material S2_ EsoCAT construct, heterologous expression and monitored transitions

Supplementary Material S3_ Identification and relative quantification of MEG proteins

Supplementary Material S4_ Descriptive statistics of QconCAT assay and protein copy per cell estimation

Supplementary Material S5_ Abundance of gastrodermis markers and statistical differences

Supplementary Table S1_ Label-free quantification and tissue signatures.xlsx

Supplementary Table S2_ QconCAT absolute quantification.xlsx

Data availability statement

The shotgun mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository [41] with the dataset identifiers PXD014872 and 10.6019/PXD014872. The SRM analyses are available on Panorama Public at the following link https://panoramaweb.org/PCTZUp.url and ProteomeXchange identifier PXD014899. Skyline exported data for all quantified peptides are available in Supplementary Table S2.

Author's contribution

LXN designed and carried out the experiments, analysed and interpreted data, and wrote the manuscript. RAW and WCB designed experiments, analysed data and wrote the manuscript. PB, VMH, SWH, RJB and CEE assisted experiment design, offered technical expertise and contributed to interpretation of data. RM offered technical expertise and contributed with data interpretation. All authors reviewed the manuscript.

Acknowledgments

The authors acknowledge Fundação Oswaldo Cruz (Centro de Pesquisas René Rachou, Belo Horizonte, Brazil) for providing cercariae for animal infection.

Funding sources

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 - by providing a PhD scholarship to LXN. WCB received funding from Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) - grant numbers APQ-00829-15 and APQ-03367-16. CEE, RJB and PB received funding from Biotechnology and Biological Sciences Research Council (BBSRC) grant number BB/M012557/1.

REFERENCES

(1) World Health Organization. Schistosomiasis http://www.who.int/news-room/fact-sheets/detail/schistosomiasis (accessed Sep 26, 2018).

(2) Kittur, N.; Castleman, J. D.; Campbell, C. H.; King, C. H.; Colley, D. G. Comparison of *Schistosoma mansoni* Prevalence and Intensity of Infection, as Determined by the Circulating Cathodic Antigen Urine Assay or by the Kato-Katz Fecal Assay: A Systematic Review. *Am J Trop Med Hyg* **2016**, *94* (3), 605–610.

(3) Secor, W. E.; Montgomery, S. P. Something Old, Something New: Is Praziquantel Enough for Schistosomiasis Control? *Future Med Chem* **2015**, pp 681–684.

(4) Molehin, A. J.; Rojo, J. U.; Siddiqui, S. Z.; Gray, S. A.; Carter, D.; Siddiqui, A. A. Development of a Schistosomiasis Vaccine. *Expert Rev Vaccines* **2016**, *15* (5), 619–627.

(5) Wilson, R. A.; Langermans, J. A.; van Dam, G. J.; Vervenne, R. A.; Hall, S. L.; Borges, W. C.; Dillon, G. P.; Thomas, A. W.; Coulson, P. S. Elimination of *Schistosoma mansoni* Adult Worms by Rhesus Macaques: Basis for a Therapeutic Vaccine? *PLoS Negl Trop Dis* **2008**, *2* (9), e290.

(6) Li, X. H.; Xu, Y. X.; Vance, G.; Wang, Y.; Lv, L. B.; van Dam, G. J.; Cao, J. P.; Wilson, R. A. Evidence That Rhesus Macaques Self-Cure from a *Schistosoma japonicum* Infection by Disrupting Worm Esophageal Function: A New Route to an Effective Vaccine? *PLoS Negl Trop Dis* **2015**, *9* (7), e0003925.

(7) Wilson, R. A.; Li, X. H.; MacDonald, S.; Neves, L. X.; Vitoriano-Souza, J.; Leite, L. C. C.; Farias, L. P.; James, S.; Ashton, P. D.; DeMarco, R.; et al. The Schistosome Esophagus Is a 'Hotspot' for Microexon and Lysosomal Hydrolase Gene Expression: Implications for Blood Processing. *PLoS Negl Trop Dis* **2015**, *9* (12), e0004272.

