Title: A single-cell RNAseq atlas of *Schistosoma mansoni* identifies a key regulator of blood feeding

Authors: George Wendt[1]*, Lu Zhao[1]*, Rui Chen[1], Chenxi Liu[2], Anthony J. O'Donoghue[2], Conor R. Caffrey[2], Michael L. Reese[1,3], James J. Collins III[1†].

Affiliations: [1]Department of Pharmacology, UT Southwestern Medical Center, Dallas, Texas 75390

[2]Center for Discovery and Innovation in Parasitic Diseases, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, 9500 Gilman Dr., La Jolla, California 92093

[3]Department of Biochemistry, UT Southwestern Medical Center, Dallas, Texas 75390

*Equal Contribution

[†]To whom correspondence should be addressed

JamesJ.Collins@UTSouthwestern.edu

UT Southwestern Medical Center

Department of Pharmacology

6001 Forest Park. Rd.

Dallas, TX 75390

United States of America

Abstract: Schistosomiasis is a neglected tropical disease that infects 240 million people. With no vaccines and only one drug available, new therapeutic targets are needed. The causative agents, schistosomes, are intravascular flatworm parasites that feed on blood and lay eggs, causing pathology. The function of the parasite's various tissues in successful parasitism are poorly understood, hindering identification of novel therapeutic targets. Using single cell RNAseq we characterize 43,642 cells from the adult schistosome, identifying 68 distinct cell populations including specialized stem cells that maintain the parasite's blood-digesting gut. These stem cells express the gene *hnf4*, which is required for gut maintenance, blood feeding, and pathology *in vivo*. Together, these data provide molecular insights into the organ systems of this important pathogen and identify potential therapeutic targets.

One Sentence Summary: Single-cell RNAseq is used to study the basic biology of and identify a novel therapeutic target against a deadly parasite

Main Text: Schistosomes dwell inside the host's circulation, often for decades, where they feed on blood and lay eggs, which become trapped in host tissues and cause disease pathology. As a metazoan comprised of multiple tissue types, understanding the schistosome's biology on a molecular level during parasitism could suggest novel therapeutic strategies. Single-cell RNAseq (scRNAseq) has been used to comprehensively describe tissue types and physiology of diverse metazoans (*1*) including larval schistosomes (*2*) but we lack a comprehensive description of the cell types present in egg-laying adults as specific molecular markers are known for only a small number of cell types (*3*-*8*).

To define the molecular signature of adult schistosome cell types, we dissociated adult *Schistosoma mansoni*, isolated cells by Fluorescence-Activated Cell Sorting (FACS), and generated scRNAseq libraries using a 10x genomics chromium controller (**Fig. S1A**). Schistosomes are dioecious and sexual maturation of the female worm's reproductive organs, including the ovary and vitellaria, requires sustained physical contact with the male worm (*9*).

Accordingly, we generated scRNAseq libraries from adult male parasites, adult sexually mature female parasites, and age-matched virgin female parasites. We then performed clustering, identifying 68 molecularly distinct clusters composed of 43,642 cells (**Figs. 1A, S1B, Table S1**). These included: three clusters of cells expressing somatic stem cell (*i.e.,* neoblast) markers such as the RNA binding protein *nanos2,* the cell surface receptor *notch,* and the receptor tyrosine kinase *fgfra* (*3*) (**Figs. 1B, S2A**); eight clusters expressing markers of tegument ("skin"-like surface) progenitors (*4, 5*) (**Fig. S2B**); two clusters of parenchymal cells (**Figs. 1C, S2C**); one cluster of ciliated flame cells that are part of the worm's protonephridial (excretory) system (**Figs. 1D, S2D**); eight clusters of muscles (**Fig. 1E**); and a cluster of esophageal gland cells (**Figs. 1F, S2E**). Despite being composed of thousands of nuclei, our analysis also identified clusters corresponding to syncytial tissues: the tegument (*4*) (**Figs. 1G, S2F**) and gut (**Figs. 1H, S2G**). We failed to identify cells from the female ootype (an organ involved in eggshell formation) (*9*) and the protonephridial ducts (*10*), possibly because of their multinucleate nature. Gene ontology (GO) analyses of these clusters (**Table S2**) confirmed expected findings (enrichment of "DNA replication" in "neoblast 1") and revealed novel biology such as the enrichment of "extracellular matrix structural components" in muscle clusters suggesting muscles are the source of extracellular matrix in schistosomes, similar to planarians (*11*).

We uncovered unexpected molecular complexity within the schistosome nervous system, identifying 30 clusters expressing the neuroendocrine protein *7b2* (**Figs. 1I**) and one apparent neuronal cluster that did not express high levels of *7b2* but expressed several of synaptic molecules (e.g. *synapsin*) (**Figs. S3A, Table S1**). Examination of genes from these neuronal cell clusters uncovered unique molecular fingerprints for several populations (**Figs. S3A, S4, Table S1**) and highly-ordered structural and regional specialization in the central and peripheral nervous systems,

including left-right asymmetry (**Fig. S3B**) and nine types of apparently ciliated neurons (**Fig. S3C,D**). This complexity is surprising given the relatively "sedentary" lifestyle of adult parasites in the portal vasculature (*9*).

Schistosome muscle is also very heterogeneous, with eight muscle clusters that possess unique expression patterns (**Fig. S5A-C**). Some populations occur diffusely throughout the animal ("muscle 1" and "muscle 2"), whereas others are anatomically restricted such as "muscle 7" cells that reside next to the gut, suggesting they are enteric muscles.

Similar to planarians (*12*), many morphogens that regulate *wnt* (**Fig. S6A-D**) and *tgfb* signaling (**Fig. S6E-H**) are expressed in muscle and neuronal cells. Homologs of many of these genes are expressed specifically in planarian muscles (*1*) and have been implicated in regeneration in planarians (*12*). Though schistosomes survive amputation (*13*), there is no evidence of whole-body regeneration. This expression pattern in a non-regenerative animal suggests these genes may regulate schistosome neoblasts during homeostasis.

The pathology of schistosome infection is driven by the host's inflammatory responses to parasite eggs(*14*). Thus, we examined the differences between male, sexually mature female, and age-matched virgin females at the cellular level (**Fig. 2A**). All adult parasites have germline stem cells (GSCs) marked by expression of *nanos1*(*6*). Our scRNAseq data revealed that GSCs have very similar gene expression regardless of sex or maturity (**Figs. 2B, S7A**). Like GSCs, GSC progeny fall into the same clusters in both male and female parasites, suggesting no major sex- or maturation-dependent differences in early gametogenesis (**Figs. 2C, S7B**). However, later germ cells cluster according to sex, with expression of "late female germ cells" markers found predominantly in mature females (**Figs. 2D and S7C**) and "late male germ cells" markers only in males (**Fig. S7D**).

The sexually mature schistosome ovary is structured such that GSCs reside at the anterior and mature oocytes at the posterior end (*6, 15*). The "GSCs" marker *nanos1* is expressed in the proliferative anterior compartment (**Figs. 2B, top, S8A-D**) whereas the "late female germ cells" marker *bmpg* is expressed most highly in the posterior ovary (**Figs. 2D, top, S8C**). Our single-cell RNAseq data shows that the "GSC progeny" cluster sits between "GSCs" and "late female germ cells" on the UMAP plot, (**Fig. 2A**), with the "GSC progeny" marker *meiob* expressed most highly between the anterior and posterior ovary (**Figs. 2C, S7B**). Concurrent visualization of these clusters reveals an organized linear architecture (**Fig. S8E**). Interestingly, both mature and virgin females express the marker *meiob* (**Fig. 2C**), suggesting that virgin female GSCs express differentiation markers without male stimulus. Thus, male parasites may regulate this developmental checkpoint by promoting survival of differentiating GSCs rather than inducing commitment, consistent with studies suggesting that male-female pairing can suppress apoptosis in the vitellaria of virgin female worms (*16*).

We also examined the vitellaria, another male-sensitive, stem-cell dependent tissue that produces the yolk cells of the parasite's eggs. Despite a different function and organization, we observed parallels between ovary and vitellaria maturation, such as an apparent lineage from stem cell to mature tissue (**Fig. S9A-D**). We also found a low frequency of vitellocyte-like cells in males (*17*) (**Fig. S9A**). Finally, we identified pairing-independent sexual tissues such as the flatworm-specific Mehlis' gland that plays an enigmatic role in egg production (*9*) (**Fig. S9E**).

In addition to sexual tissues, we observed sexual dimorphism in non-reproductive tissues as well including 3 muscle clusters (muscle 5, 6, and 8) that appear to be largely restricted to female parasites (**Table S3**), with "muscle 8" representing muscle cells that surround the ovary (**Fig. S10**). In some cases, we observed unexpected numbers of male cells in clusters of female

sexual tissues which we attribute to neoblasts expressing low levels of differentiated tissue markers, like what has been observed in planarians (*1*) (**Fig. S11, Supplementary Text**).

Egg production is the primary driver of pathology, but this pathology is exacerbated by the parasite's stem cell-mediated longevity (*3*). Previous work suggests adult neoblasts are molecularly homogeneous and predominantly give rise to cells involved in tegument production (*4, 5*) but free-living flatworms are known to possess functionally distinct neoblasts that produce specific tissues (*18*). We identified a subpopulation of neoblasts ("*eled*$^{+}$ neoblasts") that formed a putative non-tegument lineage as suggested by a linear "path" of cells from *eled*$^{+}$ neoblasts to the gut (**Figs. 3A, S12A-F**). These *eled*$^{+}$ neoblasts expressed *hnf4* (**Figs. 3A and S12B,C**), a marker of gut neoblasts in planarians (*18*). Given the importance of gut-mediated blood digestion for egg production (*19*), we sought to perturb this lineage by RNAi of genes expressed in this lineage (**Fig. S13A,B**). We found that knocking down *hnf4* resulted in a ~3.8-fold increase in *eled*$^{+}$ neoblasts (**Figs. 3B and S13C-F**) and a concomitant decrease in the expression of several gut markers (**Fig. S14A,B**). Indeed, RNAseq on *hnf4(RNAi)* animals demonstrated that over 70% of transcripts expressed in the "gut" cluster were downregulated (**Fig. S14C,D, Table S4**).

To understand whether stem cells functioned normally in *hnf4(RNAi)* animals, we first looked at apoptosis using TUNEL and found no difference in *hnf4(RNAi)* animals, ruling out increased cell death (**Fig. S15A**). Next we looked at tegument production using EdU pulse-chase approaches. We found a significant increase in tegument production compared to controls (**Fig. S15B,C**), ruling out a broad stem cell differentiation defect. Our ability to monitor new gut production by EdU pulse-chase approaches was complicated by the fact that gut marker expression was largely absent in most *hnf4(RNAi)* parasites (**Fig. S14A,B**). In cases where we could detect gut marker expression in EdU pulse-chase experiments, we found gut-like tissue was being

produced in *hnf4(RNAi)* parasites but was morphologically abnormal (**Fig. S15D**). Examination of the expression of *eled* and the gut marker *ctsb* revealed that locations where *eled*$^{+}$ neoblasts were abundant lacked normal gut tissue (**Fig. S15E**). This suggests that the impairment of gut production is at least partially responsible for the gut defects following *hnf4* RNAi.

To assess gut structure, we next supplemented the culture media of *hnf4(RNAi)* parasites with fluorescently-labeled dextran (which labels the gut lumen (*20*)). After 12 hours of culture, all control(RNAi) parasites but only 1 out of 15 *hnf4(RNAi)* parasites had dextran in the lumen (**Fig. 3C**). The dextran failed to enter the digestive tract of the *hnf4(RNAi)* parasites (**Fig. S16A**), suggesting either a complete loss of patency or a defect in the parasite's ability to coordinate dextran ingestion. We then examined *hnf4(RNAi)* animals by transmission electron microscopy (TEM). The schistosome gut is a syncytial blind tube-like structure with a microvilli-filled lumen (*21*). Though gut tissue was still present, we found a significant decrease in luminal microvilli (**Figs. 3D, S16B**) and 2 out of 4 of *hnf4(RNAi)* animals had dilated lumens compared to controls (**Fig. S16C**).

To assess the digestive capability of *hnf4(RNAi)* parasites, we added red blood cells (RBCs) to the media and observed the parasites' ability to uptake and digest the cells. *hnf4(RNAi)* parasites either failed to ingest (15/69) or digest RBCs (54/69) (**Figs. 4A, S17A**). Because we observed a decrease in the expression of proteolytic enzymes by RNAseq (**Table S4**), we studied whether *hnf4* RNAi resulted in loss of cysteine (cathepsin) protease activity (which contributes to hemoglobin digestion (*22, 23*)). Measuring cathepsin activity of lysates in *hnf4(RNAi)* parasites using a fluorogenic peptidyl substrate, we found cathepsin B activity was decreased 8.2-fold relative to control parasites (**Fig. 4B**) consistent with gene expression analyses (**Table S4**). In contrast, aspartyl protease activity was similar in control and *hnf4(RNAi)* parasites (**Fig. S17B**),

which could reflect expression of aspartic proteases in non-gut tissues that were unaffected following *hnf4* RNAi (**Table S1, S4**). Together, these data suggest *hnf4* is at least indirectly required for the digestion of hemoglobin, in part by regulating the expression of cathepsin B, a key contributor to the digestion of blood proteins including hemoglobin (*22, 23*) in *S. mansoni*.

We examined whether *hnf4* was required to cause disease in the host by transplanting control and *hnf4(RNAi)* parasites into uninfected mice and then perfusing the mice 22-30 days post-transplant. Worm recovery was statistically indistinguishable (control(RNAi) = 72% vs. *hnf4(RNAi)* = 49%, $p = 0.136$, Welch's t-test) (**Fig. S17C**). This observation is not entirely unexpected as schistosomes can acquire nutrients though their tegument (*19*). Nonetheless, mice receiving *hnf4(RNAi)* parasites had morphologically normal livers in contrast to abundant egg-induced granulomata in control parasite recipients (**Figs. 4C, S17D**). Additionally, recovered male *hnf4(RNAi)* parasites were significantly shorter than controls (2.87mm vs. 5.21mm, respectively, $p < 0.0001$, Welch's t-test) (**Figs. 4D, S17E**). These results show *hnf4* is at least indirectly required for parasite growth and egg-induced pathology *in vivo*. Together, these data suggest *hnf4* specifically and gut homeostasis generally are potential therapeutic targets to blunt the pathology caused by adult parasites.

Here we describe a comprehensive single-cell atlas of the adult schistosome, identify regulators of gut biology, and leverage this knowledge to experimentally block schistosome-induced pathology in the host. We envision these data serving as a catalyst towards understanding other aspects of schistosome biology (e.g., reproductive biology) and serving as a foundation for understanding the development of various cellular lineages during the parasite lifecycle. Indeed, our approach serves as a template for the investigation of other understudied and experimentally

challenging parasitic metazoans, improving our understanding of their biology and enabling the discovery of novel therapies for these pathogens.

References and Notes:

1. C. T. Fincher, O. Wurtzel, T. de Hoog, K. M. Kravarik, P. W. Reddien, Cell type transcriptome atlas for the planarian *Schmidtea mediterranea*. *Science* **360**, (2018).
2. B. Wang *et al.*, Stem cell heterogeneity drives the parasitic life cycle of *Schistosoma mansoni*. *Elife* **7**, (2018).
3. J. J. Collins, III *et al.*, Adult somatic stem cells in the human parasite *Schistosoma mansoni*. *Nature* **494**, 476-479 (2013).
4. G. R. Wendt *et al.*, Flatworm-specific transcriptional regulators promote the specification of tegumental progenitors in *Schistosoma mansoni*. *Elife* **7**, (2018).
5. J. J. Collins, G. R. Wendt, H. Iyer, P. A. Newmark, Stem cell progeny contribute to the schistosome host-parasite interface. *Elife* **5**, (2016).
6. J. Wang, J. J. Collins, III, Identification of new markers for the *Schistosoma mansoni* vitelline lineage. *International journal for parasitology* **46**, 405-410 (2016).
7. G. P. Dillon, J. C. Illes, H. V. Isaacs, R. A. Wilson, Patterns of gene expression in schistosomes: localization by whole mount in situ hybridization. *Parasitology* **134**, 1589-1597 (2007).
8. Z. Lu *et al.*, Schistosome sex matters: a deep view into gonad-specific and pairing-dependent transcriptomes reveals a complex gender interplay. *Sci Rep* **6**, 31150 (2016).
9. P. F. Basch, *Schistosomes: Development, Reproduction, and Host Relations*. (Oxford University Press, New York, 1991), pp. 248.
10. R. A. Wilson, L. A. Webster, Protonephridia. *Biol Rev Camb Philos Soc* **49**, 127-160 (1974).
11. L. E. Cote, E. Simental, P. W. Reddien, Muscle functions as a connective tissue and source of extracellular matrix in planarians. *Nature communications* **10**, 1592 (2019).
12. J. N. Witchley, M. Mayer, D. E. Wagner, J. H. Owen, P. W. Reddien, Muscle cells provide instructions for planarian regeneration. *Cell reports* **4**, 633-641 (2013).
13. I. Popiel, D. L. Irving, P. F. Basch, Wound healing in the trematode *Schistosoma*. *Tissue Cell* **17**, 69-77 (1985).
14. E. J. Pearce, A. S. MacDonald, The immunobiology of schistosomiasis. *Nat Rev Immunol* **2**, 499-511 (2002).
15. P. M. Nollen, R. D. Floyd, R. G. Kolzow, D. L. Deter, The timing of reproductive cell development and movement in *Schistosoma mansoni*, *S. japonicum*, and *S. haematobium,* using techniques of autoradiography and transplantation. *J Parasitol* **62**, 227-231 (1976).
16. S. E. Galanti, S. C. Huang, E. J. Pearce, Cell death and reproductive regression in female *Schistosoma mansoni*. *PLoS Negl Trop Dis* **6**, e1509 (2012).
17. M. K. Shaw, D. A. Erasmus, *Schistosoma mansoni:* the presence and ultrastructure of vitelline cells in adult males. *J Helminthol* **56**, 51-53 (1982).
18. J. C. van Wolfswinkel, D. E. Wagner, P. W. Reddien, Single-Cell Analysis Reveals Functionally Distinct Classes within the Planarian Stem Cell Compartment. *Cell Stem Cell* **15**, 326-339 (2014).
19. P. J. Skelly, A. A. Da'dara, X. H. Li, W. Castro-Borges, R. A. Wilson, Schistosome feeding and regurgitation. *PLoS pathogens* **10**, e1004246 (2014).

