

2017

# H-IPSE is a pathogen-secreted host nucleus infiltrating protein (infiltrin) 3 expressed exclusively by the *Schistosoma haematobium* egg stage

Luke Pennington

Abdulaziz Alouffi

Evaristus Mbanefo

Debalina Ray

David Heery

*See next page for additional authors*

Follow this and additional works at: [https://hsrc.himmelfarb.gwu.edu/smhs\\_uro\\_facpubs](https://hsrc.himmelfarb.gwu.edu/smhs_uro_facpubs)

 Part of the [Parasitic Diseases Commons](#), [Parasitology Commons](#), and the [Urology Commons](#)

---

## APA Citation

Pennington, L., Alouffi, A., Mbanefo, E., Ray, D., Heery, D., Jardetzky, T., Hsieh, M., & Falcone, F. (2017). H-IPSE is a pathogen-secreted host nucleus infiltrating protein (infiltrin) 3 expressed exclusively by the *Schistosoma haematobium* egg stage. *Infection and Immunity*, (). <http://dx.doi.org/10.1128/IAI.00301-17>

This Journal Article is brought to you for free and open access by the Urology at Health Sciences Research Commons. It has been accepted for inclusion in Urology Faculty Publications by an authorized administrator of Health Sciences Research Commons. For more information, please contact [hsrc@gwu.edu](mailto:hsrc@gwu.edu).

---

**Authors**

Luke Pennington, Abdulaziz Alouffi, Evaristus Mbanefo, Debalina Ray, David Heery, Theodore Jardetzky, Michael Hsieh, and Franco Falcone

1 TITLE

2 **H-IPSE is a pathogen-secreted host nucleus infiltrating protein (infiltrin)**  
3 **expressed exclusively by the *Schistosoma haematobium* egg stage**

4 RUNNING TITLE

5 H-IPSE nuclear infiltrin from *Schistosoma haematobium*

6 AUTHORS

7 Luke F. Pennington<sup>a§</sup>, Abdulaziz Alouffi<sup>b§</sup>, Evaristus C. Mbanefo<sup>c</sup>, Debalina Ray<sup>d</sup>,  
8 David M. Heery<sup>e</sup>, Theodore S. Jardetzky<sup>a</sup>, Michael H. Hsieh<sup>c,f</sup> and Franco H.  
9 Falcone<sup>b</sup>

10 <sup>§</sup> Equal contributors

11 AFFILIATIONS

12 <sup>a</sup>Stanford University School of Medicine, Stanford, CA, USA;

13 <sup>b</sup>School of Pharmacy, Division of Molecular Therapeutics and Formulation,  
14 University of Nottingham, Nottingham, UK;

15 <sup>c</sup>Biomedical Research Institute, Rockville, MD, USA

16 <sup>d</sup>Department of Pathology, University of California, San Francisco, CA, USA

17 <sup>e</sup>School of Pharmacy, Division of Biomolecular Science and Medicinal Chemistry,  
18 University of Nottingham, Nottingham, UK

19 <sup>f</sup>Department of Urology, The George Washington University, Washington, D.C., USA

20

21 CORRESPONDING AUTHOR

22

23

24 Dr Franco H. Falcone. Mailing address: University of Nottingham, School of  
25 Pharmacy, Science Road, Boots Science Building, Nottingham, NG7 2RD, United  
26 Kingdom. Tel: 44 115 84 66073, Fax: 44 115 29 951 5102; E-mail:

27 [franco.falcone@nottingham.ac.uk](mailto:franco.falcone@nottingham.ac.uk)

28

29

30 **ABSTRACT**

31 Urogenital schistosomiasis, caused by the parasitic trematode *Schistosoma*  
32 *haematobium*, affects over 112 million people worldwide. As with *S. mansoni*  
33 infections, the pathology in urogenital schistosomiasis is mainly related to the egg  
34 stage, which induces granulomatous inflammation of affected tissues. *Schistosoma*  
35 eggs and their secretions have been studied extensively for the related *S. mansoni*  
36 organism which is more amenable to laboratory studies. Indeed, we have shown that  
37 IPSE/alpha-1 (M-IPSE herein), a major protein secreted from *S. mansoni* eggs, can  
38 infiltrate host cells. Although M-IPSE function is unknown, its ability to translocate to  
39 their nucleus and bind DNA suggests a possible role in immune modulation of host  
40 cell tissues. Whether IPSE homologs are expressed in other Schistosome species  
41 has not been investigated.

42

43 Here, we describe the cloning of two paralog genes H03-IPSE and H06-IPSE  
44 which are the ortholog of M-IPSE, from the egg-cDNA of *S. haematobium*. Using  
45 PCR and immunodetection, we confirmed that expression of these genes is  
46 restricted to the egg stage and female adult worms, while H-IPSE protein is only  
47 detectable in mature eggs but not adults. We show that both H03-IPSE and H06-  
48 IPSE proteins can infiltrate HTB-9 bladder cells when added exogenously to culture  
49 medium. Monopartite C-terminal NLS motifs conserved in H03-IPSE 'SKRRRKY'  
50 and H06-IPSE 'SKRGRKY' NLS motifs, are responsible for targeting the proteins to  
51 the nucleus of HTB-9 cells, as demonstrated by site directed mutagenesis and GFP  
52 tagging. Thus, *S. haematobium* eggs express IPSE homologs that appear to perform  
53 similar functions in infiltrating host cells.

54

55 **INTRODUCTION**

56 Schistosomes are digenetic blood trematodes, which rely on their egg stage  
57 for transmission to the intermediate host, a water snail (1). In order to reach the  
58 aquatic environment, the eggs deposited by adult female worms in the blood vessels  
59 of their mammalian host have to cross several layers of host tissue before they can  
60 reach the lumen of the gut, or, in the case of *S. haematobium*, the bladder. This is a  
61 critical step in the life cycle of the parasite and is therefore very likely to have been  
62 fine-tuned to the host's immunological and tissue environment during the course of  
63 evolution. The microenvironment of the bladder and the gut are histologically and  
64 immunologically quite different, and this may be reflected in differences between the  
65 molecules produced by the eggs of *S. haematobium* and *S. mansoni* and the  
66 underlying mechanisms leading to translocation of eggs across the tissues. Most  
67 proteomic studies have concentrated on the egg stage of the more available  
68 trematode, *S. mansoni* (2–4); these have identified three major protein components  
69 (3) produced by mature eggs: omega-1 (5), kappa-5 (6) and IPSE/alpha-1 (7).

70

71 Here, we show that *S. haematobium* expresses multiple variants of a protein  
72 homologous to IPSE/alpha-1 in *S. mansoni* (M-IPSE), which we have called H-IPSE,  
73 a term which will be used here to collectively describe the different orthologs of M-  
74 IPSE in *S. haematobium*. The mRNA expression of H-IPSE is restricted to the egg  
75 and female worm stage, but only translated as protein in eggs. H-IPSE shares an  
76 important biological activity described for M-IPSE: the ability to be taken up by and  
77 translocate to the nucleus of host cells.

78

79

## 80 MATERIALS AND METHODS

81 **Cloning of IPSE transcripts.** Total RNA was isolated from *S. haematobium*  
82 eggs from the liver of infected hamsters using TRizol. Following DNase treatment  
83 and inactivation, cDNA was generated from RNA samples using the superscript III  
84 first strand cDNA synthesis kit (Invitrogen) with oligo dT primers, or custom oligo dT  
85 primers with a 5' anchoring sequence corresponding to 3' RACE reverse primers  
86 (See Table S1 in suppl. data). The resulting cDNA was then amplified with either  
87 targeted 5' and 3' primers designed from predicted H-IPSE transcripts, or amplified  
88 by a nested 3' RACE PCR with a 5' primer targeted to a conserved 5' elements in  
89 IPSE using Platinum Taq Supermix (Invitrogen). Gel purified PCR fragments were  
90 then treated with Taq polymerase to facilitate cloning into the pCR 2.1-TOPO vector  
91 included with the TOPO TA cloning kit (Invitrogen) and all positive colonies from the  
92 blue/white assay were sequenced.

