



## OPEN ACCESS

## EDITED BY

Yadong Zheng,  
Zhejiang Agriculture and Forestry  
University, China

## REVIEWED BY

Pablo Smircich,  
Instituto de Investigaciones Biológicas  
Clemente Estable (IIBCE), Uruguay  
Dong-Hui Zhou,  
Fujian Agriculture and Forestry  
University, China

## \*CORRESPONDENCE

Sergio Verjovski-Almeida,  
verjo@iq.usp.br

<sup>†</sup>These authors have contributed equally  
to this work and share first authorship

## SPECIALTY SECTION

This article was submitted to RNA,  
a section of the journal  
Frontiers in Genetics

RECEIVED 20 April 2022

ACCEPTED 24 August 2022

PUBLISHED 20 September 2022

## CITATION

Morales-Vicente DA, Zhao L,  
Silveira GO, Tahira AC, Amaral MS,  
Collins JJ III and Verjovski-Almeida S  
(2022), Single-cell RNA-seq analyses  
show that long non-coding RNAs are  
conspicuously expressed in  
*Schistosoma mansoni* gamete and  
tegument progenitor cell populations.  
*Front. Genet.* 13:924877.  
doi: 10.3389/fgene.2022.924877

## COPYRIGHT

© 2022 Morales-Vicente, Zhao, Silveira,  
Tahira, Amaral, Collins and Verjovski-  
Almeida. This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Single-cell RNA-seq analyses show that long non-coding RNAs are conspicuously expressed in *Schistosoma mansoni* gamete and tegument progenitor cell populations

David A. Morales-Vicente<sup>1,2†</sup>, Lu Zhao<sup>3†</sup>, Gilbert O. Silveira<sup>1,2†</sup>, Ana C. Tahira<sup>1</sup>, Murilo S. Amaral<sup>1</sup>, James J. Collins III<sup>3</sup> and Sergio Verjovski-Almeida<sup>1,2\*</sup>

<sup>1</sup>Laboratório de Parasitologia, Instituto Butantan, São Paulo, Brazil, <sup>2</sup>Instituto de Química, Universidade de São Paulo, São Paulo, Brazil, <sup>3</sup>Department of Pharmacology, UT Southwestern Medical Center, Dallas, TX, United States

*Schistosoma mansoni* is a flatworm that causes schistosomiasis, a neglected tropical disease that affects over 200 million people worldwide. New therapeutic targets are needed with only one drug available for treatment and no vaccine. Long non-coding RNAs (lncRNAs) are transcripts longer than 200 nucleotides with low or no protein-coding potential. In other organisms, they have been shown as involved with reproduction, stem cell maintenance and drug resistance, and they tend to exhibit tissue-specific expression patterns. *S. mansoni* expresses thousands of lncRNA genes; however, the cell type expression patterns of lncRNAs in the parasite remain uncharacterized. Here, we have re-analyzed publicly available single-cell RNA-sequencing (scRNA-seq) data obtained from adult *S. mansoni* to identify the lncRNAs signature of adult schistosome cell types. A total of 8023 lncRNAs (79% of all lncRNAs) were detected. Analyses of the lncRNAs expression profiles in the cells using statistically stringent criteria were performed to identify 74 lncRNA gene markers of cell clusters. Male gamete and tegument progenitor lineages clusters contained most of the cluster-specific lncRNA markers. We also identified lncRNA markers of specific neural clusters. Whole-mount *in situ* hybridization (WISH) and double fluorescence *in situ* hybridization were used to validate the cluster-specific expression of 13 out of 16 selected lncRNA genes (81%) in the male and female adult parasite tissues; for one of these 16 gene loci, probes for two different lncRNA isoforms were used, which showed differential isoform expression in testis and ovary. An atlas of the expression profiles across the cell clusters of all lncRNAs detected in our analysis is available as a public website resource (<http://verjolah.usp.br:8081>). The results presented here give strong support to a tissue-specific expression and to a regulated expression program of lncRNAs in *S. mansoni*. This will be the basis for further exploration of lncRNA genes as potential therapeutic targets.