20. S. L. Hall *et al.*, Insights into blood feeding by schistosomes from a proteomic analysis of worm vomitus. *Molecular and biochemical parasitology* **179**, 18-29 (2011).
21. G. P. Morris, Fine structure of the gut epithelium of *Schistosoma mansoni*. *Experientia* **24**, 480-482 (1968).
22. C. R. Caffrey, L. Goupil, K. M. Rebello, J. P. Dalton, D. Smith, Cysteine proteases as digestive enzymes in parasitic helminths. *PLoS Negl Trop Dis* **12**, e0005840 (2018).
23. M. Sajid *et al.*, Functional expression and characterization of *Schistosoma mansoni* cathepsin B and its trans-activation by an endogenous asparaginyl endopeptidase. *Mol Biochem Parasitol* **131**, 65-75 (2003).
24. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature biotechnology* **36**, 411-420 (2018).
25. T. Stuart *et al.*, Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e1821 (2019).
26. J. J. Collins, III *et al.*, Genome-Wide Analyses Reveal a Role for Peptide Hormones in Planarian Germline Development. *PLoS Biol* **8**, e1000509 (2010).
27. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
28. K. L. Howe *et al.*, WormBase 2016: expanding to enable helminth genomic research. *Nucleic Acids Res* **44**, D774-780 (2016).
29. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
30. C. R. Caffrey, A. Ruppel, Cathepsin B-like activity predominates over cathepsin L-like activity in adult *Schistosoma mansoni* and *S. japonicum*. *Parasitol Res* **83**, 632-635 (1997).
31. J. N. Collins, J. J. Collins, III, Tissue Degeneration following Loss of *Schistosoma mansoni* cbp1 Is Associated with Increased Stem Cell Proliferation and Parasite Death In Vivo. *PLoS pathogens* **12**, e1005963 (2016).
32. H. Vogel, Hermaphrodites of *Schistosoma mansoni*. *Annals of tropical medicine and parasitology* **41**, 266-277 (1947).

Acknowledgements

We thank C. Paz for technical assistance and G. Hon for expertise in scRNAseq library preparation. Schistosome Infected mice and *B. glabrata* snails were provided by the National Institute of Allergy and Infectious Diseases (NIAID) Schistosomiasis Resource Center of the Biomedical Research Institute (Rockville, MD, USA) through National Institutes of Health (NIH)-NIAID Contract HHSN272201700014Ifor distribution through BEI Resources. FACS was performed with the aid of the Moody Foundation Flow Cytometry Facility at the University of Texas Southwestern Medical Center (UTSW). TEM imaging and sample preparation was performed with the aid of the Electron Microscopy Core at UTSW. RNAseq was performed with the aid of the

McDermott Center Next Generations Sequencing Core at UTSW. **Funding:** This work was supported by the National Institutes of Health R01 R01AI121037 (J.J.C.), R01 R01AI150715 (M.L.R.), R21 R21AI133393 (A.J.O.D.), and F30 1F30AI131509-01A1 (G.R.W.), the Welch Foundation I-1948-20180324 (J.J.C.), I-1936-20170325 (M.L.R.), the National Science Foundation MCB1553334 (M.L.R.), the Burroughs Wellcome Fund (J.J.C.), the Wellcome Trust 107475/Z/15/Z (J.J.C.), and the Bill and Melinda Gates Foundation OPP1171488 (C.R.C.). **Author contributions:** Conceptualization, G.W., L.Z, R.C., C.L., A.J.O.D., C.R.C, J.J.C.; investigation, G.W., L.Z, R.C., C.L.,J.J.C.; Designing web-based resources: M.L.R.; Writing-original draft: G.W., L.Z., J.J.C; Writing-review and editing, all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** A searchable database of scRNAseq data can be accessed at www.collinslab.org/schistocyte; raw scRNAseq plots can be accessed at https://doi.org/10.5061/dryad.0k6djh9xk. Raw and some processed data from single cell RNAseq and *hnf4* RNAi RNAseq experiments have been deposited in the NCBI Gene Expression are available from NCBI GEO with accession number GSE146737.

Supplementary Materials:

Materials and Methods

Supplementary Text

Figures S1 to S18

Tables S1 to S7

References (24-32)

Fig. 1. *Schistosoma mansoni* single cell atlas.

(**A**) Uniform Manifold Approximation and Projection (UMAP) plot of the 68 scRNAseq clusters. (**B-I**), (left) UMAP plot and whole-mount *in situ* hybridization (WISH) of the indicated gene and its expression in the noted tissue in the head (middle, top) and body (middle, bottom) of a male and the ovary (right, top) and vitellaria (right, bottom) of a mature female parasite. Scale bars, 100μm. UMAP plots colored by gene expression (blue = low, red = high).

Fig. 2. The germ lineage in schistosome ovaries.

(**A**) UMAP plots of all clusters split by parasite sex. Sexual tissues are labeled. (**B-D**) (top) WISH and UMAP plot of gene expression of indicated gene in sexually mature females (m♀) (top) and in virgin females (v♀) (bottom) for the "GSCs" marker *nanos1* (**B**), the "GSC progeny" marker *meiob* (**C**), and the "late female germ cells" marker *bmpg* (**D**). Dashed line indicates boundary of ovary. Scale bars, 100μm. UMAP plots are colored by gene expression (blue = low, red = high).

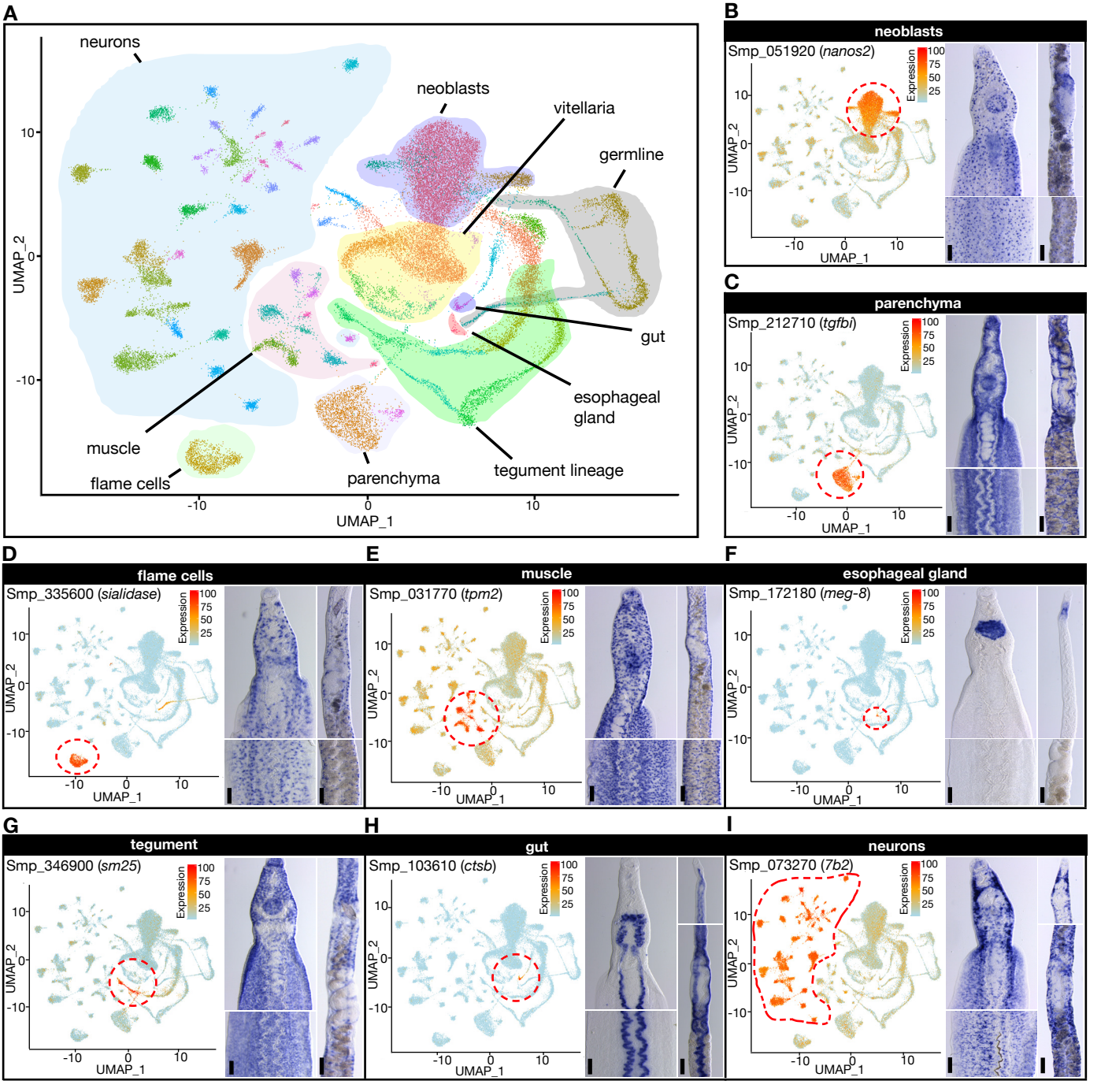
Fig. 3. An *hnf4* homolog regulates a novel gut lineage.

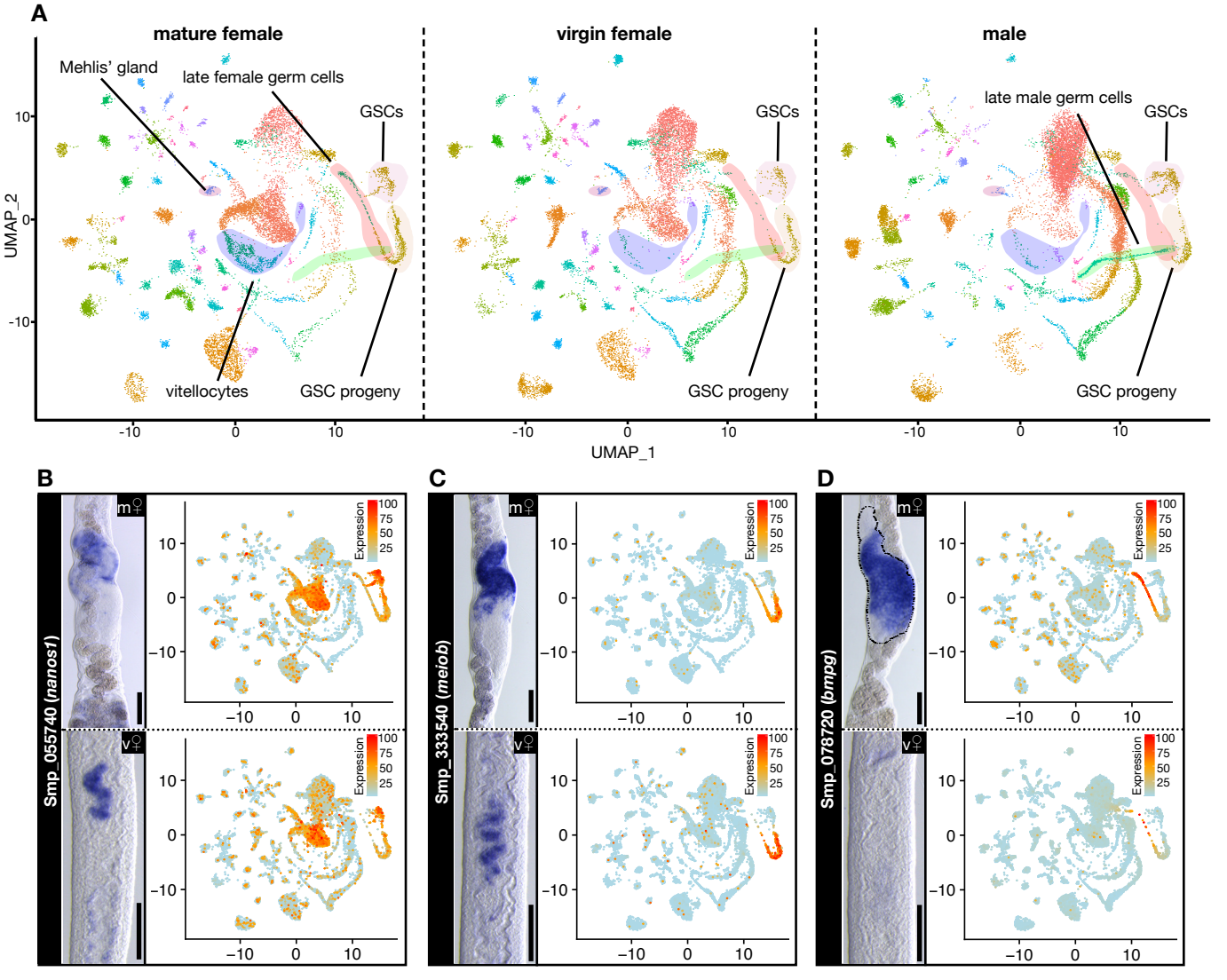
(**A**) UMAP plots of the expression pattern of the indicated gene on the (top) original dataset or the (bottom) re-clustered dataset, and (right) a colorimetric WISH of a male parasite's trunk for *eled*, *hnf4*, *prom2*, and *ctsb*. Insets: magnifications of dashed boxes. (**B**) Fluorescence *in situ* hybridization (FISH) and EdU labeling showing the expression of *eled* (green) and EdU^+ proliferative cells (yellow) in control or *hnf4(RNAi)* animals. $n = \geq 18$ parasites, two biological replicates. (**C**) FISH of *ctsb* (cyan) and fluorescent dextran (red) in the gut lumen in control(RNAi)

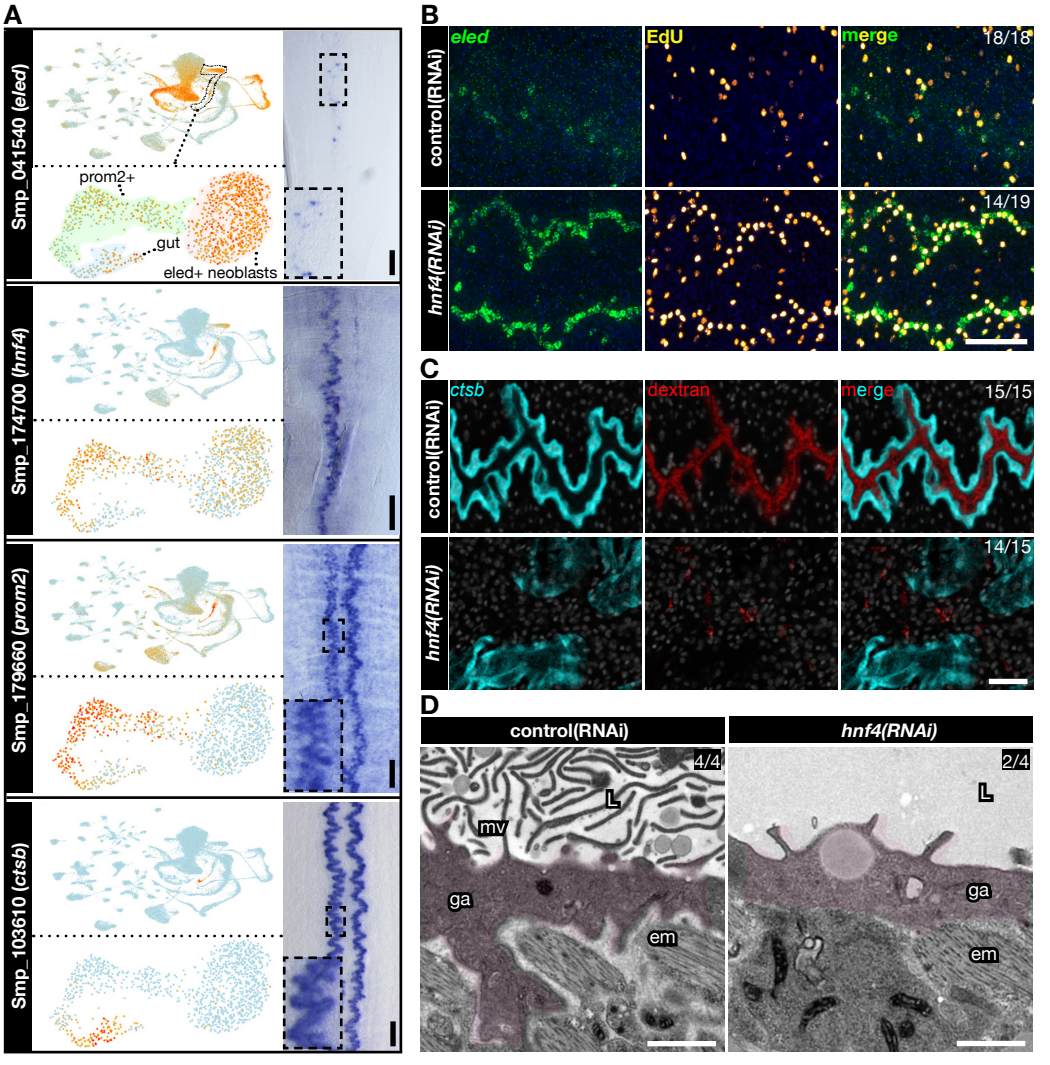
and *hnf4(RNAi)* animals. *n* = 15 parasites, three biological replicates. (**D**) TEM micrographs showing gut of control(RNAi) and *hnf4(RNAi)* animals. 'mv' microvilli, 'ga' gastrodermis, 'L' lumen, 'em' enteric muscle. *n* = 4 parasites, two biological replicates. Nuclei: blue (B) or grey (C). The number of parasites similar to the representative image is in the upper right of each panel. Scale bars: A, 100 μm, B, 50 μm, C, 20 μm, D, 1 μm. UMAP plots are colored by gene expression (blue = low, red = high).