93  
94 **Recombinant protein expression in BL21 Star DE3 *E. coli*.** The H06-IPSE  
95 protein coding sequence lacking the N-terminal signal sequence was cloned into the  
96 pET-100D Topo expression vector (Invitrogen) and transformed into BL21 Star DE3  
97 *E. coli*. 1L cell cultures were grown to an OD of 0.6 under ampicillin selection (100  
98 µg/mL), and were induced with 1mM IPTG for 3 hours before harvest. Cell pellets  
99 were then suspended in 50 mL of lysis buffer (6M guanidine hydrochloride, 10 mM  
100 imidazole, 20 mM sodium phosphate, and 500 mM NaCl at pH 7.8) and treated with  
101 EDTA-free protease inhibitor tablets (Pierce). Cells were lysed with three freeze/  
102 thaw cycles and sonicated on ice. Nickel NTA resin purifications were conducted  
103 with a binding buffer containing 8M urea, 10 mM imidazole, and PBS at pH 7.4, a  
104 wash buffer consisting of 8M urea, 25 mM imidazole, and PBS at a pH of 7.4, and an

105 elution buffer containing 8M urea, 300 mM imidazole, and PBS at a pH of 7.4. This  
106 solution was successively dialyzed against PBS pH 7.4 solutions containing 4.0 M,  
107 2.0 M, and 1.0 M urea over three days before being dialyzed overnight against PBS  
108 (see Fig. S4A in suppl. materials). Refolded protein was concentrated to <0.50  
109 mg/mL in a 3.0 kDa cutoff Centricon centrifugal concentrator (EMD Millipore).

110 **Generation of polyclonal IPSE antibodies.** Recombinant bacterial derived  
111 H06-IPSE protein (obtained as described above) was used to immunize rabbits 4  
112 times over the course of 8 weeks (ProMab Biotechnologies Inc.) (Fig. S1B in suppl.  
113 materials). Antibody was precipitated from sera using ammonium sulfate and  
114 suspended in PBS with 0.03% sodium azide prior to use.

115 **Cloning of H-IPSE into pTT5 expression vector.** To facilitate mammalian  
116 expression, codon optimized synthetic IPSE vectors were generated for H03, H06,  
117 and M-IPSE (GeneArt Invitrogen). The IPSE variants were PCR amplified from these  
118 synthetic constructs. A second insert containing the human VEGF signal sequence,  
119 an 8x His tag, and a TEV cleavage site was also amplified. These two fragments  
120 were inserted into an EcoRI/NheI-digested pTTVH8G vector (licensed from the  
121 Canadian Research Council (8)) by Gibson assembly. Subsequently, during vector  
122 optimization, the full IPSE expression cassette with N-terminal signal sequence, tag,  
123 and cleavage site was transferred from the pTTVH8G vector to the pTT5 vector by  
124 conventional restriction cloning using EcoRI and NotI. The H06 SKAAKY NLS  
125 mutant was cloned by site-directed mutagenesis of H06-pTT5 vectors using Phusion  
126 High Fidelity PCR Master mix (Invitrogen), followed by Dpn1 digestion for 1 hr.

127 **RT-PCR and stage specific expression of H-IPSE.** Total RNA was isolated  
128 from the following *S. haematobium* life cycle stages (obtained from the NIAID  
129 Schistosomiasis Resource Center for distribution through BEI Resources, NIAID,

130 NIH): purified eggs, retrieved from the liver of *S. haematobium* infected hamsters,  
131 miracidia, cercariae, schistosomula, adult females, adult males and mixed sex adult  
132 worms using the RNAzol kit (MRC, Ohio), following the manufacturer's instructions.  
133 Contaminating genomic DNA was removed by DNase treatment using TURBO  
134 DNase (Invitrogen Ambion, USA) and chemical DNase inactivation, as per  
135 manufacturer's instructions. After removal of genomic DNA contaminants, cDNA was  
136 obtained by reverse transcription using iScript cDNA synthesis kit (Bio-Rad, USA) as  
137 directed by the manufacturer. After reverse transcription, RNA sample  
138 concentrations were measured using a Nanodrop ND-1000 spectrophotometer  
139 (ThermoScientific Fisher) and adjusted to 500 ng/ $\mu$ L for all samples with molecular  
140 grade water. The PCR was performed on a BioRad CFX Connect thermocycler,  
141 using the following cycling conditions:

142 Initial denaturation step [2 min at 94 °C], followed by 35 cycles of denaturation [30  
143 sec at 94 °C], annealing [45 sec at 56 °C] and extension [1 min at 72 °C], and  
144 followed by a final extension [5 min at 72 °C]. The polymerase used was Takara Taq  
145 polymerase using 2  $\mu$ L 10x polymerase buffer, 2  $\mu$ L 10  $\mu$ M dNTPs, 1  $\mu$ L each of  
146 forward (5'-GCTCACTCTCACCACCATG-3') and reverse (5'-  
147 TCCTTCGACGTTTCGATTAC-3') primers, 2  $\mu$ L of cDNA template and 11.5  $\mu$ L  
148 molecular grade water in a 20  $\mu$ L total volume. PCR reactions were then subjected to  
149 electrophoretic separation on 1% agarose containing 0.5  $\mu$ g/mL ethidium bromide in  
150 0.5x Tris Borate EDTA (TBE) buffer. A 100 bp DNA ladder (Promega, USA) was  
151 used for sizing of the amplicon. Gels were imaged using a GelDoc XR Molecular  
152 Imager (BioRad, USA) and saved as tif files. The oligonucleotide primers used span  
153 an intron, discriminating cDNA (295 bp) from genomic DNA (402 bp).

154



155        **Recombinant protein expression using HEK 293-6E cells.** H03-IPSE and  
156 H06-IPSE were recombinantly expressed for uptake and microscopy experiments  
157 using the pTT5 HEK293-6E expression platform (L-11565) licensed from the  
158 Canadian Research Council (8). A large-scale gene expression workflow was  
159 developed using 2L vented shaking flasks. Cells were cultivated in suspension in an  
160 incubator at 37 °C in 5% CO<sub>2</sub> humidified atmosphere under constant shaking at a  
161 rate of 120 rpm. The medium consisted of Freestyle F17 medium (GIBCO, Rockville,  
162 MD) supplemented with 0.1% w/v Kolliphor P-188 (Sigma-Aldrich), with 4 mM L-  
163 glutamine and 25 µg/mL G418 (ThermoScientific Fisher). Freestyle 293 medium  
164 (GIBCO, Rockville, MD) was used interchangeably with F17 without glutamine  
165 supplementation.

166        For transient transfection, 500 mL of cell suspension were mixed with 12.5 mL  
167 of medium containing the plasmid DNA and another 12.5 mL of medium for  
168 resuspension of linear 25 kDa polyethylenimine (PEI) (PolyPlus). The final DNA  
169 amount for each pTT construct was 0.5 mg, and this was with mixed 1.5 mg of PEI  
170 (3:1 PEI:DNA ratio) and incubated for 3 minutes at room temperature. The resultant  
171 complex was then added to the cells. 2.5 mL of 20% (w/v) Tryptone N1 (TN1,  
172 TekniScience Inc., Canada) was added 24 h after transfection.

173  
174        H-IPSE wildtype or mutant proteins, secreted into the HEK293-6E serum-free  
175 cell culture medium, were harvested 7 days after transfection, followed by protein  
176 purification. This supernatant was centrifuged at 2,800 x g for 10 minutes (4°C),  
177 followed by 0.22 µm filtration to remove cell debris and aggregates, then purified by  
178 immobilized metal affinity chromatography (IMAC) using TALON Superflow cobalt  
179 affinity resin (GE Healthcare, Freiburg, Germany) or Ni-NTA Agarose (Qiagen). For

180 cobalt resin purification the binding buffer consisted of 50 mM sodium phosphate,  
181 300 mM NaCl, pH 7.4; the wash buffer contained 50 mM sodium phosphate, 300 mM  
182 NaCl, 5 mM imidazole, pH 7.4, while the elution buffer consisted of 50 mM sodium  
183 phosphate, 300 mM NaCl, 150 mM imidazole, pH 7.4. Nickel resin purifications  
184 were conducted with a binding buffer containing PBS pH 7.4, a wash buffer  
185 consisting of PBS and 10mM imidazole at a pH of 7.4, and an elution buffer  
186 containing PBS and 300mM imidazole at a pH of 7.4.