(8) Li, X. H.; DeMarco, R.; Neves, L. X.; James, S. R.; Newling, K.; Ashton, P. D.; Cao,

J. P.; Wilson, R. A.; Castro-Borges, W. Microexon Gene Transcriptional Profiles and Evolution Provide Insights into Blood Processing by the *Schistosoma japonicum* Esophagus. *PLoS Negl Trop Dis* **2018**, *12* (2), e0006235.

(9) Li, X. H.; de Castro-Borges, W.; Parker-Manuel, S.; Vance, G. M.; DeMarco, R.; Neves, L. X.; Evans, G. J. O.; Wilson, R. A. The Schistosome Oesophageal Gland: Initiator of Blood Processing. *PLoS Negl Trop Dis* **2013**, *7* (7), e2337.

(10) Skelly, P. J.; Da'dara, A. A.; Li, X. H.; Castro-Borges, W.; Wilson, R. A. Schistosome Feeding and Regurgitation. *PLoS Pathog* **2014**, *10* (8), e1004246.

(11) Neves, L. X.; Sanson, A. L.; Wilson, R. A.; Castro-Borges, W. What's in SWAP? Abundance of the Principal Constituents in a Soluble Extract of *Schistosoma mansoni* Revealed by Shotgun Proteomics. *Parasit Vectors* **2015**, *8* (1), 337.

(12) DeMarco, R.; Mathieson, W.; Manuel, S. J.; Dillon, G. P.; Curwen, R. S.; Ashton, P. D.; Ivens, A. C.; Berriman, M.; Verjovski-Almeida, S.; Wilson, R. A. Protein Variation in Blood-Dwelling Schistosome Worms Generated by Differential Splicing of Micro-Exon Gene Transcripts. *Genome Res* **2010**, *20* (8), 1112–1121.

(13) Orcia, D.; Zeraik, A. E.; Lopes, J. L. S.; Macedo, J. N. A.; Santos, C. R. D.; Oliveira, K. C.; Anderson, L.; Wallace, B. A.; Verjovski-Almeida, S.; Araujo, A. P. U.; et al. Interaction of an Esophageal MEG Protein from Schistosomes with a Human S100 Protein Involved in Inflammatory Response. *Biochim Biophys Acta* **2017**, *1861* (1 Pt A), 3490–3497.

(14) Pratt, J. M.; Simpson, D. M.; Doherty, M. K.; Rivers, J.; Gaskell, S. J.; Beynon, R. J. Multiplexed Absolute Quantification for Proteomics Using Concatenated Signature Peptides Encoded by QconCAT Genes. *Nat Protoc* **2006**, *1* (2), 1029–1043.

(15) Beynon, R. J.; Doherty, M. K.; Pratt, J. M.; Gaskell, S. J. Multiplexed Absolute Quantification in Proteomics Using Artificial QCAT Proteins of Concatenated

Signature Peptides. *Nat Methods* **2005**, *2* (8), 587–589.

(16) Teytelman, L. No More Excuses for Non-Reproducible Methods. *Nature* **2018**, *560* (7719), 411–411.

(17) Bennett, R. J.; Simpson, D. M.; Holman, S. W.; Ryan, S.; Brownridge, P.; Eyers, C. E.; Colyer, J.; Beynon, R. J. DOSCATs: Double Standards for Protein Quantification. *Sci Rep* **2017**, *7*, 45570.

(18) Zhang, J.; Xin, L.; Shan, B.; Chen, W.; Xie, M.; Yuen, D.; Zhang, W.; Zhang, Z.; Lajoie, G. A.; Ma, B. PEAKS DB: De Novo Sequencing Assisted Database Search for Sensitive and Accurate Peptide Identification. *Mol Cell Proteomics* **2012**, *11* (4), M111.010587.

(19) Götz, S.; García-Gómez, J. M.; Terol, J.; Williams, T. D.; Nagaraj, S. H.; Nueda, M. J.; Robles, M.; Talón, M.; Dopazo, J.; Conesa, A. High-Throughput Functional Annotation and Data Mining with the Blast2GO Suite. *Nucleic Acids Res.* **2008**, *36* (10), 3420–3435.