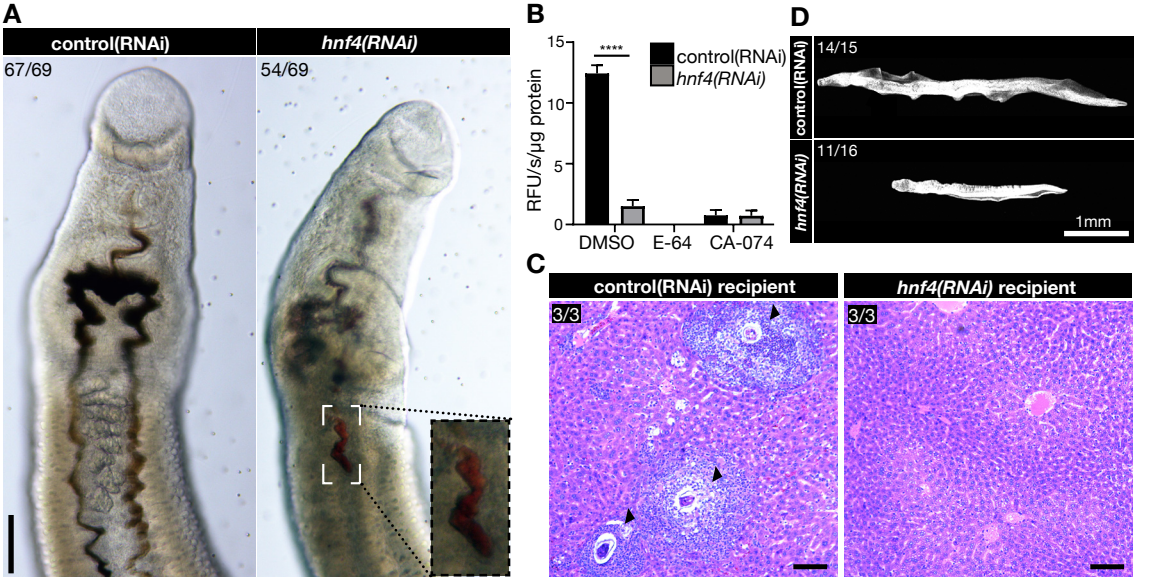
Fig. 4. *hnf4* is required for blood feeding and pathology.

(**A**) Brightfield images of control(RNAi) or *hnf4(RNAi)* animals cultured with red blood cells. Inset: magnification of boxed area. (**B**) Cathepsin activity of lysates from control(RNAi) or *hnf4(RNAi)* animals determined by cleavage of Z-FR-AMC with no inhibitor (DMSO), a broad cysteine protease inhibitor (E-64), or a cathepsin B-selective inhibitor (CA-074). *n* = 3, three biological replicates. (**C**), H&E-stained mouse liver sections 22 days post-transplant with RNAi-treated parasites. Arrows: granulomata. Sections from *n* = 3 recipients. (**D**) Parasites recovered from transplant recipients. *n* > 15 from three recipients. Nuclei: white. The number of parasites/sections similar to the representative micrograph is in the upper left of each panel. Scale bars: A,C, 100μm, D, 1 mm. ****, $p < 0.0001$ (Welch's t-test).


A
neurons
neoblasts
vitellaria
germline
gut
esophageal gland
muscle
flame cells
parenchyma
tegument lineage
UMAP_1
UMAP_2
B
neoblasts
Smp_051920 (nanos2)
C
parenchyma
Smp_212710 (tgfbi)
D
flame cells
Smp_335600 (sialidase)
E
muscle
Smp_031770 (tpm2)
F
esophageal gland
Smp_172180 (meg-8)
G
tegument
Smp_346900 (sm25)
H
gut
Smp_103610 (ctsb)
I
neurons
Smp_073270 (7b2)
Expression
100
75
50
25


A
mature female
virgin female
male
Mehlis' gland
late female germ cells
GSCs
vitellocytes
GSC progeny
late male germ cells
UMAP_1
UMAP_2
B
C
D
Smp_055740 (nanos1)
Smp_333540 (meiob)
Smp_078720 (bmpg)
m♀
v♀
Expression
100
75
50
25


A
Smp_041540 (eled)
prom2+
gut
eled+ neoblasts
Smp_174700 (hnf4)
Smp_179660 (prom2)
Smp_103610 (ctsb)
B
eled
EdU
merge
18/18
control(RNAi)
hnf4(RNAi)
14/19
C
ctsb
dextran
merge
15/15
control(RNAi)
hnf4(RNAi)
14/15
D
control(RNAi)
hnf4(RNAi)
4/4
2/4
L
mv
ga
em


A
control(RNAi)
hnf4(RNAi)
67/69
54/69
B
RFU/s/µg protein
15
10
5
0
****
control(RNAi)
hnf4(RNAi)
DMSO
E-64
CA-074
D
14/15
11/16
control(RNAi)
hnf4(RNAi)
1mm
C
control(RNAi) recipient
hnf4(RNAi) recipient
3/3
3/3

Science
AAAS

Supplementary Materials for

A single-cell RNAseq atlas of *Schistosoma mansoni* identifies a key regulator of blood feeding

George Wendt, Lu Zhao, Rui Chen, Chenxi Liu, Anthony J. O'Donoghue, Conor R. Caffrey, Michael L. Reese, James J. Collins III.

Correspondence to: JamesJ.Collins@UTSouthwestern.edu

This PDF file includes:

Other Supplementary Materials for this manuscript include the following:

Materials and Methods

Parasite acquisition and culture

Adult *S. mansoni* (NMRI strain, 6–7 weeks post-infection) were obtained from infected female mice by hepatic portal vein perfusion with 37°C DMEM (Sigma-Aldrich, St. Louis, MO) plus 10% Serum (either Fetal Calf Serum or Horse Serum) and heparin. Parasites were cultured as previously described (*5*). Unless otherwise noted, all experiments were performed with male parasites. Experiments with and care of vertebrate animals were performed in accordance with protocols approved by the Institutional Animal Care and Use Committee (IACUC) of UT Southwestern Medical Center (approval APN: 2017-102092).

Fluorescence Activated Cell Sorting

FACS sorting was performed as previously described (*4*) with minor modifications. Freshly perfused adult male and sexually mature adult female worms were separated by incubation in a 0.25% solution of tricaine (*3*) for approximately 5 minutes. Sexually immature adult virgin female worms were separately perfused from single-sex infected mice. Male, mature female, or virgin female worms were suspended in a 0.5% solution of Trypsin/EDTA (Sigma T4174) in PBS. The worms were then triturated for approximately 10 minutes (mature and virgin females) or 15 minutes (males) until the solution became turbid and no large pieces of worms were left. The trypsin was inactivated by adding an equal volume of serum-containing media. The dissociated worms were then centrifuged at 500 *g* for 10 m at 4°C. Next the worms were resuspended in 1 ml of Basch medium with 10 µL of RQ1 DNAse (Promega M6101) and incubated for 10 minutes at RT. The dissociated worms were centrifuged again at 500 *g* for 10 minutes at 4°C. The cells were then resuspended in 1mL of staining media (0.2% BSA, 2mM EDTA in PBS, pH7.40) and incubated in Hoechst 33342 (18 µg/ml) (Sigma B2261) for 1 hour at RT in the dark. 9mL of staining media was then added to the worms and then the whole suspension was filtered through a 40 µm cell strainer. The worms were centrifuged once again at 500 *g* for 10 minutes at 4°C. Worms were then resuspended in 1mL of staining media containing Hoechst 33342 (18 µg/ml) and propidium iodide (1 µg/ml) (Sigma-Aldrich P4170) and then filtered once more through a 40 µm cell strainer into a 12x75mm FACS tube. Filtered cells were then sorted on a FACSAria II custom (BD Biosystems) with 305/405/488/561/633nm lasers. Sorts were performed with a 100 µm nozzle and cells were sorted into sorting media (0.2% BSA in PBS, pH7.40). Live single cells (PI negative, singlet by comparing forward scatter height to forward scatter width) were sorted using a 100 µm nozzle and cells were sorted into sorting media (0.2% BSA in PBS, pH7.40). For all FACS experiments, a Hoechst threshold was applied to exclude debris and improve the efficiency of sorting.

Single-cell RNA sequencing

FACS-sorted cells were centrifuged again at 500 *g* for 10 minutes at 4°C then resuspended in 0.2% BSA in PBS. Libraries were created using a Chromium Controller (10x Genomics) according to manufacturer guidelines and sequenced in using a NextSeq 500 (illumina). Sequencing data was processed and mapped to the *Schistosoma mansoni* genome (v7) using Cell Ranger (10x Genomics). Unfiltered data from Cell Ranger was imported into Seurat (v3.1.1)(*24, 25*) and cells were filtered as follows: Female (nFeature_RNA (> 750), nCount_RNA (1500-20000), Percent Mitochondrial (<3%); Male/Virgin female (nFeature_RNA (> 750), nCount_RNA (1000-20000), Percent Mitochondrial (<3%)). Mitochondrial genes were identified as those with the prefix "Smp_9". Each of the 9 individual datasets (Table S5) was normalized (NormalizeData) and

variable features were identified (FindVariableFeatures, selection.method = "vst", nfeatures = 2000). From here, integration anchors were identified (FindIntegrationAnchors, dims 1:78), the data was integrated (IntegrateData, dims = 1:78, features.to.integrate = features), and scaled (ScaleData). We then ran RunPCA, RunUMAP (reduction = "pca", dims = 1:78, n.neighbors = 40), FindNeighbors (reduction = "pca", dims = 1:78), FindClusters (resolution = 5). The number of principal components (78) used for this analysis was defined by JackStraw. Analysis of the resulting single cell map found that clusters 27 and 50 contained few enriched markers, therefore we removed the 964 cells present in these clusters and reran the analysis with 78 principal components. From here we generated the final UMAP projection plot with RunUMAP (n.neighbors = 36, min.dist = 0.70, dims = 1:80). Next, we generated clusters (FindClusters, resolution = 5) and manually inspected the unique genes expressed in each of the clusters. In some cases we found that some of the 85 resulting clusters did not express a core set of unique genes, therefore, these clusters were merged into a single cluster of cells as follows: Neoblasts (clusters 0,1,2,6,7,37), Neoblast progeny (cluster 4,8), Neuron 1 (clusters 10, 60, 68), Neuron 6 (clusters 24, 26), Parenchyma (clusters 11, 12, 51), flame cells (clusters 14, 41), S1 Cells (clusters 3, 9, 32, 42) and tegument (clusters 36, 63). After merging we were left with a final map of 68 clusters of 43,642 cells. In some cases, we noted large differences or unexpected similarities in several cell populations between which we reasoned could be attributable to technical factors (i.e. male dissociation takes longer than female dissociation) which could disproportionately affect some cell types in male parasites vs. female parasites (i.e. more muscle cells were recovered from female parasites than from males, despite male parasites possessing more muscle cells per animal). In order to confirm our clustering results and also to ensure that other factors such as differences in read depth did not impact our analysis, we collapsed major cluster types (neurons, neoblasts, etc.) and reanalyzed our data using the sctransform (that uses a negative binomial normalization rather than log transformation used in our initial analysis) function from Seurat v3.1(38) (Fig. S18). Briefly, we ran SCTransform on all nine of our samples individually with default setting. Next we ran SelectIntegrationFeatures on all nine of our samples (nfeatures = 3000) to create a list of integration features ("sct features"). We then ran PrepSCTIntegration on all nine of our samples (anchor.features = "sct features") using default settings. After this we ran FindIntegrationAnchors on all nine of our samples (normalization.method = "SCT", anchor.features = "sct features") using the default settings to generate our integration anchors ("integration anchors"). We then ran IntegrateData (anchorset = "integration anchors", normalization.method = "SCT") with default settings. Next we ran RunPCA on the dataset (npcs = 200). Finally, we ran RunUMAP (n.neighbors = 50, min.dist = 0.20, dims = 1:200) and verified that our between-sample and between sex clustering results were largely the same. We still observed fewer than anticipated muscle cells in male samples, which we attribute to the relatively harsher dissociation conditions experienced by male parasites. Similar biases in cell recovery have been described in the dissociation of related flatworms (*1*).

Heat maps for Figs. S4 and S5C were generated as follows: first, all neuron clusters (Fig. 1H) or all muscle clusters (Fig. Fig. 1E) were used to make a new Seurat object with the subset() command. Then the markers of these objects were found using the FindAllMarkers() command with the parameters test.use = "roc", only.pos = TRUE. The subset of genes to use for the heatmap was then determined by using the subset() command on those markers with the parameters avg_diff>0 and power >0.9 (neuron) or >0.8 (muscle) for each gene. Heatmaps were generated with the DoHeatmap() command on the new Seurat objects with the subset of genes for the heatmap using default parameters.

The dot plot for Fig. S14D was generated as follows: first, clusters were simplified by collapsing all 31 neuron clusters into a single cluster ("Neurons"), all parenchyma clusters into a single cluster ("Parenchyma"), all muscle clusters into a single cluster ("Muscle"), all tegument clusters into a single cluster ("Tegument"), all germline clusters ("GSCs", "GSC progeny", "late female germ cells", and "late male germ cells") into a single cluster ("Germline"), all neoblast progeny cells ("neoblast progeny" and "hes2+") into a single cluster ("Neoblast Progeny"), all tegument progenitor cells ("early tsp-2+", "egc+", "meg-1+", "zfp-1-1+", and "sm13+") into a single cluster ("Tegument Progenitors"), all vitelline cells ("S1", "S1 progeny", "early vitellocytes", "late vitellocytes", and "mature vitellocytes") into a single cluster ("Vitellaria"), and two neoblast clusters ("neoblast 1" and "neoblast 2") into a single cluster ("Neoblasts"). Next, a dotplot was generated using the DotPlot() function in Seurat v3.1 with the 25 most down-regulated genes following *hnf4* RNAi (see Table S4). The size of the dot corresponds to the percentage of the cells in the cluster (indicated on the vertical axis) that express the given gene (indicated on the horizontal axis), whereas the color of the dot indicates the average expression level of the gene in the cluster.

Parasite labeling and imaging

Colorimetric and fluorescence in situ hybridization analyses were performed as previously described (*3, 5*) with the following modification. To improve signal-to-noise for colorimetric in situ hybridization, all probes were used at 10 ng/mL in hybridization buffer. *In vitro* EdU labeling and detection was performed as previously described(*3*). For dextran labeling of the parasite gut, 10 male RNAi-treated parasites were given 10μL/mL of 5 mg/mL (in water) solution of biotin-TAMRA-dextran (Life Technologies D3312) and cultured 12 hours. The parasites were then fixed in fixative solution (4% formaldehyde in PBSTx (PBS + 0.3% triton-X100)) for 4 hours in the dark with mild agitation. Worms were then washed with 10 ml of fresh PBSTx for 10 minutes, then dehydrated in 100% methanol and stored at -20dC until used in fluorescence in situ hybridization as described(*3, 5*). All fluorescently labeled parasites were counterstained with DAPI (1 μg/ml), cleared in 80% glycerol, and mounted on slides with Vectashield (Vector Laboratories).

Transmission electron microscopy samples were prepared from RNAi-treated parasites that were immersed in fixative (2.5% glutaraldehyde in 0.1M sodium cacodylate buffer pH 7.4 with 2mM $CaCl_2$) and then amputated at the head and the tail in order to retain ~5mm of trunk. After three rinses with 0.1 M sodium cacodylate buffer, the parasite trunks were embedded in 3% agarose and sliced into small blocks ($1mm^3$), rinsed with the fixative three times and post-fixed with 1% osmium tetroxide and 0.8 % Potassium Ferricyanide in 0.1 M sodium cacodylate buffer for one and a half hours at room temperature. Samples were rinsed with water and *en bloc* stained with 4% uranyl acetate in 50% ethanol for two hours. They were then dehydrated with increasing concentration of ethanol, transitioned into propylene oxide, infiltrated with Embed-812 resin and polymerized in a 60°C oven overnight. Blocks were sectioned with a diamond knife (Diatome) on a Leica Ultracut 7 ultramicrotome (Leica Microsystems) and collected onto copper grids, post stained with 2% aqueous Uranyl acetate and lead citrate. Images were acquired on a Tecnai G2 spirit transmission electron microscope (FEI, Hillsboro, OR) equipped with a LaB_6 source at 120kV using a Gatan ultrascan CCD camera.

Blood in the parasite gut and Haematoxylin and Eosin stained-samples were imaged with brightfield light using a Zeiss AxioZoom V16 equipped with a transmitted light base and a Zeiss AxioCam 105 Color camera.

Confocal imaging of fluorescently labeled samples was performed on a Nikon A1 Laser Scanning Confocal Microscope. Unless otherwise mentioned, all fluorescence images represent maximum intensity projection plots. To perform cell counts, cells were manually counted in maximum intensity projection plots derived from confocal stacks. In order to normalize counts, we collected confocal stacks and normalized the number of cells counted to the length of the parasite in the imaged region. Brightfield images were acquired on a Zeiss AxioZoom V16 equipped with a transmitted light base and a Zeiss AxioCam 105 Color camera.

RNA interference

For detailed schematic of RNAi experiments, see Table S6. Generally, all experiments utilized freshly perfused male parasites (separated from females) unless otherwise noted. dsRNA treatments were all carried out at 30 μg/ml in Basch Medium 169. dsRNA was generated by *in vitro* transcription and was replaced as indicated in Table S4. EdU pulses were performed at 5μM for 4 hours before either fixation or chase as previously described(*3*).