187

188 **SDS-PAGE, Coomassie stain and Western blotting.** 15  $\mu$ L aliquots of the  
189 eluted purified protein fractions were separated by SDS-PAGE using Bio-Rad Mini-  
190 Protean ready gels (4–20% gradient TGX) in a Mini-Protean electrophoresis cell, as  
191 recommended by the manufacturer. The gels were then incubated for half an hour in  
192 Instant Blue (Expedeon, Harston, UK) for Coomassie staining, followed by washing  
193 in deionised water. For Western Blotting, the gradient TGX gels were transferred to a  
194 0.2  $\mu$ m nitrocellulose membranes using Trans-Blot® Turbo™ Transfer System, as  
195 per the manufacturer's protocol (Bio-Rad). The membranes were blocked by  
196 blocking buffer (5% (w/v) dried skimmed milk, 0.01% (v/v) Tween 20 and TBS) with  
197 shaking for 1 hour at room temperature. Next, membranes were incubated with the  
198 primary mouse anti-His antibody (GE Healthcare) as primary antibody, which was  
199 diluted 1:5000, at 4°C overnight, followed by three washes in Tris-buffered saline  
200 solution (TBS) containing 1% Tween (T) for 10 minutes each. The membranes were  
201 then incubated with anti-mouse IgG (whole molecule), HRP-conjugated antibody  
202 (Sigma-Aldrich, UK) as a secondary antibody (1:4000) for one hour at room  
203 temperature, followed by washing in the manner described above. Membranes were  
204 imaged using a Fuji LAS4000 imager with chemiluminescence-luminol reagent (3  $\mu$ L

205 of 30% H<sub>2</sub>O<sub>2</sub>, Tris/HCl 0.1 mM, pH, 8; 2.5 mM luminol and 400 μM coumaric acid).  
206 For detection of native IPSE protein in adult worm extract and egg-derived samples  
207 approximately 40 μg of parasite-derived material was loaded per well as determined  
208 by each samples' A<sub>280</sub> using a spectrophotometer. These samples were run on 4-  
209 20% ExpressPlus Page gels (Genescript) as directed by manufacturer, and stained  
210 with Coomassie Brilliant Blue G250 (BioRad). For western blotting gels were  
211 transferred to a 0.22 μm PDVF membrane after preactivation with methanol. The  
212 membranes were blocked by blocking buffer (5% (w/v) dried skimmed milk, 0.01%  
213 (v/v) Tween-20 and PBS) with shaking for 1 hour at room temperature. Next, blots  
214 were incubated with polyclonal anti-H06-IPSE rabbit antibodies at a 1:500 dilution  
215 and stained overnight at 4°C. After three washes in PBS-T, blots were then  
216 developed with polyclonal HRP conjugated goat anti-rabbit secondary antibody  
217 (EMD Millipore) at a dilution of 1:5000. The gels were washed 3 additional time in  
218 PBS-T and developed with SignalFire ECL reagent (Cell Signaling).

219  
220 **Cloning of predicted NLS and mutants into pTetra-EGFP.** The predicted  
221 NLS for each protein was subcloned into the Tetra-EGFP vector, which carries a  
222 kanamycin resistance gene (9). This vector encodes four EGFP repeats with a  
223 multiple cloning site inserted between the third and fourth EGFP sequence. The  
224 nucleotides encoding the predicted NLS in H03-IPSE and H06-IPSE, as well as  
225 predicted NLS mutant and the canonical Sv40 NLS, were inserted into pTetra-EGFP  
226 using oligonucleotide primers and specific restriction enzymes (see Table S1 in  
227 suppl. mat). This leads to a construct, which codes for a tetra-EFGP fusion protein of  
228 approximately 113 kDa that due to its large size is completely excluded from the  
229 nucleus in the absence of a functional NLS. Initially, 5'-phosphorylated pairs of

230 matching oligonucleotides were designed to code putative NLSs containing GATC  
231 overhangs by using 1  $\mu$ L of each oligo (100 $\mu$ M) mixed with 98  $\mu$ L of 10 mM Tris-HCl,  
232 1 mM EDTA, pH 8.0 followed by denaturation at 95°C for 7 minutes, then 3 minutes  
233 at 5°C. Next, the double-stranded oligonucleotides and the pTetra-EGFP vector were  
234 digested with the restriction enzyme *Bgl*II (New England Biolabs), according to the  
235 manufacturer's protocol. The ends of the linearized pTetra-EGFP vector were  
236 dephosphorylated with Antarctic Phosphatase (New England BioLabs) to avoid re-  
237 ligation to itself, following the manufacturer's instructions. Ligation was performed  
238 using 1  $\mu$ L vector, 3  $\mu$ L insert, 1  $\mu$ L 10X Buffer T4-ligase, 4  $\mu$ L molecular biology  
239 grade water and 1 U T4 DNA ligase (Promega) in a 10  $\mu$ L reaction mixture and  
240 incubated overnight at 16°C. DNA sequencing (Source BioScience, UK) using T7  
241 primers determined successful insertion in the correct orientation.

242

243 **HTB-9 Cell culture and transfection.** The human bladder cancer cell line  
244 HTB-9 (5637 ATCC) was grown in T75 flasks (Sarstedt, Germany) at 37°C in a  
245 humidified 5% CO<sub>2</sub> incubator, with Minimum Essential Medium Eagle (MEM; Sigma-  
246 Aldrich) supplemented with 5% heat-inactivated fetal bovine serum (FBS, GIBCO), 2  
247 mM L-glutamine, 100 unit/mL penicillin and 100  $\mu$ g/mL streptomycin (Sigma-Aldrich,  
248 UK). Transient transfections of HTB-9 cells were performed using X-tremeGENE9  
249 DNA transfection reagent (Roche Applied Science, Germany) according to the  
250 manufacturer's protocol. Cells were plated onto 5 mg/mL rat-tail collagen I-coated  
251 glass cover slips (Invitrogen, UK, 15mm diameter, # 1 thickness) in 6-well plates,  
252 and transfected with the different Tetra-EGFP plasmids at 60-70% confluency.

253

254        **Cell Fixation and Fluorescence Microscopy.** One-day after transfection the  
255 cells were washed with Dulbecco's phosphate-buffered saline (DPBS, Gibco) and  
256 fixed at room temperature for 10-15 min in 4% paraformaldehyde. The cells were  
257 then washed three times with DPBS and incubated with 0.5 g/mL Hoechst 33342  
258 stain (Sigma-Aldrich) at room temperature for 8-15 minutes, before washing again  
259 three times with DPBS. Slides were mounted with mounting medium (Sigma-  
260 Aldrich). The transfected cells were visualized by fluorescence microscopy (EVOS *fl*,  
261 Advanced Microscopy Group, USA) or confocal microscopy (LSM510 META, ZEISS,  
262 Germany), and analyzed using Zeiss LSM Image Browser software (version  
263 4.2.0.121).

264

265        **Cellular Uptake of H03-IPSE.** HTB9 cells were seeded onto Lab-Tek 8-well  
266 chambered cover glass (Nalgene Nunc International) at a density of  $5 \times 10^5$  cells in  
267 order to achieve 50-60% confluency after 24 hours. Cells were then incubated with  
268 15-0.40 nM of recombinant proteins (H03-IPSE WT or mut) in serum-free  
269 internalization medium (HEPES buffered Ham's F12 medium (SigmaAldrich, UK)  
270 containing 10 mM  $\text{NaHCO}_3$  and 2 mg/mL (bovine serum albumin) (Fraction V Biomol  
271 GmbH, Germany), followed by fixation at room temperature for 10-15 min in 4%  
272 paraformaldehyde solution. The cells were then washed 4-5 times with DPBS and  
273 incubated with 0.5  $\mu\text{g/mL}$  Hoechst 33342 or 5  $\mu\text{M}$  DRAQ5 (Thermo Fisher Scientific)  
274 nuclear stain for 15 min and permeabilized with 0.2% Triton X-100 in DPBS for 10  
275 min. The cells were washed 4-5 times with DPBS and incubated separately at room  
276 temperature for 30 min with two different primary antibodies diluted 1:5000, either  
277 mouse anti-His antibody (GE Healthcare). The cells were then washed thrice and  
278 labelled with the secondary antibody, Alexa Fluor 555-conjugated goat anti-mouse

279 IgG (H+L; Molecular Probes), diluted 1:500 by incubation at room temperature for 30  
280 min, followed by three final washes with DPBS.