(20) Szklarczyk, D.; Gable, A. L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N. T.; Morris, J. H.; Bork, P.; et al. STRING V11: Protein–Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. *Nucleic Acids Res.* **2019**, *47* (D1), D607–D613.

(21) Hall, S. L.; Braschi, S.; Truscott, M.; Mathieson, W.; Cesari, I. M.; Wilson, R. A. Insights into Blood Feeding by Schistosomes from a Proteomic Analysis of Worm Vomitus. *Mol Biochem Parasitol* **2011**, *179* (1), 18–29.

(22) Eyers, C. E.; Lawless, C.; Wedge, D. C.; Lau, K. W.; Gaskell, S. J.; Hubbard, S. J. CONSeQuence: Prediction of Reference Peptides for Absolute Quantitative Proteomics Using Consensus Machine Learning Approaches. *Mol Cell Proteomics*

2011, *10* (11), M110 003384.

(23) Cheung, C. S.; Anderson, K. W.; Wang, M.; Turko, I. V. Natural Flanking Sequences for Peptides Included in a Quantification Concatamer Internal Standard. *Anal Chem* **2015**, *87* (2), 1097–1102.

(24) Brownridge, P. J.; Harman, V. M.; Simpson, D. M.; Beynon, R. J. Absolute Multiplexed Protein Quantification Using QconCAT Technology. In *Quantitative Methods in Proteomics*; Marcus, K., Ed.; Humana Press: Totowa, NJ, **2012**; pp 267–293.

(25) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. Skyline: An Open Source Document Editor for Creating and Analyzing Targeted Proteomics Experiments. *Bioinformatics* **2010**, *26* (7), 966–968.

(26) Reiter, L.; Rinner, O.; Picotti, P.; Huttenhain, R.; Beck, M.; Brusniak, M. Y.; Hengartner, M. O.; Aebersold, R. mProphet: Automated Data Processing and Statistical Validation for Large-Scale SRM Experiments. *Nat Methods* **2011**, *8* (5), 430–435.

(27) Brownridge, P.; Holman, S. W.; Gaskell, S. J.; Grant, C. M.; Harman, V. M.; Hubbard, S. J.; Lanthaler, K.; Lawless, C.; O'Cualain, R.; Sims, P.; et al. Global Absolute Quantification of a Proteome: Challenges in the Deployment of a QconCAT Strategy. *Proteomics* **2011**, *11* (15), 2957–2970.

(28) Rofatto, H. K.; Parker-Manuel, S. J.; Barbosa, T. C.; Tararam, C. A.; Alan Wilson, R.; Leite, L. C. C.; Farias, L. P. Tissue Expression Patterns of *Schistosoma mansoni* Venom Allergen-Like Proteins 6 and 7. *Int J Parasitol* **2012**, *42* (7), 613–620.

(29) Dillon, G. P.; Illes, J. C.; Isaacs, H. V; Wilson, R. A. Patterns of Gene Expression in Schistosomes: Localization by Whole Mount in Situ Hybridization. *Parasitology* **2007**,

134 (Pt 11), 1589–1597.

(30) Berriman, M.; Haas, B. J.; LoVerde, P. T.; Wilson, R. A.; Dillon, G. P.; Cerqueira, G. C.; Mashiyama, S. T.; Al-Lazikani, B.; Andrade, L. F.; Ashton, P. D.; et al. The Genome of the Blood Fluke *Schistosoma mansoni*. *Nature* **2009**, *460* (7253), 352–358.

(31) Philippsen, G. S.; Wilson, R. A.; DeMarco, R. Accelerated Evolution of Schistosome Genes Coding for Proteins Located at the Host-Parasite Interface. *Genome Biol Evol* **2015**, *7* (2), 431–443.

(32) Xu, X.; Sun, J.; Zhang, J.; Wellems, D.; Qing, X.; McCutchan, T.; Pan, W. Having a Pair: The Key to Immune Evasion for the Diploid Pathogen *Schistosoma japonicum*. *Sci Rep* **2012**, *2* (1), 346.