As a negative control for RNAi experiments, we used a non-specific dsRNA containing two bacterial genes (*26*). cDNAs used for RNAi and in situ hybridization analyses were cloned as previously described (*26*); oligonucleotide primer sequences are listed in Table S7.

qPCR and RNAseq

RNA collection was performed as previously described(*5*) with the following modifications. Parasites were treated with dsRNA as described in Table S6 ("strategy 4") and whole parasites were collected in Trizol. RNA was purified from samples utilizing Direct-zol RNA miniprep kits (Zymo Research R2051). Quantitative PCR analyses were performed as previously described (*3, 5*). cDNA was synthesized using iScript™ cDNA synthesis kit (Bio-Rad 1708891) and qPCR was performed as previously described(*4*) utilizing iTaq™ Universal SYBR® Green Supermix (Bio-Rad 1725122) and a QuantStudio 3 Real-Time PCR System (Applied Biosystems); oligonucleotide primer sequences used for qPCR are listed in Table S7. RNAseq on *hnf4(RNAi)* parasites was performed as previously described(*4*) using TruSeq Stranded mRNA Library Prep (illumina 20020594) to prepare libraries, which were sequenced on a NextSeq 550 (illumina). The total number of reads per gene was determined by mapping the reads to the *S. mansoni* genome (v7) using STAR (version 020201)(*27*). *S. mansoni* genome sequence and GTF files used for mapping were acquired from Wormbase Parasite(*28*). Pairwise comparisons of differential gene expression were performed with DESeq2 (version 1.12.2)(*29*). Volcano plots were made with using the "volc" function from ggplot2. To remove genes expressed at lower levels, genes with a base-mean expression value less than 50 were excluded from analysis. Furthermore, genes that were differentially expressed ($p_{adj} < 0.05$) that were not assigned to the automatically assigned to the "gut" cluster during initial clustering were manually examined in the single-cell RNAseq data and those that were expressed in the gut were reclassified to the "gut" cluster. Raw data from *hnf4* RNAi RNAseq experiments are available at NCBI GEO with accession number GSE146737.

Protease activity assays

To measure cysteine protease cathepsin activity (*30*), five worms of each RNAi condition (see Table S6 "strategy 7") were ground and sonicated in 300 μL assay buffer (0.1 M citrate-phosphate, pH 5.5). The lysate was centrifuged at 15,000*g* for 5 minutes and the pellet was discarded. The total protein concentration was calculated using the bicinchoninic acid assay with bovine serum albumin as the protein standard. Each well in the assay had 1 μg of protein. The assay buffer was

0.1 M citrate-phosphate, pH 5.5 with 2 mM DTT. CA-074 (Cayman Chemical, 24679-500) and E-64 (Alfa Aesar, J62933) controls were set up by incubating the sample with 10 μM of each inhibitor for 30 min at room temperature. The final substrate concentration of Z-FR-AMC (R&D Systems, ES009) was 10 μM. The release of the AMC fluorophore was recorded in a Synergy HTX multi-mode reader (BioTek Instruments, Winooski, VT) with excitation and emission wavelengths at 340nm and 460nm, respectively.

To measure aspartic protease cathepsin activity, five worms of each RNAi condition (See Table S6 "strategy 7") were ground and sonicated in 300 μL assay buffer (0.1 M citrate-phosphate, pH 5.5). The lysate was centrifuged at 15,000*g* for 5 mins and the pellet was discarded. Each well in the assay had 1 μg of protein. The assay buffer was 0.1 M citrate-phosphate, pH 3.5. Pepstatin A (MP Biomedicals, 0219536805) and E-64 controls were set up by incubating the sample with 10μM of either inhibitor for 30 minutes at room temperature. The final substrate concentration of mca-GKPILFFRL-K(dnp) (CPC Scientific, SUBS-017A) was 10μM. The release of the AMC fluorophore was recorded in a Synergy HTX multi-mode reader (BioTek Instruments, Winooski, VT) with excitation and emission wavelengths at 320nm and 400nm, respectively.

Surgical transplantation of schistosomes

Surgical transplantation was performed as previously described(*31*) with the following modifications. Seven days prior to surgery, 5-week-old parasites were recovered from mice and treated with 30 μg/ml dsRNA for 7 days in Basch Medium 169 (see Table S6 "strategy 8"). Before mice were anesthetized, 10 pairs (male and female) were sucked into a 1ml syringe, the syringe was fitted with a custom 25G extra thin wall hypodermic needle (Cadence, Cranston, RI), the air and all but ~200 μL of media were purged from the needle, and the syringe was placed needle down in a test tube to settle the parasites to the bottom of the syringe. Mice were kept on infrared heating pads for the duration of the surgery. Following wound closure, mice received a single subcutaneous 20 μL dose of a 1 mg/mL solution of Buprenorphine SR-LAB CIII for analgesia and were allowed to recover on a warm heating pad. Mice were group housed and individual mice were tracked by ear punches. On either day 22 or day 30 post-transplantation mice were sacrificed and perfused to recover parasites. Male and female parasites were counted and fixed for 4 hours in 4% formaldehyde in PBSTx. Recipient livers were removed and fixed for 72 hours in 4% formaldehyde in PBS. The percentage parasite recovery was determined by dividing the total (male and female) number of worms transplanted by the total number of parasites recovered following perfusion. Livers from individual mice were sectioned and processed for Haematoxylin and Eosin staining by the UT Southwestern Molecular Pathology Core.

Statistical analysis

All two-way comparisons were analyzed using Welch's t-test. All three-way comparisons were analyzed using one-way ANOVA. RNAseq data was analyzed by the Benjamini-Hochberg corrected Wald test in DeSeq2. *p* values are indicated in the figure legends or in Table S2. Power calculations for transplantation experiments were performed based on previously published data (*31*) (75% recovery vs. 0% recovery, alpha 0.05, beta 0.8)

Supplementary Text

Though our scRNAseq atlas identified many of the anticipated biological differences between male and female parasites for many reproductive tissues (i.e. "late male germ cells" being present only in males), some cells that belong to clusters that should largely be restricted to female parasites appear with unexpectedly high frequency in male parasites, namely the "late female germ cells", "mature vitellocytes", and "Mehlis' gland" (Fig. S11). In order to understand this phenomenon, we examined the expression of cluster-specific markers for each cluster individually. This examination revealed that these clusters were not homogenous, but rather composed of essentially two populations: one that expressed high levels of the cluster-specific markers and a separate population that expressed lower levels. Given the observation that planarian neoblasts sometimes express markers of differentiated tissues at low levels (*1*), we hypothesized that the cells expressing low levels of cluster-specific markers might be neoblast-like cells. Upon examining the expression of the neoblast marker *nanos2*, we did indeed find this to be the case for all three clusters.

Examination of the "late female germ cells" cluster (Fig. S11A) reveals that *bmpg*$^{+}$ late female germ cells make up almost the entire cluster in sexually mature females whereas the *nanos2*$^{+}$ neoblast-like cells make up the majority of the cells in both males and virgin females. This agrees with the known biology; sexually mature female parasites possess a functional ovary that robustly produces mature oocytes. In contrast, virgin female parasites produce minimal late germ cells (Fig. 2D). The few "late female germ cell"-like cells in males likely come from rare hermaphroditic ovary-bearing males (*32*).

Examination of the "mature vitellocyte" cluster (Fig. S11B) reveals a similar phenomenon. *ataxin2*$^{+}$ mature vitellocytes are highly abundant in sexually mature females but rare in virgin females and males. *nanos2*$^{+}$ neoblast-like cells showed the opposite pattern: sparse in sexually mature females and relatively abundant in virgin females and males. Once again, this agrees with the biology: sexually mature female parasites should robustly produce mature vitellocytes but these cells should be rare within virgin females and males (Fig. S9A).

Finally, examining the "Mehlis' gland" cluster (Fig. S11C) revealed similar findings. *Smp_343210*$^{+}$ Mehlis' gland cells dominate the cluster in sexually mature and virgin female parasites but are sparse in male parasites. In contrast, *nanos2*$^{+}$ neoblast-like cells were abundant in male parasites (and to a lesser extent virgin female parasites). This, once again, agrees with the biology: the Mehlis' gland is a pairing-independent tissue found in sexually mature and virgin females but not appreciably in males (Fig. S9E). The *Smp_343210*$^{+}$ cells in male parasites could represent a primordial Mehlis' gland.

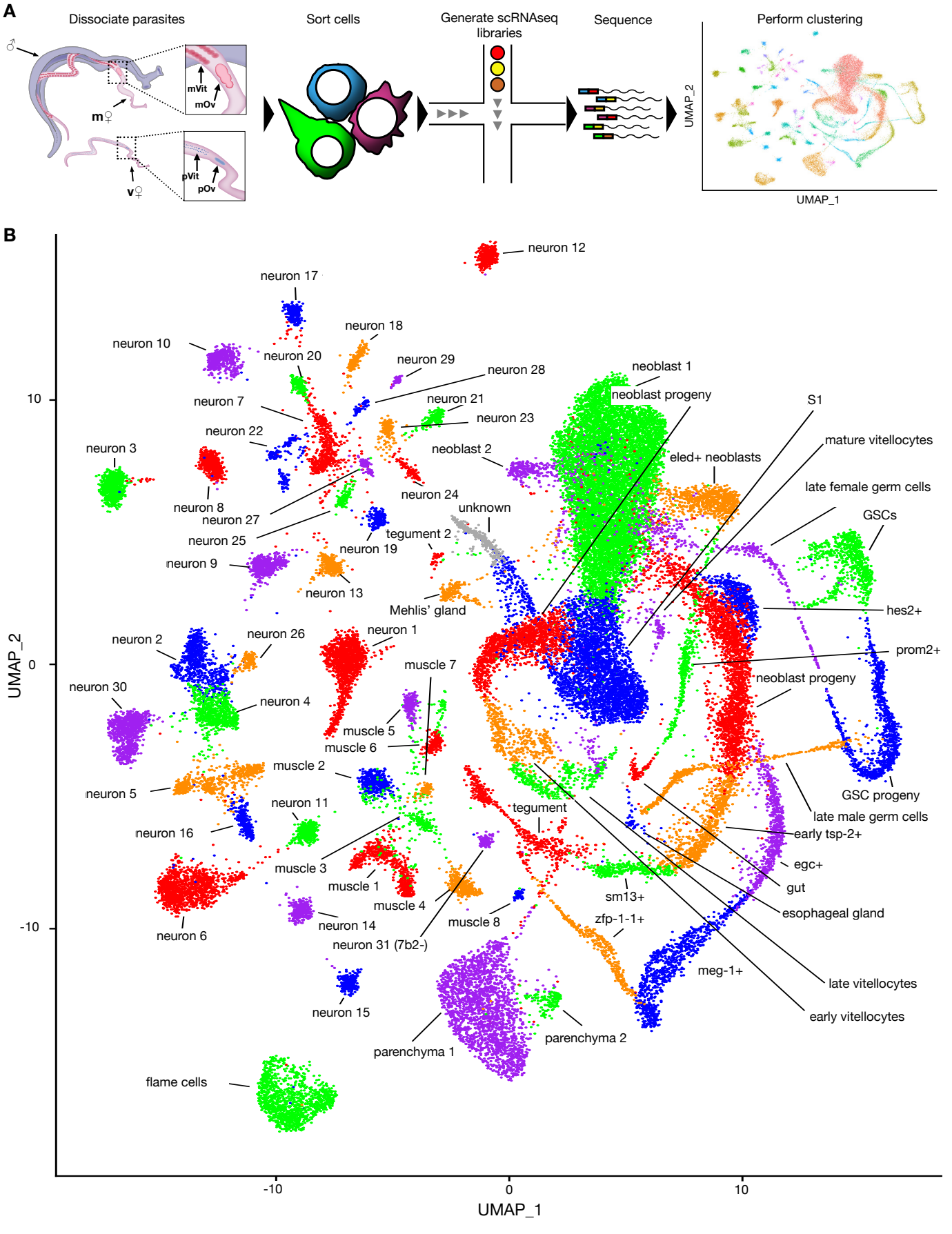


Fig. S1. Overview of single-cell RNA sequencing of schistosome cells. (**A**) Schematic diagram of single-cell RNA sequencing workflow. Cartoon to left depicts male paired with a mature female worm (m♀) that possess a mature ovary (mOv) and vitellaria (mVit); unpaired virgin female worms (v♀) possess a primordial ovary (pOv) and vitellaria (pVit). (**B**) Uniform Manifold Approximation and Projection (UMAP) plot of all clusters with labels.

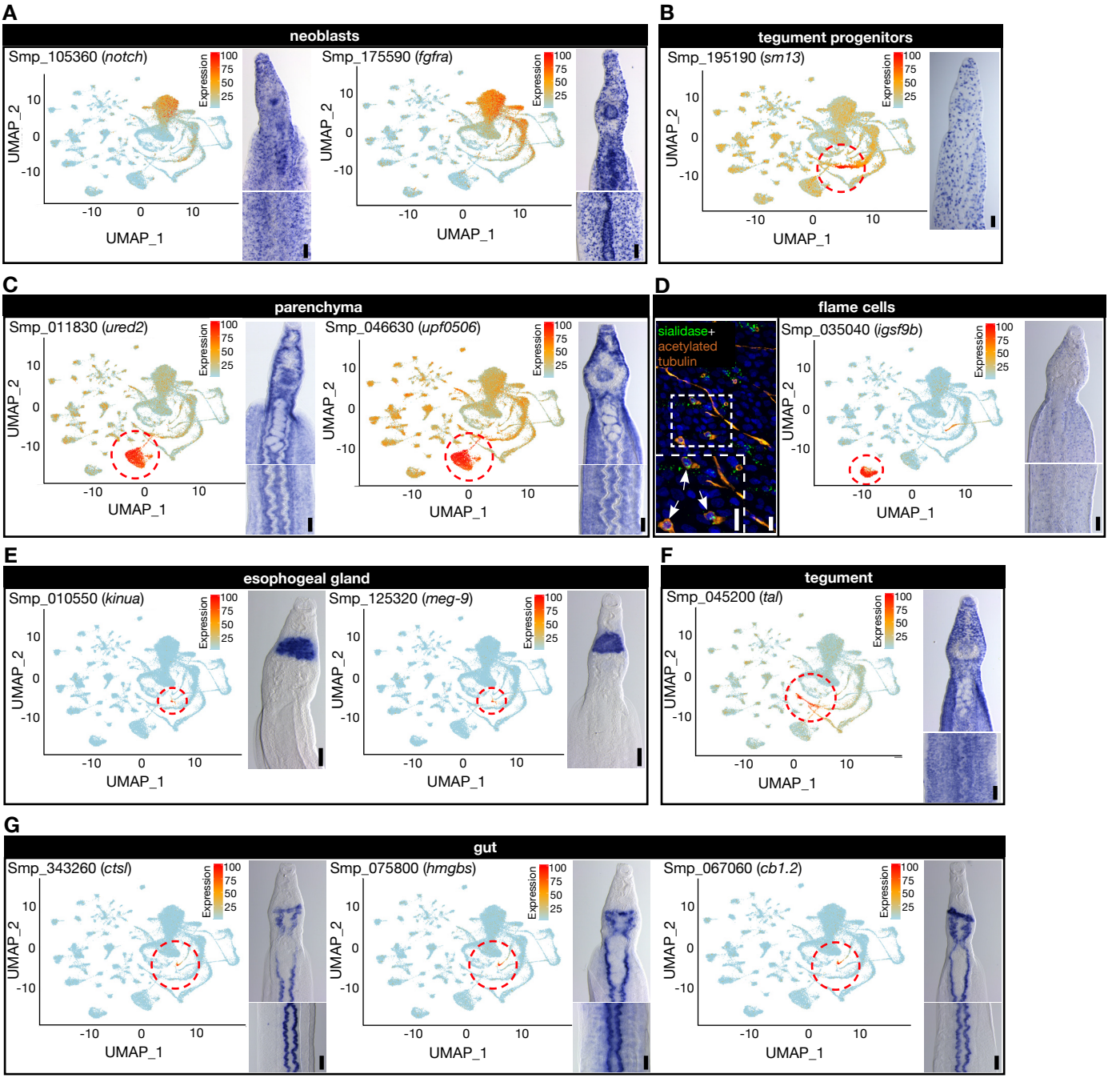


Fig. S2. Additional somatic tissue-specific genes. (**A**) (left) UMAP plot and (right) whole-mount *in situ* hybridization (WISH) of neoblast-specific genes *notch* and *fgfra*. (**B**) (left) UMAP plot and (right) WISH of tegument progenitor-specific gene *sm13*. (**C**) (left) UMAP plot and (right) WISH of parenchyma-specific genes *ured2* and *upf0506*. (**D**) (left) fluroscence *in situ* hybridization (FISH) with acetylated tubulin immunofluorescence to label cilia, (middle) UMAP plot, and (right) WISH of flame cell-specific gene *igsf9b*. (**E**) (left) UMAP plot and (right) WISH of esophageal gland-specific genes *kinua* and *meg-9*. (**F**) (left) UMAP plot and (right) WISH of tegument-specific gene *tal*. (**G**) (left) UMAP plot and (right) WISH of gut-specific genes *ctsl*, *hmgbs*, and *cb1.2*. Scale bars, D, left panel: 10 μm; all others: 100 μm. UMAP plots are colored by gene expression (blue = low, red = high).