281

## 282 RESULTS

283

284 **Identification of H-IPSE variants.** To identify homologs of *S. mansoni* IPSE in  
285 the *S. haematobium* genome, we performed a BLAST search at Wormbase.org  
286 using the predicted transcript of M-IPSE (Smp\_112110.1) (10–12). This analysis  
287 identified three paralogs within the *S. haematobium* genome, all with predicted  
288 transcripts with high identity to M-IPSE (Gene ID, % amino acid identity: C\_00050,  
289 67%: C\_00244, 63%: B\_00796, 56%). To verify these transcripts we isolated cDNA  
290 from *S. haematobium* eggs, and employed two strategies. The first employed 5' and  
291 3' primers designed to amplify transcripts predicted from the C\_00244 locus, and the  
292 second employed a 3' RACE cloning strategy with a 5' primer targeting highly  
293 conserved regions of all H-IPSE variants (Table S1). In total, 14 IPSE transcript  
294 sequences were obtained, and 8 Sanger sequencing runs contained data sufficient  
295 for unambiguous base calling throughout the ORF`s. These transcripts clustered with  
296 two of the three H-IPSE paralogs (Figure 1A), and one transcript from each cluster  
297 (H03 and H06) was selected for further study. Sequence variations within these  
298 clones suggest that H-IPSE genes are polymorphic (see Figures S1-S3 for  
299 supplementary information).

300

301 <Figure 1 here>

302

303 Alignment of H03-/H06-IPSE with M-IPSE (Fig. 1B) demonstrates conservation  
304 of the seven cysteines, known to form three intramolecular disulfide bonds and one  
305 intermolecular bond, resulting in a homodimeric structure. Two potential N-linked  
306 glycosylation consensus motifs are also present with small variations. A 20 amino  
307 acid long N-terminal classical secretory sequence (CSS) is predicted for both H03-  
308 and H06-IPSE by SignalP 4.1 (13). To verify the presence of IPSE protein in *S.*  
309 *haematobium* parasite-derived material, we expressed and refolded H06-IPSE from  
310 insoluble inclusion bodies, and used this bacterially derived H06-IPSE to generate  
311 polyclonal anti-H-IPSE antibodies in rabbits (Figure S1). On western blots, anti H-  
312 IPSE antibodies bind a ~40 kDa protein species in both *S. haematobium* egg  
313 secreted protein and soluble egg antigens, but not in adult worm extract (Figure 1C).  
314 This corresponds to the expected size for dimeric, glycosylated H-IPSE variants.

315

316 **H03- and H06-IPSE variants have a predicted nuclear localization**  
317 **sequence (NLS).** Several algorithms (cNLSMapper (14), pSORTII (15)  
318 NLStradamus (16) and NucPred (17)) predict a C-terminal nuclear localization  
319 sequence (NLS) in H03-IPSE close to the C-term (data not shown), similarly to the  
320 NLS described for M-IPSE (18). Intriguingly, the H06-IPSE paralog carried an  
321 R128G variant within the nuclear localization sequence corresponding to the  
322 validated NLS (18) in M-IPSE (Fig. 1B). Such variants have also been observed in *S.*  
323 *mansoni* studies. For example ESP3-6, a protein later recognized as an M-IPSE  
324 variant, is 97% homologous to the published M-IPSE sequence (GenBank Acc. Nr.:  
325 AAK26170.1), but contains an R132L variant within the NLS (ESP3-6 GenBank Acc.  
326 Nr.: AF527011). Positively charged amino acids in an NLS are key to its nuclear  
327 targeting activity mediated by binding to cytosolic importin- $\alpha$  (19), thus such a

328 replacement will have an impact on the protein's ability to translocate to the nucleus,  
329 ranging from less efficient translocation to no translocation at all, depending on the  
330 exact position in the NLS (19). Substitutions in the NLS will potentially also have an  
331 effect on DNA binding specificity. This is also reflected in the less certain prediction  
332 of an NLS in H06-IPSE by cNLSmapper and the other tested algorithms (data not  
333 shown).

334

335 **The NLS in H03- and H06-IPSE is able to direct a large fluorescent protein**  
336 **to the nucleus of mammalian cells.** Therefore, in order to assess functionality of  
337 the predicted NLS in H03-IPSE and the potential impact of the R128G substitution in  
338 H06-IPSE, we cloned the oligonucleotides encoding the NLS predicted sequences  
339 into the previously described Tetra-EGFP vector (18). The resulting constructs were  
340 then transfected into HTB-9 uroepithelial carcinoma cells as a model of host cells  
341 relevant to *S. haematobium* infection. The results of the transfection of the Tetra-  
342 EGFP vector encoding the NLS sequences of H03-IPSE (SKRRRKY) and H06-IPSE  
343 (SKRGRKY), as well as the predicted NLS mutant SAAGAAY (Figure 2) confirm that  
344 the H03-IPSE NLS is fully functional, resulting in complete translocation of the large  
345 Tetra-EGFP protein into the nucleus. In the H06-IPSE NLS, the presence of an  
346 uncharged G in the charged KRRRK H03-IPSE core sequence of the NLS appears  
347 to weaken its strength as nuclear targeting signal, as documented by the mixed  
348 cytosolic/nuclear localization in contrast to the exclusive nuclear localization with the  
349 SKRRRKY H03-IPSE NLS sequence (Fig. 2B). This difference is consistent with the  
350 results obtained with different prediction algorithms (not shown). Substitution of  
351 Lysine and Arginine with Alanine (SAAGAAY) results in a non-functional NLS, which  
352 is no longer able to translocate the Tetra-EGFP protein into the nucleus (Fig. 2A and



353 B). These results clearly show that while the predicted NLS in H03-IPSE is fully  
354 functional, the G substitution in the positively charged core compromises this  
355 function at least in part. Further substitutions almost fully ablate NLS functionality.

356

357 <Figure 2 here>

358

359 **Recombinant M-IPSE and H-IPSE can be expressed in high yields in**

360 **HEK293-6E cells grown in suspension.** For subsequent experiments, we cloned

361 and expressed H-IPSE using HEK293-6E cells. This system uses a serum-free

362 medium adapted clone and allows HEK293 cell culture in suspension, enabling high

363 cell densities and recombinant protein yields. Proteins generated using these cells

364 are glycosylated, which more closely parallels glycosylation of native proteins.

365 Recombinant protein can be harvested from culture supernatants after transient

366 transfection and can be purified e.g. via IMAC using the 8xHis-Tag in the construct

367 (Figure 3A).

368

369 <Figure 3 here>

370

371 Results shown in Figure 3B demonstrate successful expression of H03-IPSE as

372 a mostly dimeric protein of approximately 38-40 kDa molecular weight, with small

373 amounts of monomeric protein of about 20 kDa, in line with what we have previously

374 described for M-IPSE (1,2). The double bands are presumably due to glycosylation

375 variants, which are well described for M-IPSE (21). Purity in eluted fractions after

376 IMAC was high, and did not require any additional purification steps for downstream

377 experiments. In our hands, IPSE proteins produced a range of yields, with H03-IPSE

378 exhibiting the lowest final yield (~5-10 mg/L) and M-IPSE and H06-IPSE SKAAAKY  
379 NLS mutant producing the highest yield (~15-25 mg/L) (Figure S5). Attempts to  
380 concentrate protein to higher concentration above 0.5 mg/mL resulted in formation of  
381 aggregates, not seen by SDS, but appearing in size-exclusion chromatography.

382 **Recombinant H03-IPSE added exogenously is taken up by HTB-9**  
383 **uroepithelial cells and efficiently translocates to the nucleus.** To mimic *S.*  
384 *haematobium* infection conditions, we added recombinant 8xHis tagged H03-IPSE to  
385 the culture medium of proliferating HTB-9 uroepithelial carcinoma cells as described  
386 in Materials and Methods. After 24 hours, fixed cells were immunostained with anti-  
387 His tag antibody. As shown in Figure 4, this revealed a highly efficient uptake of  
388 exogenous H03-IPSE, which was present in HTB-9 nuclei, as revealed by co-  
389 staining of nuclear DNA with DRAQ5. The nuclear staining pattern suggests that  
390 H03-IPSE was largely excluded from the nucleolar regions. This result indicates that  
391 H03-IPSE protein can infiltrate the vast majority of cells with remarkable efficiency  
392 and localize to the nucleus. The only cells we observed that did not stain for  
393 recombinant H03-IPSE were those actively undergoing mitosis, in which the nuclear  
394 membrane is broken down

395

396 <Figure 4 here>

397

398 **The NLS in H03- and H06-IPSE is essential for nuclear translocation, but**  
399 **not for cellular uptake.** Next, we compared the ability of both H-IPSE variants with  
400 the ability of the NLS Alanine mutant, to gain access to the nuclear compartment of  
401 HTB-9 cells when added to cell culture medium. The results, visualized on a  
402 fluorescence microscope, are shown in Figure 5. When assessing the subcellular

403 localization of the recombinant molecules with a molecular weight of approximately  
404 40 kDa, both H03- and H06-IPSE variants seemed similarly efficient in translocating  
405 across the nuclear membrane.