(33) Lopes, J. L. S.; Orcia, D.; Araujo, A. P. U.; DeMarco, R.; Wallace, B. A. Folding Factors and Partners for the Intrinsically Disordered Protein Micro-Exon Gene 14 (MEG-14). *Biophys J* **2013**, *104* (11), 2512.

(34) Parker-Manuel, S. J.; Ivens, A. C.; Dillon, G. P.; Wilson, R. A. Gene Expression Patterns in Larval *Schistosoma mansoni* Associated with Infection of the Mammalian Host. *PLoS Negl Trop Dis* **2011**, *5* (8), e1274.

(35) Li, X. H.; Wilson, R. A. Alimentary Tract of Schistosoma. In *Schistosoma: Biology, Pathology and Control*; Jamieson, B. G. M., Ed.; CRC Press, **2016**; Vol. 1, pp 239–271.

(36) Hofmann, A.; Osman, A.; Leow, C. Y.; Driguez, P.; McManus, D. P.; Jones, M. K. Parasite Annexins--New Molecules with Potential for Drug and Vaccine Development. *Bioessays* **2010**, *32* (11), 967–976.

(37) Collins, J. J.; King, R. S.; Cogswell, A.; Williams, D. L.; Newmark, P. A. An Atlas for *Schistosoma mansoni* Organs and Life-Cycle Stages Using Cell Type-Specific Markers and Confocal Microscopy. *PLoS Negl Trop Dis* **2011**, *5* (3), e1009.

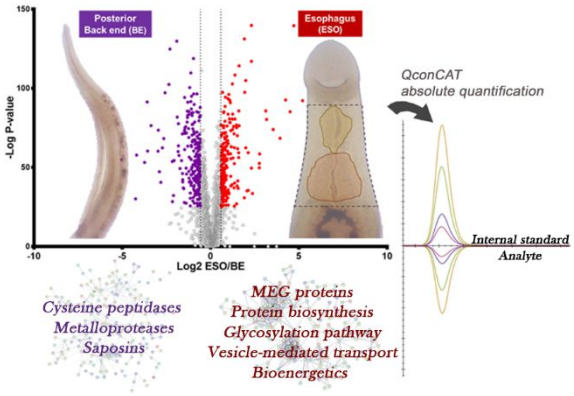
(38) Bogitsh, B. J.; Dalton, J. P.; Brady, C. P.; Brindley, P. J. Gut-Associated Immunolocalization of the *Schistosoma mansoni* Cysteine Proteases, SmCL1 and SmCL2. *J Parasitol* **2001**, *87* (2), 237–241.

(39) Dvořák, J.; Mashiyama, S. T.; Sajid, M.; Braschi, S.; Delcroix, M.; Schneider, E. L.; McKerrow, W. H.; Bahgat, M.; Hansell, E.; Babbitt, P. C.; et al. SmCL3, a Gastrodermal Cysteine Protease of the Human Blood Fluke S*chistosoma mansoni*. *PLoS Negl Trop Dis* **2009**, *3* (6), e449.

(40) Brady, C. P.; Brindley, P. J.; Dowd, A. J.; Dalton, J. P. *Schistosoma mansoni*: Differential Expression of Cathepsins L1 and L2 Suggests Discrete Biological Functions for Each Enzyme. *Exp Parasitol* **2000**, *94* (2), 75–83.

(41) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; et al. The PRIDE Database and Related Tools and Resources in 2019: Improving Support for Quantification Data. *Nucleic Acids Res.* **2019**, *47* (D1), D442–D450.

For TOC Only


Posterior
Back end (BE)
Esophagus
(ESO)
-Log P-value
Log2 ESO/BE
QconCAT
absolute quantification
Internal standard
Analyte
Cysteine peptidases
Metalloproteases
Saposins
MEG proteins
Protein biosynthesis
Glycosylation pathway
Vesicle-mediated transport
Bioenergetics