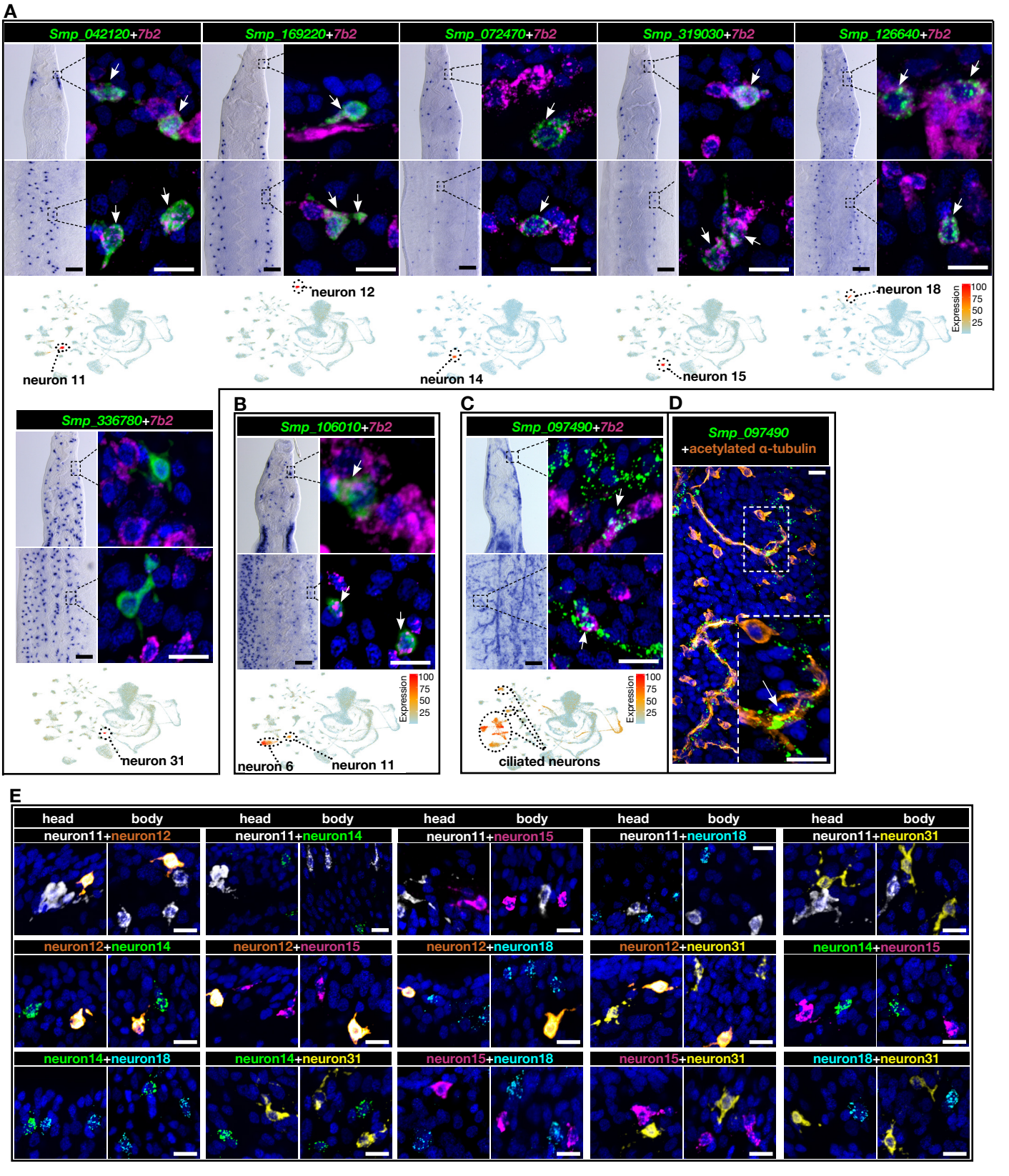


Fig. S3. Schistosome neurons display complex heterogeneity. (**A**) For each of 6 different neuron cluster-specific genes (top left) WISH of head, (top right) double FISH of region of head indicated to the left with cluster specific gene (green) and *7b2* (magenta), (middle left) WISH of body, (middle right) double FISH of region of body indicated to the left with cluster specific gene (green) and *7b2* (magenta), and (bottom) UMAP plot. (**B** and **C**) (top left) WISH of head, (top right) double FISH of region of head indicated to the left with (**B**) "neuron 6"- and "neuron 11"- enriched gene *Smp_106010* (green) or (**C**) the ciliated neuron-enriched gene *Smp_097490* (green) and *7b2* (magenta), (middle left) WISH of body, (middle right) double FISH of region of body indicated to the left with *Smp_106010* (green) and *7b2* (magenta), and (bottom) UMAP plot. (**D**) FISH of *Smp_097490* (green) with immunofluorescent labeling of acetylated tubulin (orange) and (bottom) UMAP plot. (**E**) Double FISH with the indicated combination of neuron cluster-specific markers. Nuclei: blue. Scale bars, all FISH: 10 μm; all colorimetric WISH: 100 μm. UMAP plots are colored by gene expression (blue = low, red = high).

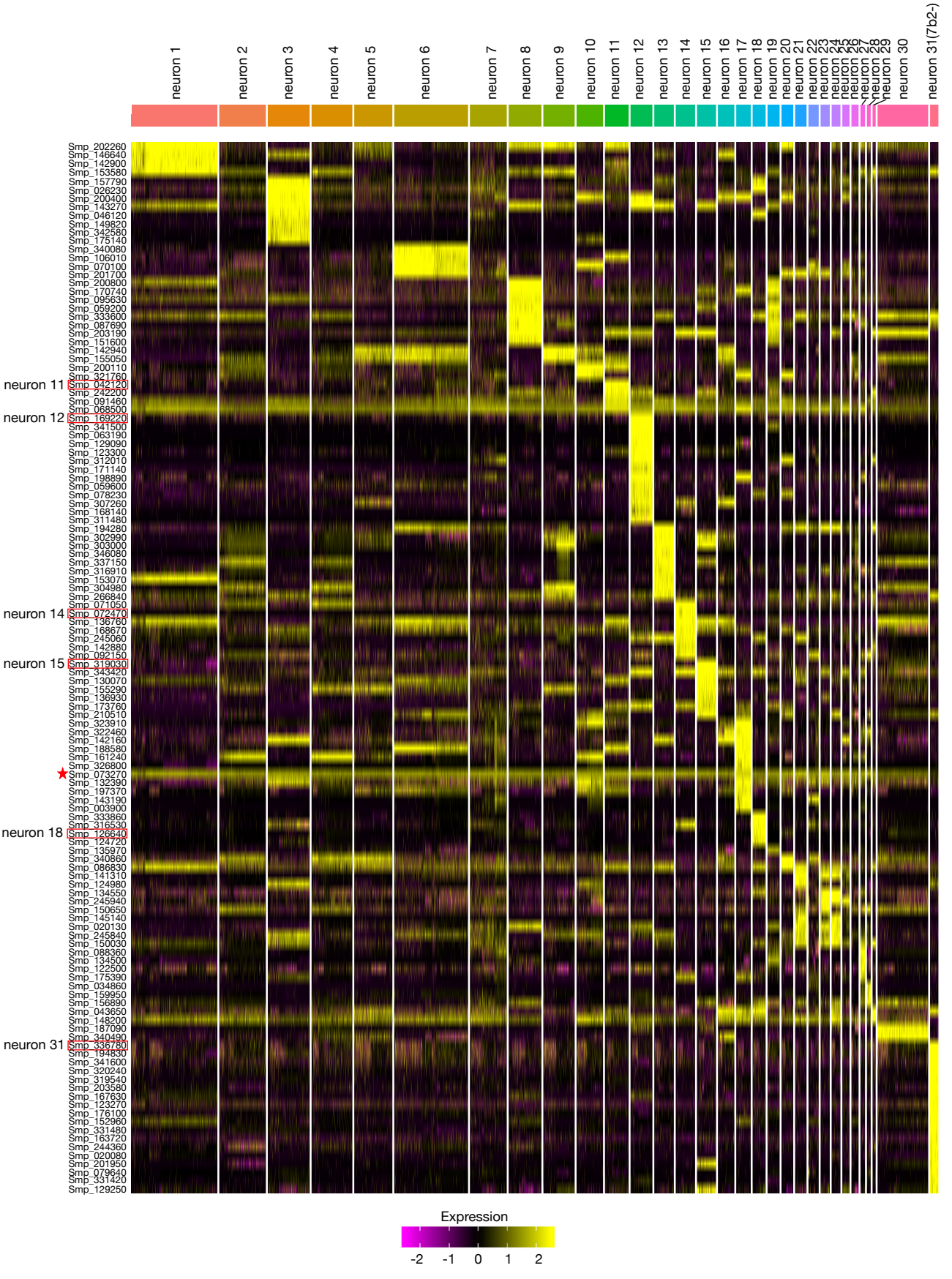
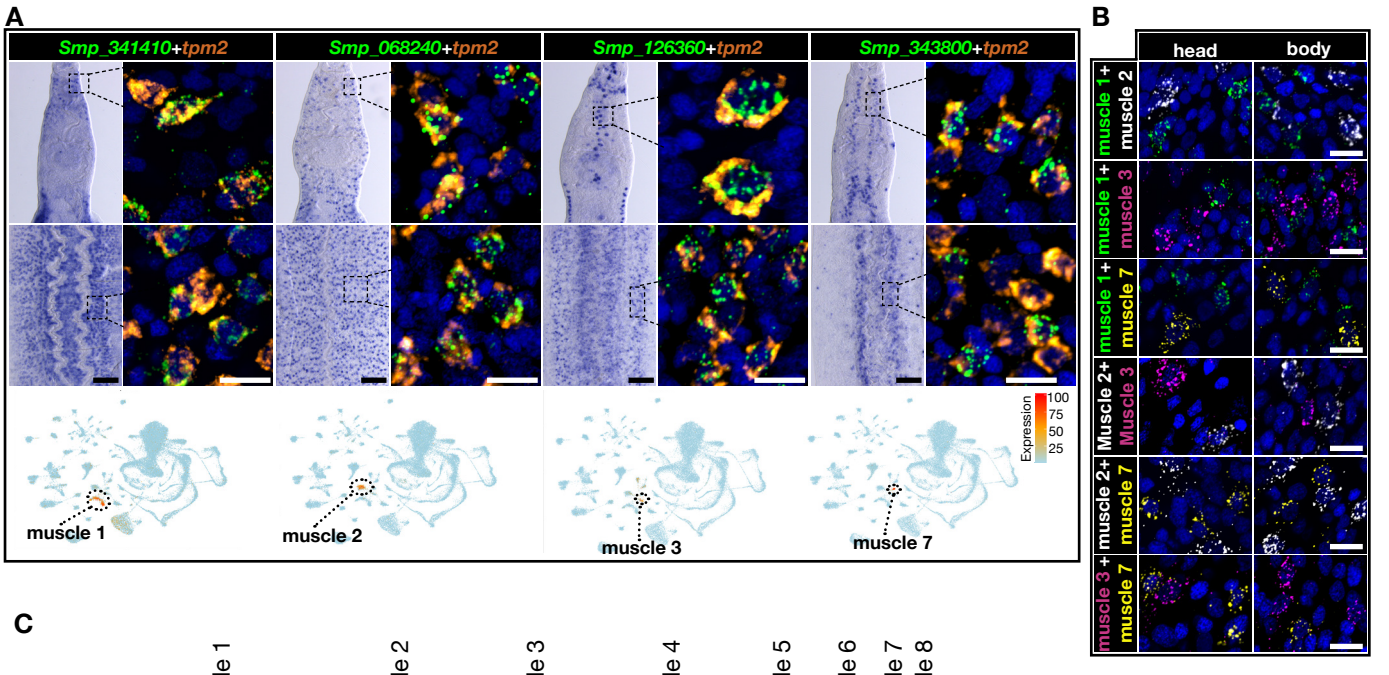


Fig. S4. Heatmap of neuronal marker genes. The six different neuron cluster-specific genes shown in Fig. S3A are labeled and indicated by a red rectangle. The general neuronal marker (for 30/31 clusters) *7b2* (Smp_073270) is indicated with red asterisk. Each row represents a gene and each column represents a cell. Clusters are indicated on the top of the panel. Heatmaps are colored by gene expression (purple = low, yellow = high).



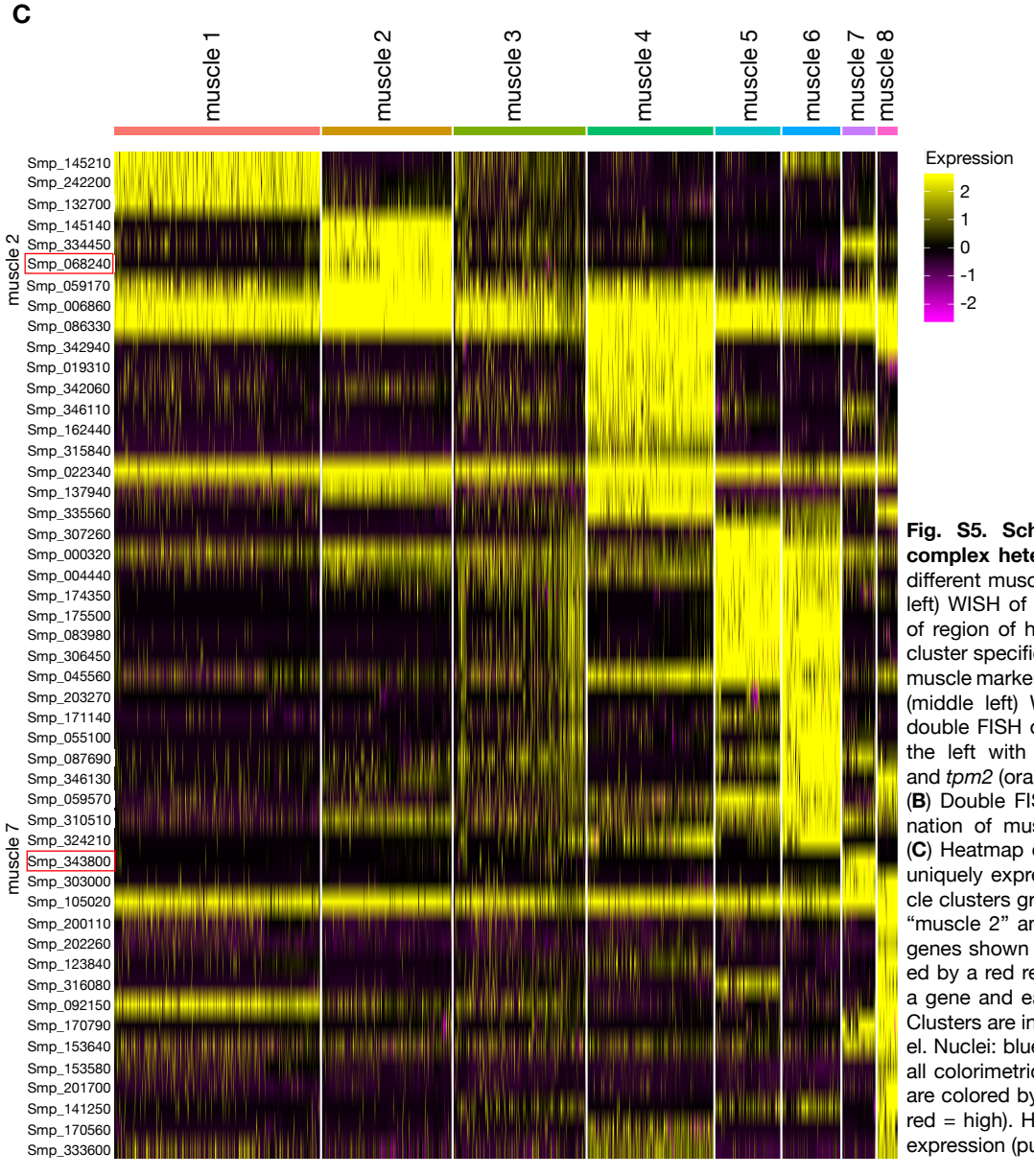


Fig. S5. Schistosome muscles display complex heterogeneity. (A) For each of 4 different muscle cluster-specific genes: (top left) WISH of head, (top right) double FISH of region of head indicated to the left with cluster specific gene (green) and the general muscle marker *tropomyosin2* (*tpm2*, orange), (middle left) WISH of body, (middle right) double FISH of region of body indicated to the left with cluster-specific gene (green) and *tpm2* (orange), and (bottom) UMAP plot. (**B**) Double FISH with the indicated combination of muscle cluster-specific markers. (**C**) Heatmap of the expression of the most uniquely expressed genes from the 8 muscle clusters grouped by cluster number. The "muscle 2" and "muscle 7" cluster-specific genes shown in (A) are labeled and indicated by a red rectangle. Each row represents a gene and each column represents a cell. Clusters are indicated on the top of the panel. Nuclei: blue. Scale bars, all FISH: 10 µm, all colorimetric WISH: 100 µm. UMAP plots are colored by gene expression (blue = low, red = high). Heatmaps are colored by gene expression (purple = low, yellow = high).