406

407 <Figure 5 here>

408

409 Unlike what was observed for the Tetra-EGFP-NLS constructs (Fig. 2A and B),  
410 we did not find any reduction in translocation efficiency in H06-IPSE compared with  
411 H03-IPSE, suggesting that these differences may only become apparent with larger  
412 proteins (such as Tetra-EGFP) and thus may not be relevant in the context of  
413 molecular crosstalk between H-IPSE and the host cells. In contrast, the SKAAAKY  
414 NLS H03 mutant, despite potentially being able to cross nuclear pores due to its low  
415 molecular weight, remained completely excluded from the nucleus (Fig. 5 H03-IPSE  
416 mut). The H03 mutant appears to be located in vacuoles or endosome-like  
417 structures, mainly located around the nucleus, rather than diffuse in the cytoplasm.  
418 This suggests that an intact NLS might be an important feature needed e.g. for  
419 endosomal escape.

420 The lack of uptake of the Ala mutant shows that the NLS in H-IPSE is  
421 monopartite and can be described as necessary and sufficient, i.e. no nuclear  
422 translocation occurs in its absence, and it is the only NLS in the molecule. This is  
423 consistent with the lack of prediction of additional nuclear translocation signals  
424 elsewhere in the molecule. These data also demonstrate that the NLS is not required  
425 for H-IPSE transport into cells.

426

427 **H-IPSE mRNA is expressed in *S. haematobium* adult females and the egg**  
428 **stage.** Next, we investigated the expression of H-IPSE across the different life cycle  
429 stages, using conventional RT-PCR. The results are shown in Figure 6. RT-PCR  
430 data indicated adult worm cDNA preparations were more contaminated with genomic  
431 DNA, in comparison with other life cycles in control experiments; however, DNase  
432 treatment completely removed genomic DNA. Using DNase-treated samples, RT-  
433 PCR indicated expression of H-IPSE mRNA in purified eggs, female adult worms  
434 (AdF), and weak expression in mixed gender adult worms (Ad Mix), but none in  
435 cercariae (cer), schistosomula (som), miracidia (mir) or male adult worms (AdM).

436

437 <Figure 6 here>

438

439 Overall, this result is consistent with an expression pattern restricted to the egg  
440 stage (since female worms often contain immature eggs), and similar to what was  
441 described for M-IPSE (7). Interestingly, despite the detection of H-IPSE transcripts in  
442 AdF, we detected no H-IPSE protein by western blot in AWA preparations (Figure  
443 1b).

444

445

## 446 DISCUSSION

447 The concurrent presence of a classical secretory sequence (CSS) and a  
448 nuclear localization sequence (NLS) on the same protein, two apparently  
449 contradictory signals in terms of subcellular targeting, is a rare feature. Only four of  
450 19 tested algorithms correctly identified the presence of a C-terminal, monopartite  
451 NLS in H03-IPSE (cNLSmapper (14), PSORT II (15), NLStradamus (16) and

452 NucPred (17)), while most other tested programs predicted a secretory pathway,  
453 some unexpectedly also after removal of the signal sequence. Thus prediction of an  
454 NLS by algorithms is still insufficient and such predictions need to be verified  
455 experimentally. At least three properties need to be fulfilled in order to confirm  
456 functionality of an NLS; a) a functional NLS needs to be able to direct the protein to  
457 the nucleus, and b) the NLS in isolation should also be able to direct heterologous  
458 proteins to the nucleus. Finally, c) mutation of one or several basic amino acids  
459 should lead to loss of translocation, or in the case of a single substitution, at least  
460 weakening of NLS functionality.

461

462 Our data confirm that the nuclear localization signals present in H03- and H06-  
463 IPSE are functional and essential for translocation of IPSE into the nuclei of host  
464 cells. This is somewhat surprising, as 40 kDa is well below the known limit for  
465 passive diffusion across the nuclear pore complex, which has been described as  
466 'quite larger than 60 kDa' (23). A possible explanation for such behavior could be yet  
467 to be characterized interactions with cellular structures or soluble proteins in the  
468 cytosol, making the resulting complex too large for passive diffusion into the nucleus.

469

470 Kosugi and coauthors described six classes of NLS with different specificities  
471 for the binding grooves of the karyopherin importin  $\alpha$  (19). Based on their  
472 classification, the H03-IPSE SKRRRKY would be considered a Class I classical  
473 NLS, characterized by a stretch of at least four consecutive basic amino acids (either  
474 K or R). In contrast, the SKRGRKY NLS in H06-IPSE conforms with being a Class II  
475 classical NLS with the consensus sequence K(K/R)X(K/R), in which one non-basic  
476 amino acid interrupts the adjacency of the basic amino acids found in Class I signals,

477 reducing the signal to an interrupted sequence of three basic amino acids. The only  
478 difference from Kosugi's predicted canonical patterns is that both H03- and H06-  
479 IPSE appear to possess one additional basic residue in their NLS. Both classes  
480 would be predicted to bind to the large major binding pocket of Importin  $\alpha$  (19).

481

482       Having shown that the NLS in H-IPSE is fully functional, and bearing in mind  
483 that the sequence cannot be functional within the schistosome eggs themselves, as  
484 the N-terminal CSS will target the protein for secretion well before the N-terminal  
485 NLS is synthesized, the key question is what the biological function of such a protein  
486 might be. The restriction of its expression to a single stage of the parasitic life cycle  
487 (the egg stage), which is in line with the M-IPSE homolog's restricted expression (7),  
488 suggests a specialized function needed only during a specific phase of egg  
489 embryogenesis, or a function needed to govern the very important interaction with  
490 the host cells and tissues. The former is unlikely due to the secretory nature of the  
491 molecule. To further explore this possibility, we first need to summarize the fate of  
492 eggs after oviposition by the female schistosomes. Newly deposited eggs do not  
493 have the complex morphology found in mature eggs; this has been described in  
494 detail for *S. mansoni* by Neill (24), Ashton (25), Jurberg (26) and their coauthors.  
495 Fewer details are available regarding *S. haematobium*. Immature eggs are smaller  
496 and characterized by the absence of the two envelopes surrounding the miracidia in  
497 mature eggs: the outer envelope (Reynold's layer, RL) and the inner envelope (von  
498 Lichtenberg's layer, vLL). The RL is enriched with tightly packed rough endoplasmic  
499 reticulum structures and is therefore thought to be a major site of protein synthesis.  
500 Under the electron microscope, the RL appears richly surrounded by granular  
501 materials in a 1  $\mu\text{m}$  wide space underneath the egg shell (25). Eggs deposited *in vitro*

502 by *ex vivo* worms have been shown to take about a week to fully develop into  
503 mature, infective miracidia (27), but development in the host is likely to be more  
504 rapid.

505 The production of the M-IPSE in *S. mansoni* eggs has been clearly shown to occur in  
506 the subshell area within the fully formed RL and vLL. This has been demonstrated by  
507 immunohistochemical staining with a monoclonal antibody to M-IPSE, and by in situ  
508 hybridization with labeled antisense transcripts of full-length M-IPSE cDNA (7). M-  
509 IPSE can also be seen in contact with the tissues around the eggs (confirming that it  
510 is secreted by the eggs) (20) and has been also seen inside surrounding host cells  
511 (7). Thus, the emerging picture is that immature eggs initially do not produce IPSE,  
512 but that this protein is produced as eggs mature while migrating through the tissues,  
513 releasing it into the tissues, where it is able to enter host cells. Inside host cells, it  
514 rapidly (in less than half an hour (18)) translocates to the nucleus, where it binds  
515 DNA (F. Falcone, unpublished data) with yet-to-be described downstream effects.  
516 It is very clear however, that translocation across the host tissues, in order for the  
517 eggs to reach the lumen of the bladder or gut, is an event of paramount importance  
518 in the life cycle of the parasite. Hence, it can be assumed that molecules secreted  
519 exclusively by more mature forms of this life stage may be involved in the egress  
520 process, and have evolved under high evolutionary pressure.

521 The first step after oviposition is the escape of eggs from the venules in which  
522 they were deposited; *in vitro*, human cells obtained from umbilical venous endothelial  
523 cells (HUVECs) rapidly overgrow eggs directly oviposited onto a monolayer (within 4  
524 hours) and a similar behavior is seen when eggs are inserted into umbilical veins  
525 (28). More recently, de Walick and coauthors demonstrated deposition of van  
526 Willebrand factor and other prothrombic plasma proteins onto the eggshell of *S.*

527 *mansoni* (29). However, such steps occur immediately after oviposition, hence IPSE  
528 and other molecules secreted only after a few days of egg maturation cannot play a  
529 role in this initial process. Indeed, it was reported that this process was slower in  
530 mature eggs obtained from infected rodent livers in comparison with freshly  
531 deposited eggs, which may depend on the presence of uterine secretions covering  
532 the freshly deposited eggs (28). Once the eggs have reached the perivascular  
533 tissues, it takes another few days during which they need to cross several robust  
534 layers, including the submucosa, the outer muscularis mucosa, consisting of circular  
535 and longitudinal muscle, and the inner mucosa, complete with basal membrane and  
536 a very tight epithelial cell layer, before they can reach the lumen. A seemingly  
537 impossible journey?

538         It is well-accepted that the granulomatous reaction induced by the eggs plays  
539 a key role in this process (30, 31); much of the past research has focused on the  
540 interactions between immune cells and the schistosome eggs. In *S. mansoni*, eggs  
541 have been proposed to exploit gut lymphoid structures known as Peyer's patches as  
542 a preferential route of egress into the gut lumen (32), however a comparable route is  
543 not available to *S. haematobium* in the bladder. This leads to the question whether in  
544 addition to the inflammatory granulomatous response involving immune cells, any  
545 direct interactions with non-immune cells, such as fibroblasts, muscle cells or  
546 epithelial cells are also involved in facilitating egg translocation. In this context, it is  
547 interesting to note that in *S. haematobium*-infected animals, uroplakins and claudins  
548 involved in epithelial tight junction formation are downregulated after bladder  
549 exposure to eggs (33). Whether this downregulation, which is likely to aid egg egress  
550 by disrupting the integrity of the epithelium is due to the effects of H-IPSE or other  
551 egg-derived components, remains to be established. The nuclear translocation and



552 direct effects on gene transcription of H-IPSE are currently under active investigation  
553 in our laboratories.

554

555         Perhaps the most surprising result was to see to what extent an exogenously  
556 added parasitic molecule was able to enter host cells. This uptake does not appear  
557 to be very selective for specific cell types or animal species, as we have seen uptake  
558 of M-IPSE in human Huh-7, U2-OS and hamster CHO cell lines (18) as well as  
559 HUVEC and human monocyte derived dendritic cells (Falcone et al., unpublished  
560 data). H-IPSE was taken up by human HTB-9, Huh-7 and monkey Cos-7 cell lines.  
561 This raises the question as to whether or which receptors are involved. For M-IPSE,  
562 the uptake mechanism has been shown to involve the carbohydrate residues on the  
563 protein and C-type lectin receptors such as the mannose receptor, the dendritic cell-  
564 specific ICAM3-grabbing non-integrin (DC-SIGN), and a macrophage galactose-type  
565 lectin and the mannose receptor (34). Similar receptors have been shown to be  
566 involved in the uptake of other Schistosoma molecules such as omega-1 (5) and  
567 kappa-5 (34). Uptake of a secreted molecule by dendritic cells and macrophages has  
568 also been shown for *S. japonicum* Sj16 (35), but there is no information regarding  
569 the receptors involved in this process; the expression of Sj16 in *E. coli* used for the  
570 described experiments however suggests that protein glycosylation does not play a  
571 role in uptake of this molecule.

572

573         Thus secretion of a molecule by a parasitic life stage which is in intimate  
574 contact with host tissues, and its subsequent uptake by host cells, may be a more  
575 common feature in the host-parasite relationship than hitherto assumed, at least as  
576 far as trematodes are concerned. We would like to propose the term 'infiltrin' to

577 denote the ability of such molecules to enter host cells in the course of crossing  
578 several barriers (the cell membrane, presumably the endosomal membrane, and in  
579 some cases, the nuclear membrane). The archetypal nuclear infiltrins, characterized  
580 by the simultaneous presence of a classical secretory and a nuclear localization  
581 signal (CSS/NLS) signal (11), would be M-IPSE and H-IPSE, while the archetypal  
582 cytosolic infiltrin would be omega-1 (5). The ability of exogenous polypeptides to  
583 enter human cells crossing biological membranes is not a new finding. This was  
584 shown for the first time for trans-activating transcriptional activator (TAT) of human  
585 immunodeficiency virus 1 (HIV-1) in 1988 (36) and the 60 amino acid peptide  
586 encoded by antennapedia gene homeobox in *Drosophila* (37). However, with the  
587 exception of our previous report (18), such a principle has not been described for  
588 molecules secreted by macroparasites, which are too large to enter host cells. HIV-1  
589 Tat can also carry heterologous proteins across the cell membrane (38), a process  
590 now understood to be mediated via a caveolin-dependent uptake route (39).  
591 Interestingly, both HIV-1 Tat and *Drosophila* antennapedia homeobox peptide exhibit  
592 DNA binding activities, which are also predicted in silico for H-IPSE. Whether H-  
593 IPSE has similar properties is also under investigation. We believe that these  
594 observations make a compelling case warranting more in-depth studies of parasitic  
595 infiltrins and their potential roles as pathogen-derived nuclear transcription factors.

596

597 Finally, it needs to be noted that H-IPSE's ability to enter host cells is not  
598 dependent on an intact NLS, as the H03-IPSE NLS mutant, as well as the previously  
599 described M-IPSE NLS mutants (18), are also able to enter mammalian cells. The  
600 same is true for HIV-1 Tat, where the regions responsible for cellular uptake and  
601 nuclear translocation are distinct (40, 41). Our data suggest that in the absence of an

602 intact NLS, H-IPSE is able to enter host cells, but remains trapped in endosome-like  
603 vesicles with a perinuclear distribution. Whether NLS mutants retain their ability to  
604 bind host DNA remains to be established.

605

606 Taken together, we suggests that nuclear infiltrins, by acting e.g. as  
607 transcription factors, might play a central role in controlling the host-parasite  
608 relationship at the molecular level.

609

## 610 REFERENCES

611

- 612 1. Gryseels B, Polman K, Clerinx J, Kestens L. 2006. Human schistosomiasis.  
613 Lancet 368:1106–18.
- 614 2. Dewalick S, Bexkens ML, van Balkom BWM, Wu Y-P, Smit CH, Hokke CH, de  
615 Groot PG, Heck AJR, Tielens AGM, van Hellemond JJ. 2011. The proteome of  
616 the insoluble *Schistosoma mansoni* eggshell skeleton. Int J Parasitol 41:523–  
617 32.
- 618 3. Mathieson W, Wilson RA. 2010. A comparative proteomic study of the  
619 undeveloped and developed *Schistosoma mansoni* egg and its contents: the  
620 miracidium, hatch fluid and secretions. Int J Parasitol 40:617–28.
- 621 4. Cass CL, Johnson JR, Califf LL, Xu T, Hernandez HJ, Stadecker MJ, Yates  
622 JR, Williams DL. 2007. Proteomic analysis of *Schistosoma mansoni* egg  
623 secretions. Mol Biochem Parasitol 155:84–93.
- 624 5. Fitzsimmons CM, Schramm G, Jones FM, Chalmers IW, Hoffmann KF,  
625 Grevelding CG, Wuhrer M, Hokke CH, Haas H, Doenhoff MJ, Dunne DW.  
626 2005. Molecular characterization of omega-1: A hepatotoxic ribonuclease from