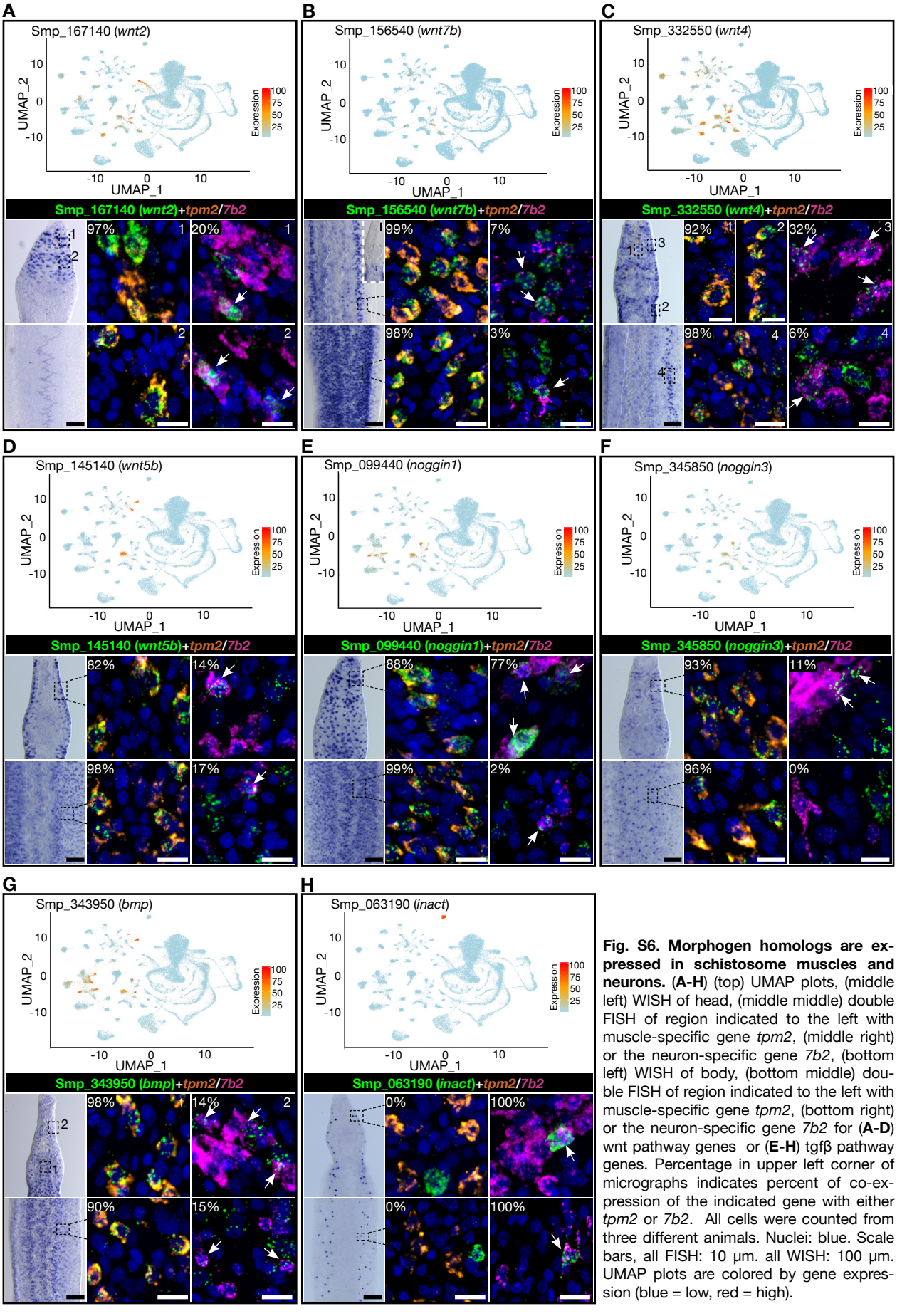


Fig. S6. Morphogen homologs are expressed in schistosome muscles and neurons. (**A-H**) (top) UMAP plots, (middle left) WISH of head, (middle middle) double FISH of region indicated to the left with muscle-specific gene *tpm2*, (middle right) or the neuron-specific gene *7b2*, (bottom left) WISH of body, (bottom middle) double FISH of region indicated to the left with muscle-specific gene *tpm2*, (bottom right) or the neuron-specific gene *7b2* for (**A-D**) wnt pathway genes or (**E-H**) tgfβ pathway genes. Percentage in upper left corner of micrographs indicates percent of co-expression of the indicated gene with either *tpm2* or *7b2*. All cells were counted from three different animals. Nuclei: blue. Scale bars, all FISH: 10 μm. all WISH: 100 μm. UMAP plots are colored by gene expression (blue = low, red = high).

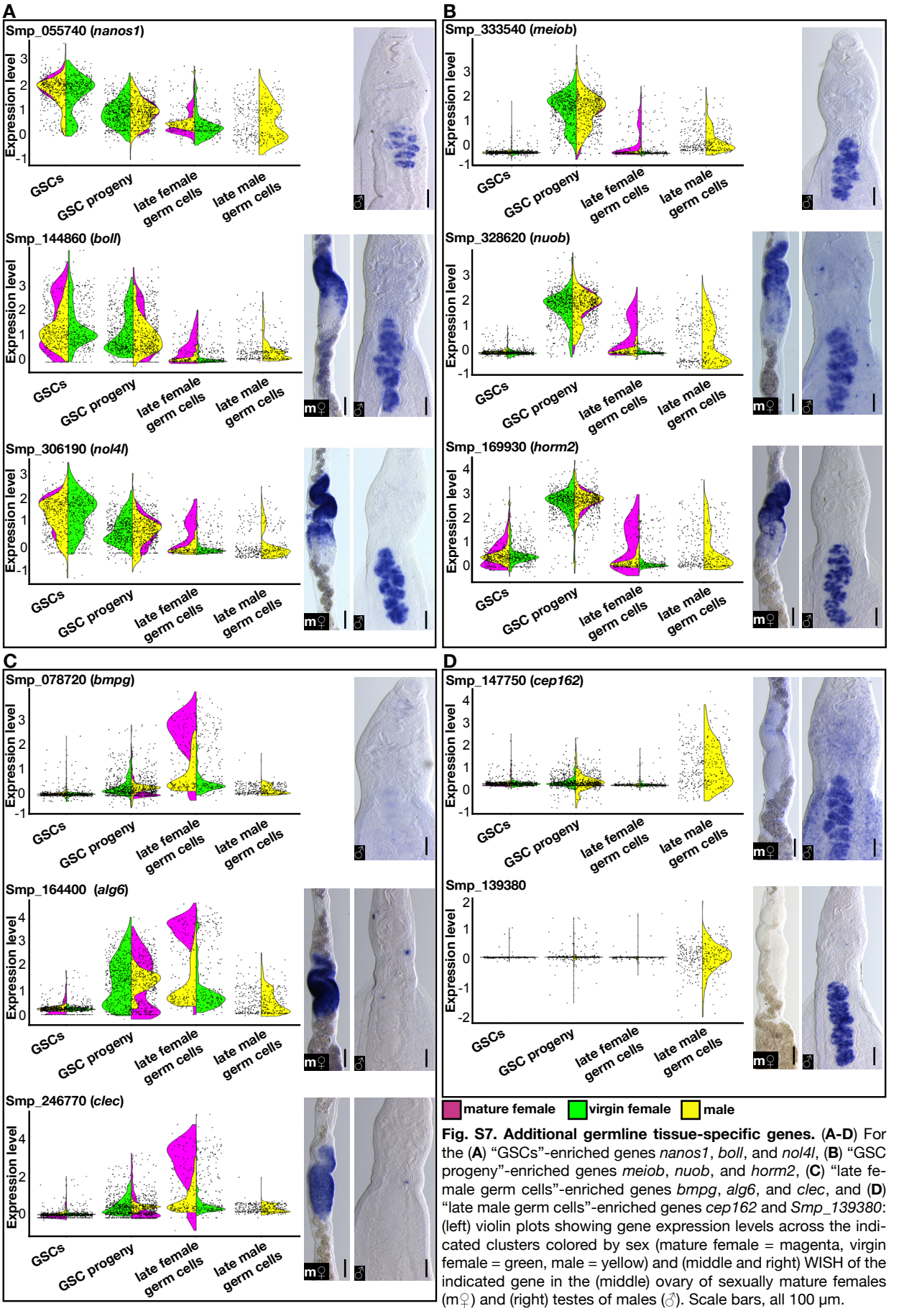


Fig. S7. Additional germline tissue-specific genes. (**A-D**) For the (**A**) "GSCs"-enriched genes *nanos1*, *boll*, and *nol4l*, (**B**) "GSC progeny"-enriched genes *meiob*, *nuob*, and *horm2*, (**C**) "late female germ cells"-enriched genes *bmpg*, *alg6*, and *clec*, and (**D**) "late male germ cells"-enriched genes *cep162* and *Smp_139380*: (left) violin plots showing gene expression levels across the indicated clusters colored by sex (mature female = magenta, virgin female = green, male = yellow) and (middle and right) WISH of the indicated gene in the (middle) ovary of sexually mature females (m♀) and (right) testes of males (♂). Scale bars, all 100 µm.

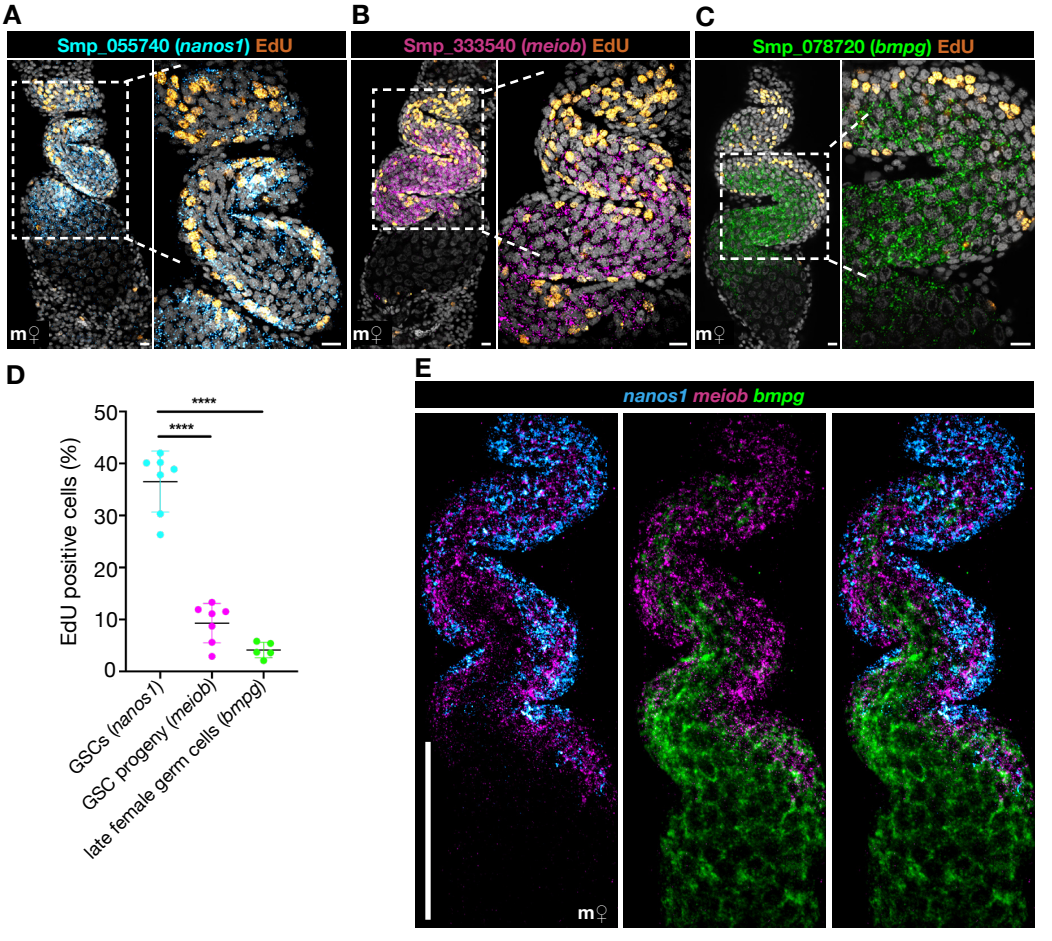


Fig. S8. Description of germ lineage in schistosome ovary. (**A-C**) FISH of the GSC marker *nanos1* (cyan) (**A**), the "GSC progeny"-enriched gene *meiob* (magenta) (**B**), or the "late female germ cells"-enriched gene *bmpg* (green) (**C**) in conjunction with a 30-minute EdU pulse (orange) to label the actively proliferating cells of the ovary of a sexually mature female (m♀). Nuclei: grey. (**D**) Graph showing quantification of percentage of *nanos1*$^+$, *meiob*$^+$, or *bmpg*$^+$ cells that are EdU$^+$ following a 30-minute EdU pulse. (**E**) Triple FISH of *nanos1*, *meiob* and *bmpg* in the ovary of a sexually mature female (m♀). Scale bars, all 100 μm.****$p < 0.0001$ (one-way ANOVA test).

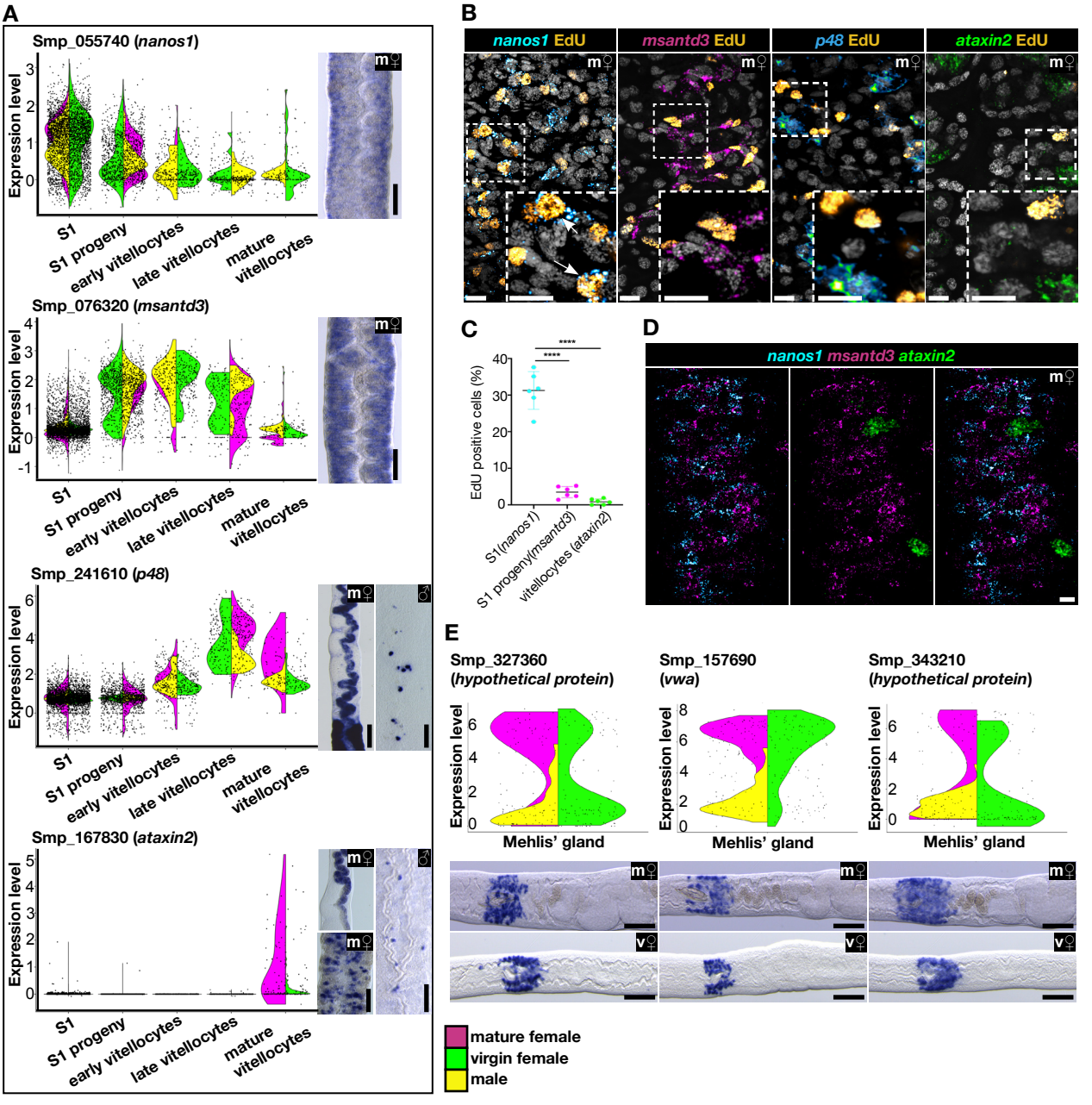


Fig. S9. Description of germ lineage in schistosome vitellaria. (**A-B**) For the "S1" marker *nanos1*, the "S1 progeny" marker *msantd3*, the "late vitellocytes" marker *p48*, and the "mature vitellocytes" marker *ataxin2*: (**A**) (left) violin plots showing gene expression levels across the indicated clusters colored by sex (mature female = magenta, virgin female = green, male = yellow) and (right) WISH of the indicated gene in the vitellaria of mature females (m♀) and the midline of males (♂) as indicated on the image or (**B**) FISH for the indicated gene with EdU labeling of proliferative cells (orange) in the vitellaria of a sexually mature female. Nuclei: grey. (**C**) Graph showing quantification of percentage of *nanos1*$^+$, *msantd3*$^+$, or *ataxin2*$^+$ cells that are EdU$^+$ following a 30-minute EdU pulse. (**D**) Representative micrograph of triple FISH of *nanos1*, *msantd3* and *ataxin2* in the vitellaria of a sexually mature female. (**E**) For the "Mehlis' gland" marker *Smp_327360*, *vwa*, and *Smp_343210*: (top) violin plots showing gene expression levels in the "Mehlis' gland" cluster colored by sex (mature female = magenta, virgin female = green, male = yellow) and (bottom) WISH of the indicated gene in region anterior to the ovary in sexually mature females (m♀) and virgin females (v♀) as indicated on the image. Scale bars, A, E, 100 µm; B, D, 10 µm. ****$p < 0.0001$ (one-way ANOVA test).