- 627 *Schistosoma mansoni* eggs. Mol Biochem Parasitol 144:123–127.
- 628 6. Schramm G, Hamilton JV, Balog CIA, Wuhrer M, Gronow A, Beckmann S,  
629 Wippersteg V, Grevelding CG, Goldmann T, Weber E, Brattig NW, Deelder  
630 AM, Dunne DW, Hokke CH, Haas H, Doenhoff MJ. 2009. Molecular  
631 characterisation of kappa-5, a major antigenic glycoprotein from *Schistosoma*  
632 *mansoni* eggs. Mol Biochem Parasitol 166:4–14.
- 633 7. Schramm G, Gronow a, Knobloch J, Wippersteg V, Grevelding CG, Galle J,  
634 Fuller H, Stanley RG, Chiodini PL, Haas H, Doenhoff MJ. 2006. IPSE/alpha-1:  
635 a major immunogenic component secreted from *Schistosoma mansoni* eggs.  
636 Mol Biochem Parasitol 147:9–19.
- 637 8. Delafosse L, Xu P, Durocher Y. 2016. Comparative study of polyethylenimines  
638 for transient gene expression in mammalian HEK293 and CHO cells. J  
639 Biotechnol 227:103–111.
- 640 9. Beetz C, Brodhun M, Moutzouris K, Kiehntopf M, Berndt A, Lehnert D, Deufel  
641 T, Bastmeyer M, Schickel J. 2004. Identification of nuclear localisation  
642 sequences in spastin (SPG4) using a novel Tetra-GFP reporter system.  
643 Biochem Biophys Res Commun 318:1079–1084.
- 644 10. Howe KL, Bolt BJ, Cain S, Chan J, Chen WJ, Davis P, Done J, Down T, Gao  
645 S, Grove C, Harris TW, Kishore R, Lee R, Lomax J, Li Y, Muller H-M,  
646 Nakamura C, Nuin P, Paulini M, Raciti D, Schindelman G, Stanley E, Tuli MA,  
647 Van Auken K, Wang D, Wang X, Williams G, Wright A, Yook K, Berriman M,  
648 Kersey P, Schedl T, Stein L, Sternberg PW. 2016. WormBase 2016:  
649 expanding to enable helminth genomic research. Nucleic Acids Res 44:D774-  
650 80.
- 651 11. Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. 2016. WormBase

- 652 ParaSite - a comprehensive resource for helminth genomics. Mol Biochem  
653 Parasitol.
- 654 12. Young ND, Jex AR, Li B, Liu S, Yang L, Xiong Z, Li Y, Cantacessi C, Hall RS,  
655 Xu X, Chen F, Wu X, Zerlotini A, Oliveira G, Hofmann A, Zhang G, Fang X,  
656 Kang Y, Campbell BE, Loukas A, Ranganathan S, Rollinson D, Rinaldi G,  
657 Brindley PJ, Yang H, Wang J, Wang J, Gasser RB. 2012. Whole-genome  
658 sequence of *Schistosoma haematobium*. Nat Genet 44:221–5.
- 659 13. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0:  
660 discriminating signal peptides from transmembrane regions. Nat Methods  
661 8:785–786.
- 662 14. Kosugi S, Hasebe M, Tomita M, Yanagawa H. 2009. Systematic identification  
663 of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by  
664 prediction of composite motifs. Proc Natl Acad Sci U S A 106:10171–10176.
- 665 15. Nakai K, Horton P. 1999. PSORT: A program for detecting sorting signals in  
666 proteins and predicting their subcellular localization. Trends Biochem Sci.
- 667 16. Nguyen Ba AN, Pogoutse A, Provar N, Moses AM. 2009. NLStradamus: a  
668 simple Hidden Markov Model for nuclear localization signal prediction. BMC  
669 Bioinformatics 10:202.
- 670 17. Brameier M, Krings A, MacCallum RM. 2007. NucPred - Predicting nuclear  
671 localization of proteins. Bioinformatics 23:1159–1160.
- 672 18. Kaur I, Schramm G, Everts B, Scholzen T, Kindle KB, Beetz C, Montiel-Duarte  
673 C, Blindow S, Jones AT, Haas H, Stolnik S, Heery DM, Falcone FH. 2011.  
674 Interleukin-4-inducing principle from *Schistosoma mansoni* eggs contains a  
675 functional C-terminal nuclear localization signal necessary for nuclear  
676 translocation in mammalian cells but not for its uptake. Infect Immun 79:1779–

- 677 1788.
- 678 19. Kosugi S, Hasebe M, Matsumura N, Takashima H, Miyamoto-Sato E, Tomita  
679 M, Yanagawa H. 2009. Six classes of nuclear localization signals specific to  
680 different binding grooves of importin alpha. *J Biol Chem* 284:478–85.
- 681 20. Schramm G, Falcone FH, Gronow A, Haisch K, Mamat U, Doenhoff MJ,  
682 Oliveira G, Galle J, Dahinden C a, Haas H. 2003. Molecular characterization of  
683 an interleukin-4-inducing factor from *Schistosoma mansoni* eggs. *J Biol Chem*  
684 278:18384–92.
- 685 21. Wuhrer M, Balog CIA, Catalina MI, Jones FM, Schramm G, Haas H, Doenhoff  
686 MJ, Dunne DW, Deelder AM, Hokke CH. 2006. IPSE/alpha-1, a major  
687 secretory glycoprotein antigen from schistosome eggs, expresses the Lewis X  
688 motif on core-difucosylated N-glycans. *FEBS J* 273:2276–2292.
- 689 22. Shatkay H, Höglund A, Brady S, Blum T, Dönnnes P, Kohlbacher O. 2007.  
690 SherLoc: High-accuracy prediction of protein subcellular localization by  
691 integrating text and protein sequence data. *Bioinformatics* 23:1410–1417.
- 692 23. Wang R, Brattain MG. 2007. The maximal size of protein to diffuse through the  
693 nuclear pore is larger than 60 kDa. *FEBS Lett* 581:3164–3170.
- 694 24. Neill PJG, Smith JH, Doughty BL, Kemp M. 1988. The ultrastructure of the  
695 *Schistosoma mansoni* egg. *Am J Trop Med Hyg* 39:52–65.
- 696 25. Ashton PDD, Harrop R, Shah B, Wilson RA. 2001. The schistosome egg:  
697 development and secretions. *Parasitology* 122:329–338.
- 698 26. Jurberg AD, Gonçalves T, Costa TA, Mattos ACA, Pascarelli BM, Manso PPA,  
699 Ribeiro-Alves M, Pelajo-Machado M, Peralta JM, Coelho PMZ, Lenzi HL.  
700 2009. The embryonic development of *Schistosoma mansoni* eggs: Proposal  
701 for a new staging system. *Dev Genes Evol* 219:219–234.

- 702 27. Michaels RM, Prata A. 1968. Evolution and characteristics of *Schistosoma*  
703 *mansoni* eggs laid in vitro. J Parasitol 54:921–930.
- 704 28. File S. 1995. Interaction of schistosome eggs with vascular endothelium. J  
705 Parasitol 81:234–8.
- 706 29. deWalick S, Hensbergen PJ, Bexkens ML, Grosserichter-Wagener C, Hokke  
707 CH, Deelder AM, de Groot PG, Tielens AGM, Van Hellemond JJ. 2014.  
708 Binding of von Willebrand factor and plasma proteins to the eggshell of  
709 *Schistosoma mansoni*. Int J Parasitol 44:263–268.
- 710 30. Doenhoff MJ, Hassounah O, Murare H, Bain J, Murare H. 1986. The  
711 schistosome egg granuloma: Immunopathology in the cause of host protection  
712 or parasite survival? Trans R Soc Trop Med Hyg 80:503–514.
- 713 31. Pearce EJ. 2005. Priming of the immune response by schistosome eggs.  
714 Parasite Immunol.
- 715 32. Turner JD, Narang P, Coles MC, Mountford AP. 2012. Blood Flukes Exploit  
716 Peyer's Patch Lymphoid Tissue to Facilitate Transmission from the  
717 Mammalian Host. PLoS Pathog 8:e1003063.
- 718 33. Ray D, Nelson TA, Fu CL, Patel S, Gong DN, Odegaard JI, Hsieh MH. 2012.  
719 Transcriptional Profiling of the Bladder in Urogenital Schistosomiasis Reveals  
720 Pathways of Inflammatory Fibrosis and Urothelial Compromise. PLoS Negl  
721 Trop Dis 6:e1912.
- 722 34. Meevissen MHJ, Driessen NN, Smits HH, Versteegh R, van Vliet SJ, van  
723 Kooyk Y, Schramm G, Deelder AM, Haas H, Yazdanbakhsh M, Hokke CH.  
724 2012. Specific glycan elements determine differential binding of individual egg  
725 glycoproteins of the human parasite *Schistosoma mansoni* by host C-type  
726 lectin receptors. Int J Parasitol 42:269–277.

- 727 35. Sun X, Yang F, Shen J, Liu Z, Liang J, Zheng H, Fung M, Wu Z. 2016.  
728 Recombinant Sj16 from *Schistosoma japonicum* contains a functional N-  
729 terminal nuclear localization signal necessary for nuclear translocation in  
730 dendritic cells and interleukin-10 production. *Parasitol Res* 115:4559–4571.
- 731 36. Frankel AD, Pabo CO. 1988. Cellular uptake of the tat protein from human  
732 immunodeficiency virus. *Cell* 55:1189–1193.
- 733 37. Joliot A, Pernelle C, Deagostini-Bazin H, Prochiantz A. 1991. Antennapedia  
734 homeobox peptide regulates neural morphogenesis. *Proc Natl Acad Sci U S A*  
735 88:1864–8.
- 736 38. Fawell S, Seery J, Daikh Y, Moore C, Chen LL, Pepinsky B, Barsoum J. 1994.  
737 Tat-mediated delivery of heterologous proteins into cells. *Proc Natl Acad Sci U*  
738 *S A* 91:664–668.
- 739 39. Ferrari A, Pellegrini V, Arcangeli C, Fittipaldi A, Giacca M, Beltram F. 2003.  
740 Caveolae-mediated internalization of extracellular HIV-1 Tat fusion proteins  
741 visualized in real time. *Mol Ther* 8:284–294.
- 742 40. Efthymiadis A, Briggs LJ, Jans DA. 1998. The HIV-1 tat nuclear localization  
743 sequence confers novel nuclear import properties. *J Biol Chem* 273:1623–  
744 1628.
- 745 41. Kuppuswamy M, Subramanian T, Srinivasan A, Chinnadurai G. 1989. Multiple  
746 functional domains of Tat, the trans-activator of HIV-1, defined by mutational  
747 analysis. *Nucleic Acids Res* 17:3551–61.
- 748
- 749



750 **ACKNOWLEDGMENTS**

751 We would like to thank Dr Christian Beetz for the kind donation of the  
752 TetraEGFP plasmid and Dr Hilary Collins for help with confocal microscopy. This  
753 work was funded by an NIH R56 grant (R56AI119168) awarded to MHH and TSJ.  
754 The funders had no role in study design, data collection and interpretation, or the  
755 decision to submit the work for publication.

756

757 **Figure Legends**

758

759

760

761

762 **Figure 1:** *S. haematobium* expresses multiple forms of H-IPSE. A) Amino acid  
763 sequences of predicted H-IPSE paralogs, M-IPSE, and sequenced transcripts from  
764 egg cDNA were globally aligned using a Blosum62 cost matrix and a tree was built  
765 using the neighbor-joining method in Geneious 7.1.4. Scale bar represents amino  
766 acid substitutions per site. H06 and H03 variants chosen for expression are  
767 highlighted in green, and their respective amino acid identity and identity to M-IPSE  
768 are shown. All variants identified through 3' RACE cloning are denoted with  
769 asterisks. B) Alignment of amino acid sequences of H03-IPSE (top row) and H06-  
770 IPSE (middle row) with the homolog in *S. mansoni* (IPSE/alpha-1, named M-IPSE  
771 here, bottom row). These H-IPSE clones retain ~63-68% amino acid identity and  
772 several previously features described in M-IPSE (20), including: a 20 amino acid  
773 classical signal sequence, seven cysteine residues involved in disulfide bonds, two  
774 N-linked glycosylation consensus motifs, and a predicted nuclear localization  
775 sequence (data not shown). Residues colored in green are identical, residues in  
776 yellow share properties (e.g. hydrophobicity, polarity), and residues in red lack  
777 similarity. C) SDS-PAGE gel (left) and western blot with anti-H06-IPSE antiserum  
778 (right). Lanes contain parasite-derived adult worm antigen (AWA), egg secretory  
779 protein (ESP), or soluble egg antigen (SEA).

780

781

782 **Figure 2:** Effect of multiple amino acid substitutions on NLS in H-IPSE. A) The  
783 nucleotides encoding the H03/H06-IPSE nuclear localization sequence (SKRRRKY  
784 and SKRGRKY, respectively) were inserted into the pTetra-EGFP construct (2,3).  
785 pTetra-EGFP encodes a tetrameric EGFP construct resulting in the expression of a  
786 fluorescent protein which due to its size (>100 kDa) is excluded from the nucleus in  
787 the absence of a functional NLS (Tetra-EGFP) or imported into the nucleus in the  
788 presence of a functional NLS (canonical SV40 NLS, H03/H06-IPSE NLS). Nuclei  
789 were stained with DAPI and green fluorescence measured with the GFP light cube  
790 on an EVOS *fl* microscope, 24 hours after transfection. Bar is 100  $\mu$ m. B)  
791 Comparison of wild-type H06-IPSE, H03-IPSE and H03-IPSE mutant NLS effect on  
792 nuclear localisation of Tetra-EGFP fusion protein. One hundred transfected HTB9  
793 cells were evaluated under the EVOS *fl* microscope for each transfection and the  
794 percentage of cells displaying exclusive nuclear fluorescence, as opposed to  
795 cytosolic only or mixed cytosolic/nuclear localization, recorded. Positive control:  
796 Sv40 canonical NLS sequence; negative control: unmodified Tetra-EGFP vector  
797 (Tetra-EGFP).  
798

799 **Figure 3:** Expression of M-IPSE and H-IPSE in HEK293-6A cells. A) Schematic  
800 diagram of pTT5 H03/06-IPSE expression cassette. eCMV=Cytomegalovirus (CMV)  
801 enhancer sequence; pCMV=CMV promoter; TPL=tripartite leader sequence from  
802 adenovirus; eMLP=enhancer element from the adenovirus major late promoter  
803 (MLP); hVEGF=human vascular endothelial growth factor signal sequence;  
804 8xHis=octahistidine tag; TEV=Tobacco Etch Virus protease cleavage site; STOP:  
805 stop codon; pA:  $\beta$ -globin polyadenylation signal. B) Coomassie-stained 4-20% SDS-  
806 PAGE gradient gel and C) Western Blotting of recombinant H03-IPSE (and M-IPSE,  
807 used as comparison) expressed in HEK 293SF-3F6 cells and purified by IMAC from  
808 serum-free culture supernatant, and run under non-reducing (NR) or reducing (R)  
809 conditions.

810 **Figure 4:** Recombinant H03-IPSE is taken up by HTB-9 host cells and translocates  
811 to the nucleus. HTB-9 cells, incubated for 24 hours with 0.40 nM recombinant H03-  
812 IPSE, were stained with 5  $\mu$ M DRAQ5 nuclear stain for 15 minutes at room  
813 temperature, followed by staining with a mouse anti-His antibody and Alexa Fluor®  
814 555 conjugated Goat anti-Mouse IgG (H+L) as secondary antibody. The right column  
815 shows the overlay of the two channels. The uptake in HTB-9 cells was visualized by  
816 confocal microscopy. The primary anti-His antibody was omitted in the control lane.  
817

818 **Figure 5:** Fluorescence microscopy of HTB-9 cells incubated with recombinant H03-  
819 IPSE (NLS: SKRRRKY), H06-IPSE (NLS: SKRGRKY) or H03-IPSE mutant (NLS:  
820 SKAAAKY). HTB-9 cells were stained with Hoechst 33342 nuclear stain for 15  
821 minutes at room temperature, followed by staining with a mouse anti-His antibody  
822 and Alexa Fluor® 555-conjugated Goat anti-Mouse IgG (H+L) as secondary  
823 antibody. The right column shows the overlay of the two channels. The primary anti-  
824 His antibody was omitted in the control lane. Bar size is 100 µm.

825

826

827

828 **Figure 6:** Stage-specific expression of H-IPSE mRNA. RT-PCR results for H-IPSE  
829 obtained from cDNAs, prepared by reverse transcription of DNase-treated RNA of  
830 various life stages of *S haematobium*. Ladder: 100 basepair (Bp) DNA ladder; egg:  
831 *S. haematobium* egg cDNA; mir: miracidial cDNA; cer: cercarial cDNA; som: *in vitro*  
832 mechanically transformed schistosomula cDNA; Ad, F, M mixed cDNA from female,  
833 male or mixed adult worms, respectively. ShTub: control housekeeping gene, *S.*  
834 *haematobium* tubulin.