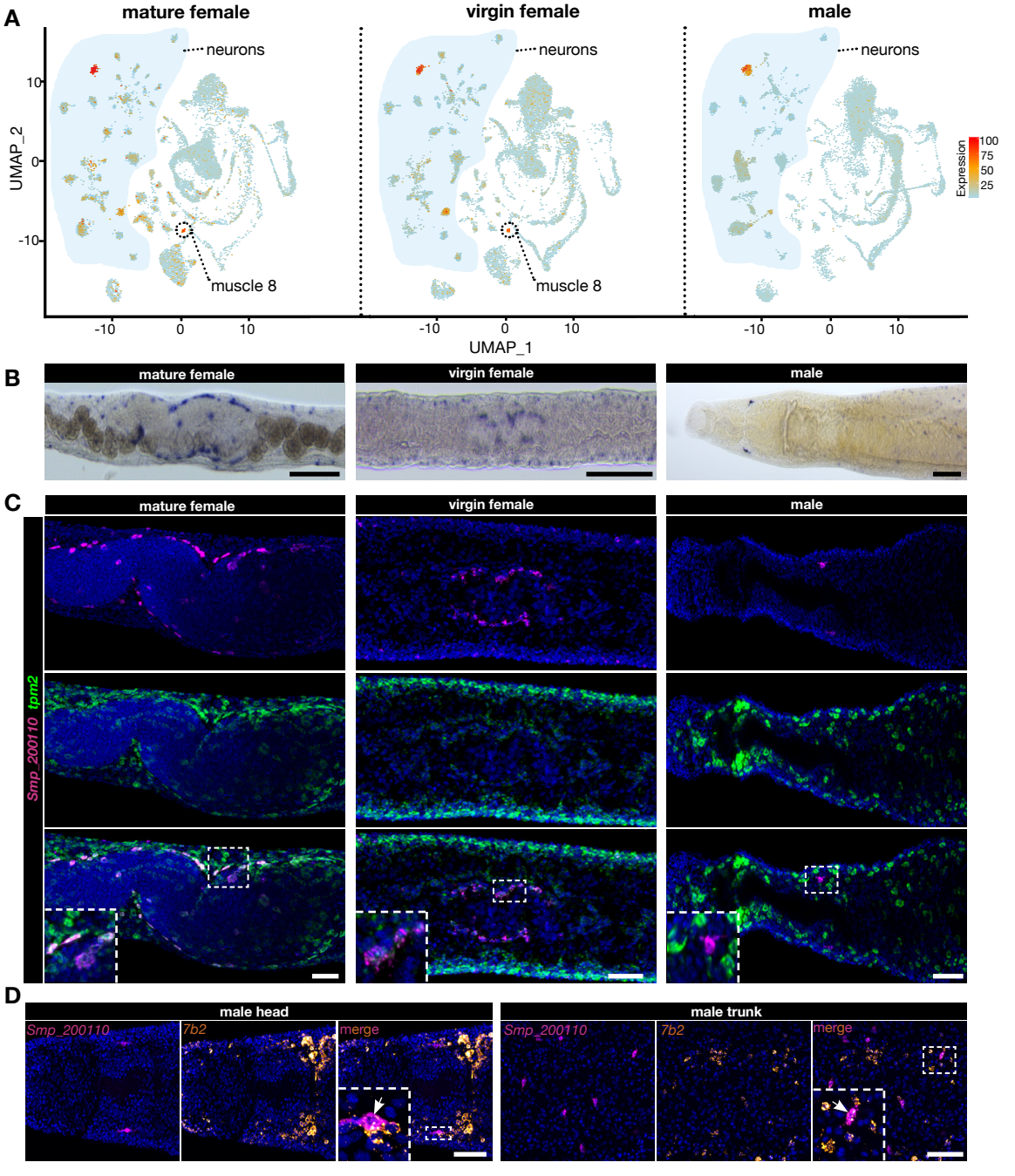


Fig. S10. Description of a female-specific muscle type. (**A**) UMAP plot showing expression of the "muscle 8" marker *Smp_200110* in mature female (left), virgin female (middle) but not near male gonads (right). (**B**) WISH of *Smp_200110* in mature female (left), virgin female (middle) and male (right). (**C**) Double FISH of *Smp_200110* (magenta) and the general muscle marker *tpm2* (green). Left and middle insets highlight co-expression of *Smp_200110* and *tpm2* around the ovary in mature and virgin female parasites. Right inset highlights lack of co-expression of *Smp_200110* and *tpm2* in male parasites. (**D**) Double FISH of *Smp_200110* (magenta) and general neuronal marker *7b2* (orange) in the (left) head and (right) trunk of male parasites. Insets show co-expression of *Smp_200110* and *7b2* through the male. UMAP plots are colored by gene expression (blue = low, red = high). Scale bars, B, 100 µm; C, D, 50 µm.

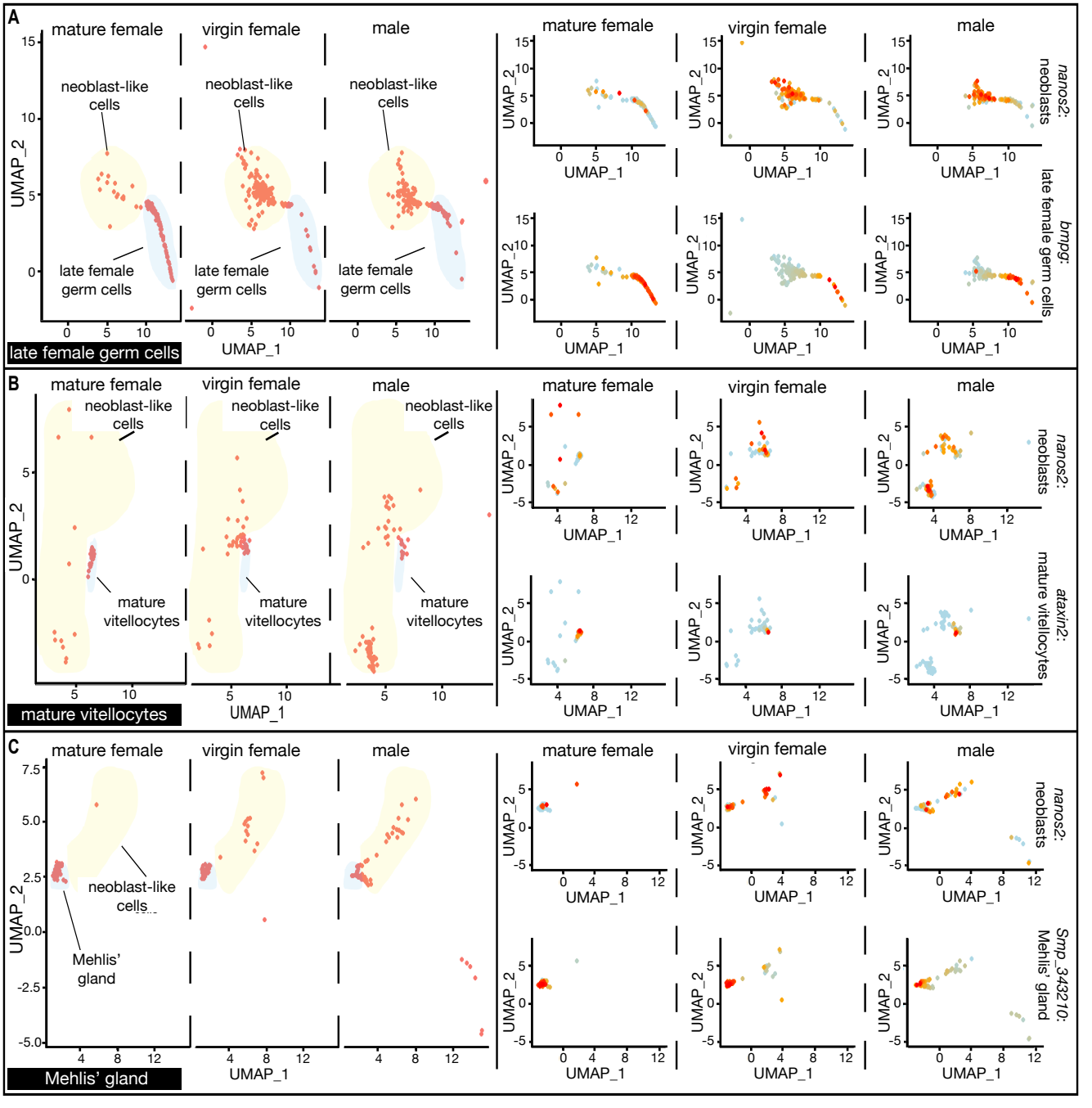


Fig. S11. Male cells in female reproductive tissues. (**A-C**) (Left) UMAP plot showing all cells in the indicated cluster split by parasite sex and maturity. (top right) UMAP plot showing the expression of the neoblast marker *nanos2* in the indicated cluster split by parasite sex and maturity and (bottom right) UMAP plot showing the expression of the cluster marker in the indicated cluster split by parasite sex and maturity for (**A**) "late female germ cells", (**B**) "mature vitellocytes" and (**C**) "Mehlis' gland". UMAP plots are colored by gene expression (blue = low, red = high).

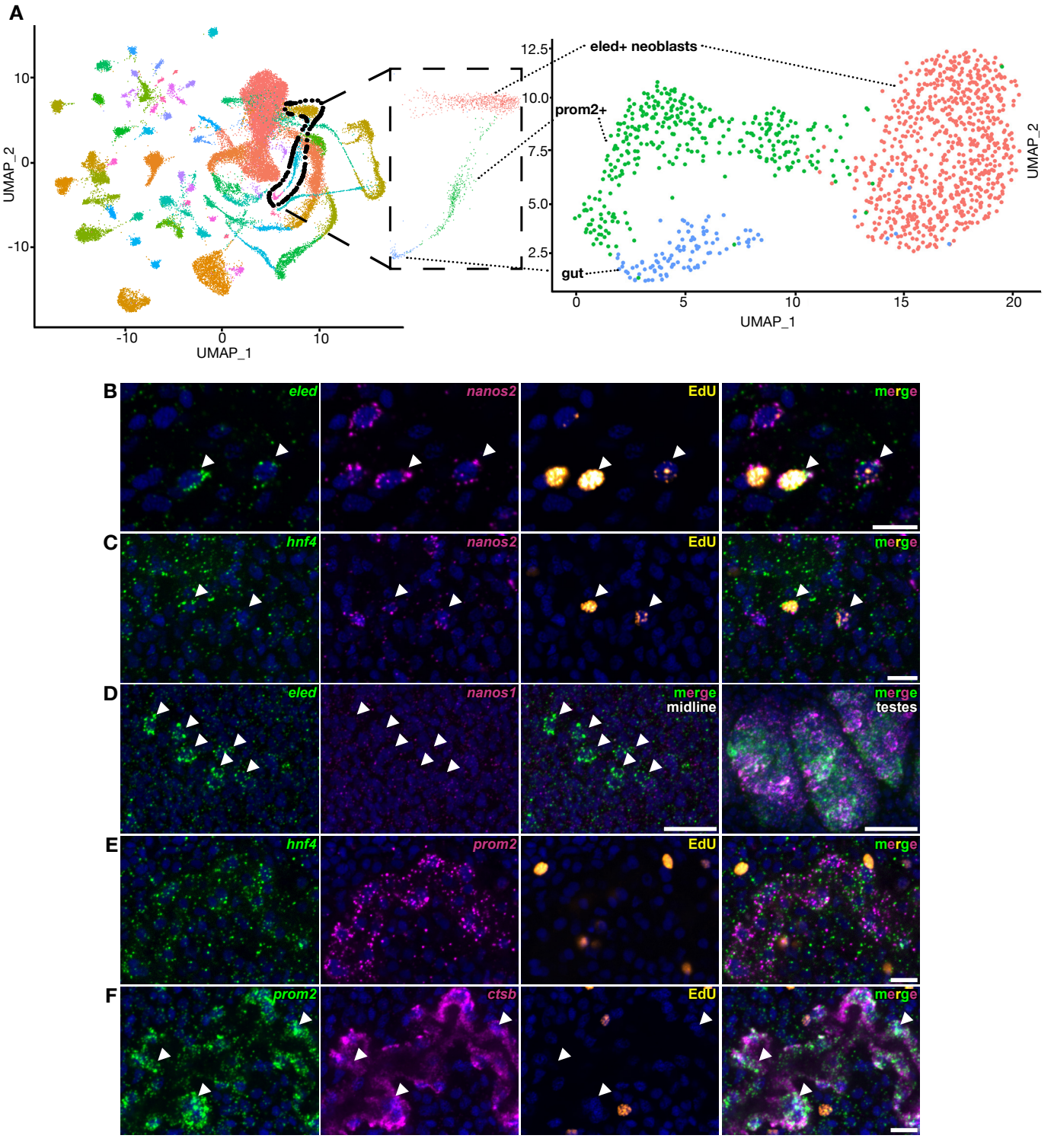


Fig. S12. A putative schistosome gut lineage. (**A**) Schematic of the re-clustering of the putative gut lineage from the single cell RNAseq data. (**B**) Double FISH of *eled* and the neoblast marker *nanos2* in EdU⁺ proliferative cells. Arrows indicate triple-positive cells. (**C**) Double FISH of *hnf4* and *nanos2* in EdU⁺ proliferative cells. Arrows indicate triple-positive cells. (**D**) Double FISH of *eled* and the GSC marker *nanos1*. Arrows indicate single-positive *eled*⁺ cells along the midline. (**E**) Double FISH of *hnf4* and *prom2*. (**F**) Double FISH of *prom2* with the gut marker *ctsb*. Arrows indicate foci of high *prom2* expression and low *ctsb* expression. Nuclei: blue. Scale bars, 10 μm.

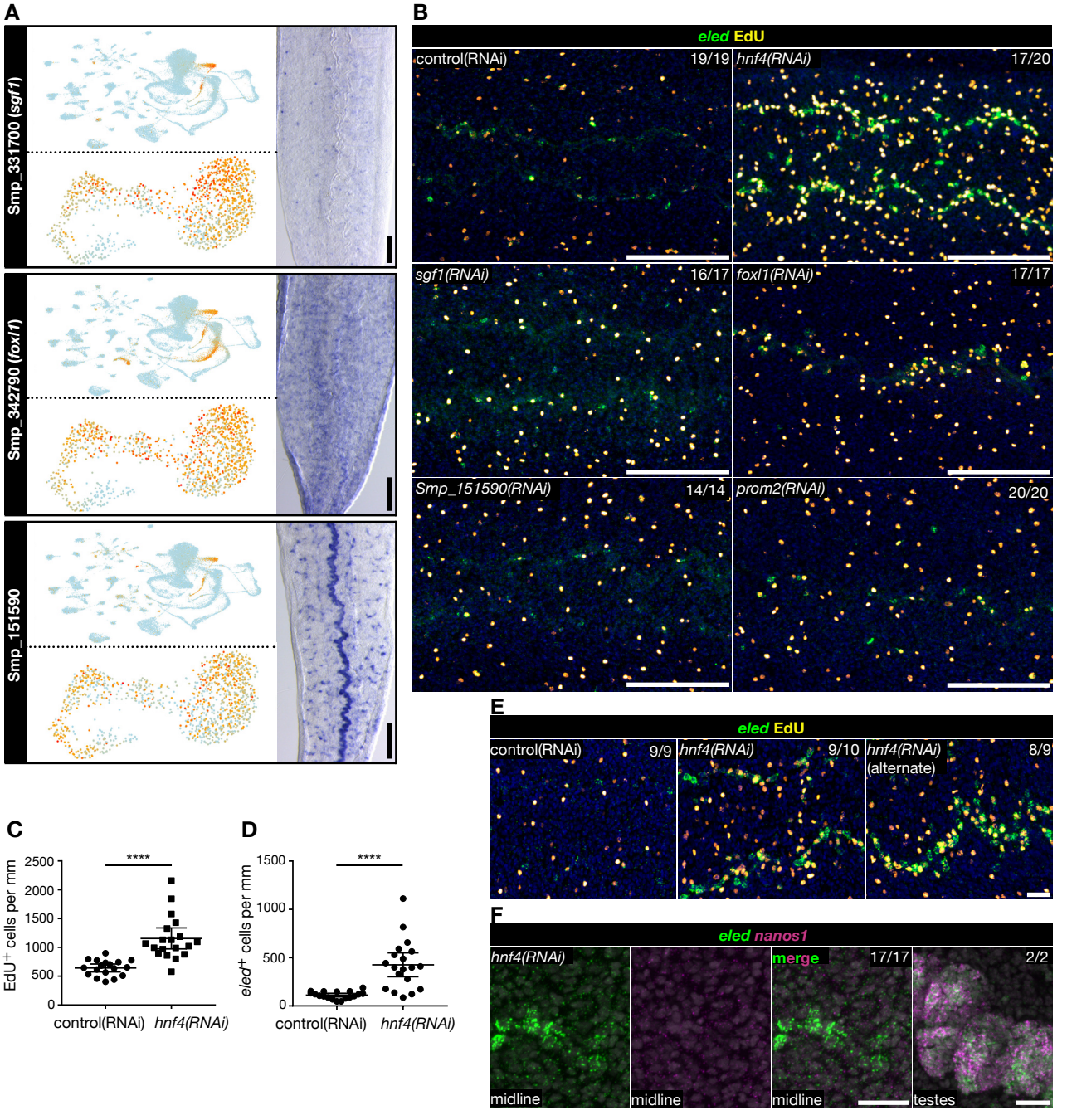


Fig. S13. An RNAi screen identifies *hnf4* as a regulator of *eled*⁺ neoblasts. (**A**) For each of the "*eled*⁺ neoblast"-enriched genes *sgf1*, *foxl1*, and *Smp_151590*: UMAP plots of the expression pattern of the indicated gene on the (top) original and (bottom) re-clustered dataset from Fig. S12A, and (right) WISH of the indicated gene. (**B**) FISH of *eled* with EdU pulse detection showing the location of *eled*⁺ neoblasts (green) and EdU⁺ proliferative cells (yellow) in the indicated RNAi condition. RNAi target gene name is indicated in the upper left. $n \geq 14$ parasites, two biological replicates. (**C** and **D**) Graph showing quantification of the number of EdU⁺ proliferative cells (**C**) or *eled*⁺ cells (**D**) per mm of parasite from Fig. 3B in either control(RNAi) or *hnf4(RNAi)* animals. $n \geq 18$ parasites, two biological replicates. (**E**) FISH of *eled* with EdU pulse detection showing the location of *eled*⁺ neoblasts (green) and EdU⁺ proliferative cells (yellow) in either control RNAi conditions ("control RNAi"), *hnf4* RNAi conditions ("*hnf4(RNAi)*"), or *hnf4* RNAi conditions using a separate, non-overlapping construct ("*hnf4(RNAi)* (alternate)"). $n \geq 9$ parasites, one biological replicate. (**F**) Double FISH of *eled* and *nanos1* demonstrating no co-expression along the parasite's midline but strong co-expression of *eled* and *nanos1* in reproductive organs like the testes in *hnf4* RNAi conditions. $n = 17$ *hnf4(RNAi)* animals from two biological replicates. The number of parasites similar to the representative images is indicated in the upper-right of each panel. Nuclei: blue (A, B) or grey (F). Scale bars, A, B, 100 μm; E, F, 20 μm. UMAP plots are colored by gene expression (blue = low, red = high). ****$p < 0.0001$ (Welch's t-test).

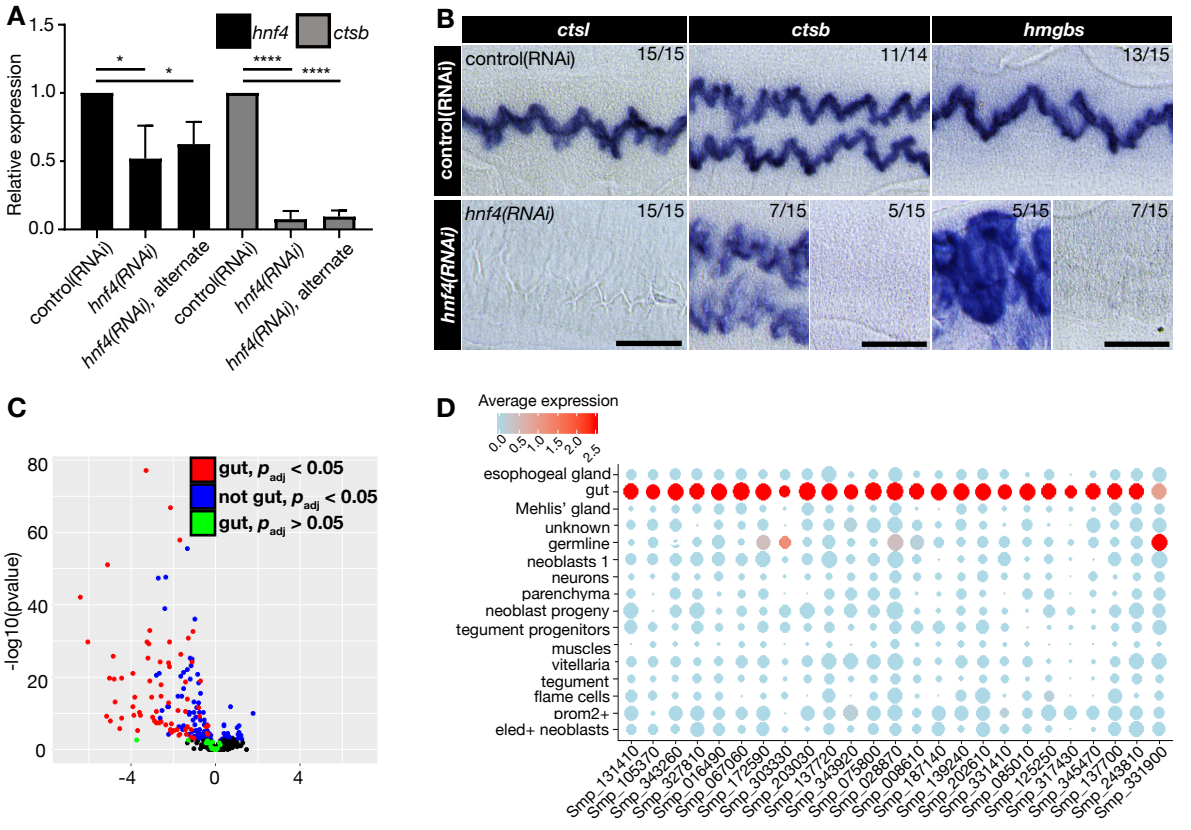


Fig. S14. *hnf4* RNAi results in transcriptional gut abnormalities. (**A**) Graph of relative quantification of *hnf4* mRNA (black) or *ctsb* mRNA (grey) as determined by qPCR in either "control(RNAi)", "*hnf4(RNAi)*", or "*hnf4(RNAi)* alternate" animals. Four biological replicates. (**B**) For the "gut"-specific genes *ctsl, ctsb* and *hmgbs*: WISH of the indicated gene in either control RNAi conditions or *hnf4* RNAi conditions. The number of parasites similar to representative images is indicated in the upper right of each panel. $n \geq 14$ parasites, three biological replicates. (**C**) Volcano plot of data from an RNAseq experiment comparing gene expression of control(RNAi) animals to that of *hnf4(RNAi)* animals. "gut", genes expressed in the "gut" cluster, "not gut", genes not expressed in the "gut" cluster. Significance determined as $p_{adj} < 0.05$ by Benjamini and Hochberg-corrected Wald test. (**D**) A dot-plot summarizing the cluster-specific expression of each of the top 25 down-regulated genes in *hnf4(RNAi)* animals. Cluster IDs are on the vertical axis and gene IDs are on the horizontal axis. Expression levels are colored by gene expression (blue = low, red = high). Percentage of cells in the cluster expressing the gene is indicated by the size of the circle (small = few, large = many). Scale bars, 100 µm. *$p < 0.05$, ****$p < 0.0001$ (Welch's t-test).

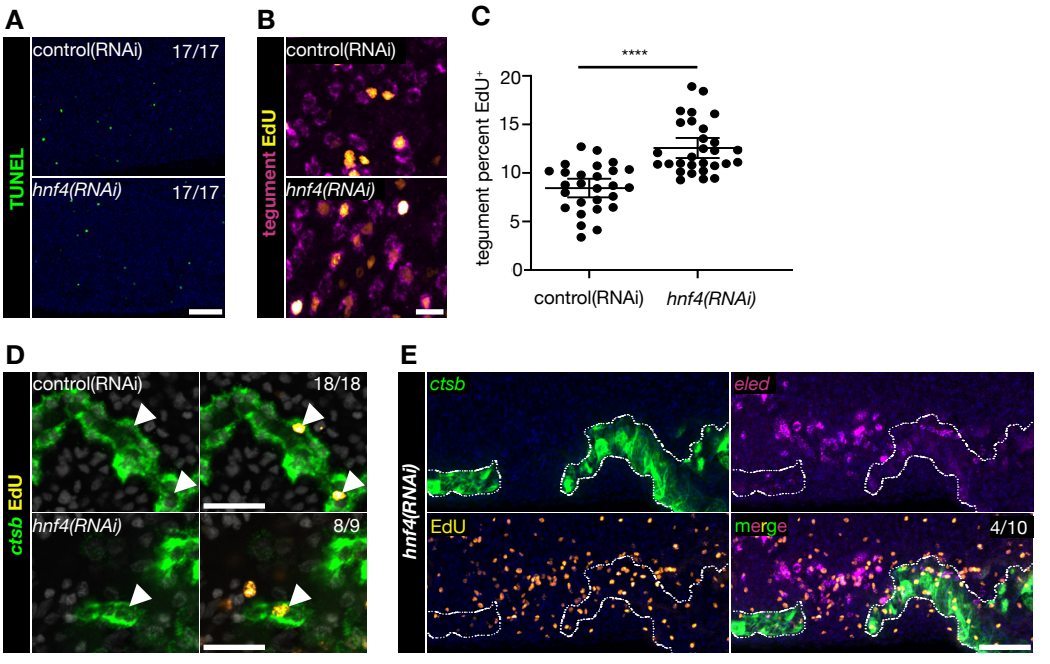


Fig. S15. *hnf4* RNAi results in gut neoblast abnormalities. (**A**) Fluorescent TUNEL experiment showing apoptotic cells (green) in either control RNAi or *hnf4* RNAi conditions. n = 17 parasites, two biological replicates. (**B**) FISH with a pooled mix of four tegument-specific mRNAs (magenta) with detection of EdU after a 7-day pulse-chase showing the location of EdU$^+$ progeny cells (yellow) in either control RNAi or *hnf4* RNAi conditions. (**C**) Quantification of the percentage of tegument cells that are EdU$^+$ from (B). $n \geq 27$, three biological replicates. (**D**) FISH of the gut marker *ctsb* (green) with detection of EdU after a 7-day pulse-chase showing location of EdU$^+$ progeny cells (yellow) in either control RNAi or *hnf4* RNAi conditions. $n \geq 9$, two biological replicates. (**E**) Double FISH of the gut marker *ctsb* and *eled* with an EdU pulse showing the location of EdU$^+$ proliferative cells (yellow) in *hnf4* RNAi conditions. The dashed line indicates the approximate boundary of the residual gut-like tissue found in *hnf4(RNAi)* animals. The number of parasites similar to the representative image is indicated in the upper right of each panel. Nuclei: blue (A, B, E) or grey (D). Scale bars, A, E, 50 µm; B, 10 µm; D, 20 µm. ****$p < 0.0001$ (Welch's t-test).

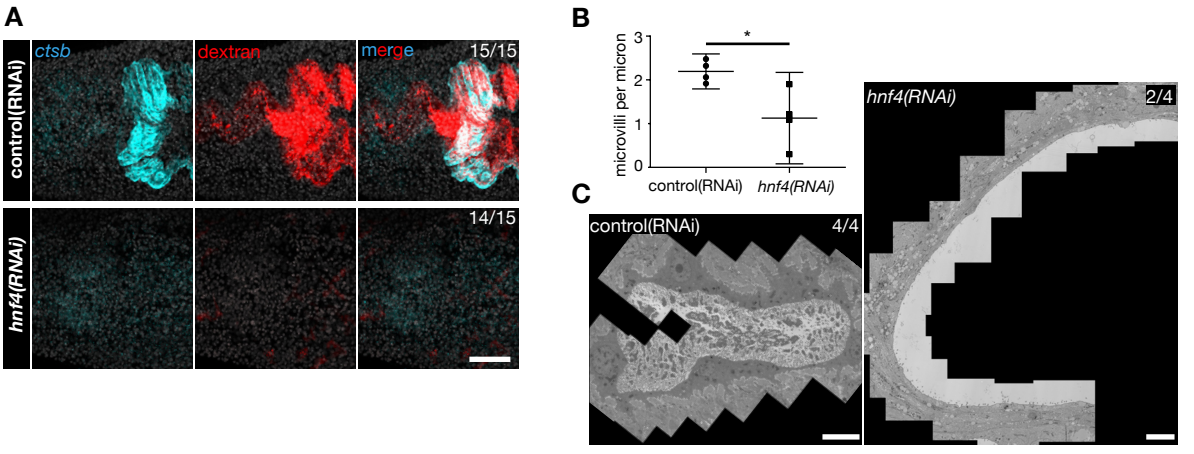


Fig. S16. *hnf4* RNAi results in structural abnormalities in the gut. (**A**) Double FISH of the gut marker *ctsb* (cyan) and fluorescently-labeled dextran (red) in the gut lumen in either control RNAi or *hnf4* RNAi conditions. *n* =15 animals, three biological replicates. (**B**) Graph showing quantification of the number of microvilli per micron of gut surface from Fig. 3C. Numbers are the average of four different sections of gut from each of four animals. (**C**) Stitched TEM micrographs from either control(RNAi) animals, or *hnf4(RNAi)* animals. *n* = 4 animals, two biological replicates. The number of parasites similar to representative images is indicated in the upper right of each panel. Nuclei: grey. Scale bars, A, 20 µm; C, 5 µm. $^{*}p < 0.05$ (Welch's t-test).

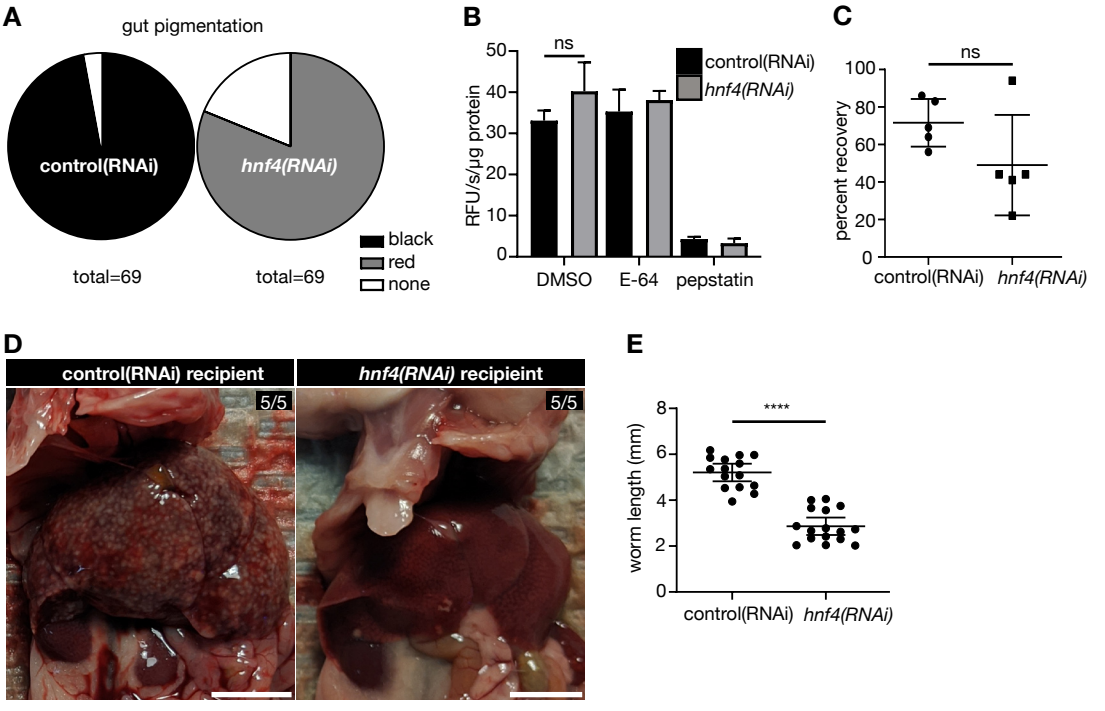


Fig. S17. ***hnf4*** **is required for blood feeding.** (**A**) Pie chart depicting the frequency of different gut pigmentation of animals from Fig. 4A. $n = 69$ animals, three biological replicates. (**B**) Graph of the aspartyl protease activity of lysates from control(RNAi) or *hnf4(RNAi)* parasites as determined by the ability to cleave the fluorogenic substrate, mca-GK-PILFFRLK-K(dnp) in the presence of no inhibitor (DMSO), the general cysteine protease inhibitor E-64 (E-64), or the aspartyl protease inhibitor pepstatin A (pepstatin). (**C**) Graph quantifying the recovery rate of worms from transplant recipients. $n = 5$ recipients, two biological replicates. (**D**) Images of livers of mice 30 days after transplant with RNAi-treated parasites. $n = 5$ recipients, two biological replicates. (**E**) Graph showing quantification of male worm length from Fig. 4D. $n \geq 15$ animals, three separate recipients. The number of livers similar to the representative image is indicated in the upper right each panel. Scale bars, 1 cm. ns, not significant, $^{****}p < 0.0001$ (Welch's t-test).

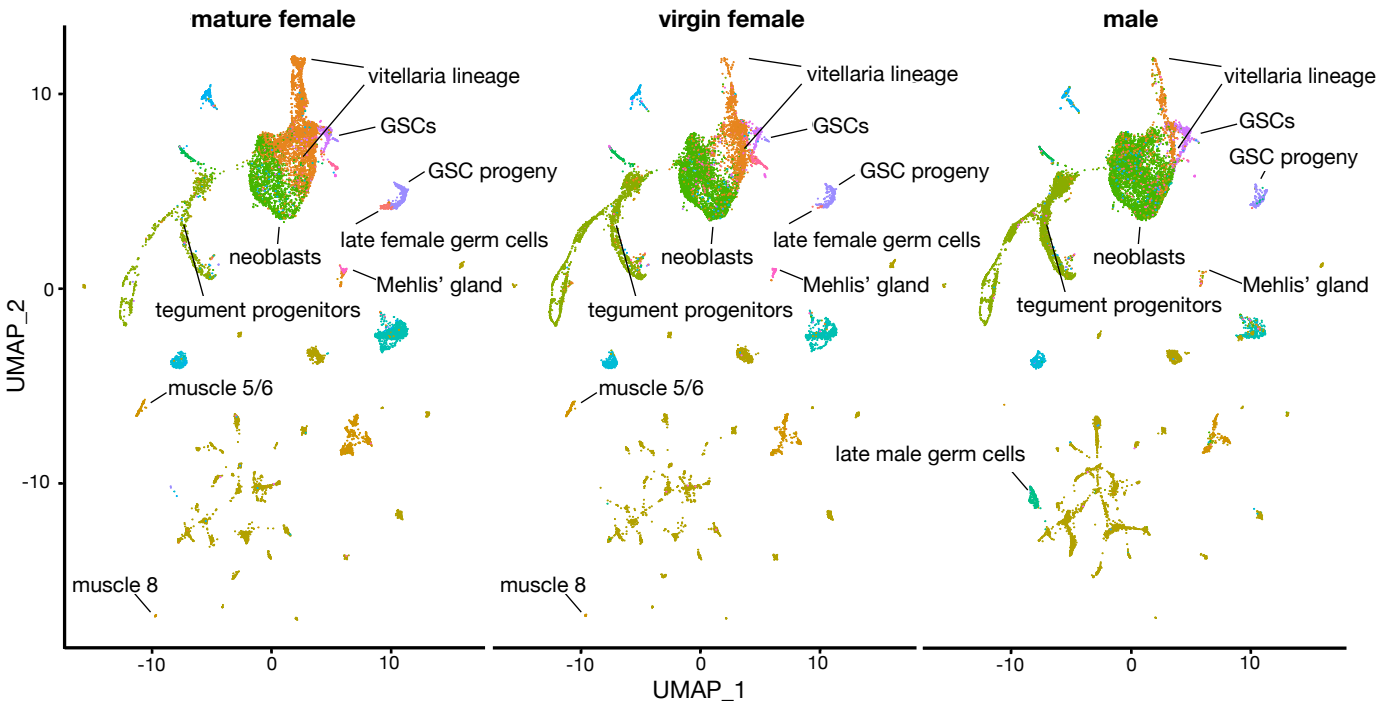


Fig. S18. Different normalization methods produce same results. UMAP plot showing the scRNAseq data reclustered using sctransform instead of the standard Seurat workflow. Relevant clusters are labeled.

Captions for Tables S1-S10 (Tables are in separate excel files).

Table S1.
List of genes significantly enriched in indicated cluster (see Fig. S1 for clusters)

Table S2.
Number of cells per cluster per sample from scRNAseq data

Table S3.
GO terms enriched in clusters from scRNAseq data

Table S4.
Differentially expressed genes following hnf4 RNAi

Table S5.
Information regarding sample material for single-cell RNAseq library generation

Table S6.
Strategies for performing RNAi experiments

Table S7.
List of oligos, gene abbreviations, and markers described in this study