## KEYWORDS

parasitology, RNA-seq, RNA sequencing, single-cell sequencing data analysis, *Schistosoma mansoni* adult worms, long non-coding RNAs, single-cell expression profiles

## 1 Introduction

Schistosomiasis is a neglected tropical disease that affects more than 200 million people worldwide (Colley et al., 2014; Mcmanus et al., 2018). Controlling the disease is still a challenge, as no vaccine is currently available (Tebeje et al., 2016; Molehin, 2020). In addition, treatment is restricted to a single drug, praziquantel, which does not act on juvenile worms and against which there are reports of parasite tolerance (Bergquist et al., 2017a; Vale et al., 2017; Kittur et al., 2019). Therefore, the search for new therapeutic targets is needed and understanding the schistosome's biology on a molecular level could suggest new therapeutic alternatives (Bergquist et al., 2017b).

Single-cell RNA sequencing (scRNA-seq) has been used to advance knowledge of the schistosome's biology through the identification of specific protein-coding molecular markers of different tissue types in *Schistosoma mansoni* sporocysts (Wang et al., 2018), schistosomula (Diaz Soria et al., 2020) and juvenile/adult worms (Tarashansky et al., 2019; Wendt et al., 2020; Li et al., 2021). Importantly, these works have provided comprehensive protein-coding gene expression cell type atlases at different stages of parasite development. However, the spatial distribution of long non-coding RNAs (lncRNAs) across tissues and cell types has not been assessed yet in *Schistosoma*, even though it is well known that lncRNAs can define cell clusters in other multicellular organisms (Liu et al., 2016; Zhou et al., 2019).

lncRNAs are RNAs longer than 200 nucleotides with low or no protein-coding potential that have been implicated in different biological processes (Ransohoff et al., 2018). They are responsible for a wide range of functions, including regulation of protein-coding gene expression (Jandura and Krause, 2017; Rinn and Chang, 2020) and stem cell maintenance (Chen et al., 2020). Because of their versatile functions and tissue-specific expression, lncRNAs have been proposed as pharmacological targets, especially in human neurodegenerative disorders and cancers (Jiang et al., 2019; Nath et al., 2019).

In *S. mansoni*, we have published a catalogue of thousands of lncRNAs expressed in several stages of the parasite (Maciel et al., 2019), serving as the basis for further studies of these lncRNAs at different conditions. Recently, we have also shown that lncRNAs are potential new therapeutic targets in *S. mansoni* (Silveira et al., 2022). Here, we show for the first time the single-cell landscape of lncRNA distribution across adult *S. mansoni* cell types. We have re-analyzed public scRNA-seq data obtained from *S. mansoni* adult male and immature and mature adult female and identified

the lncRNAs signature of schistosome cell types. Analyses of the lncRNAs expression profiles in the cells have identified 74 lncRNA gene markers of cell clusters, many of which were validated with WISH. The results presented here give strong support to a tissue-specific expression and to a regulated expression program of lncRNAs in the parasite, which will be the basis for the exploration of lncRNA genes as potential therapeutic targets in the future.

## 2 Materials and methods

### 2.1 scRNA-seq processing

Single-cell raw fastq files from Wendt et al. (2020) SRA project PRJNA611777 were downloaded via fasterq-dump with the following arguments “-S -e 94 --include-technical”. The integrity of the raw fastq files was checked using vdb-validate, and all files were identified as consistent. To quantify the gene expression of the single-cell data set, we used STARsolo version 2.7.9a (Kaminow et al., 2021) along with a merged gene annotation file containing protein-coding genes, pseudogenes (*Schistosoma mansoni* WormBase gene annotation version 16 (Howe et al., 2017)) and lncRNA genes (Maciel et al., 2019) from a gtf file downloaded from <http://verjolab.usp.br/public/schMan/schMan3/macielEtAl2019/files/>, along with the genome assembly Smansoni\_v7 from WormBase (Howe et al., 2017) with the following parameters “--soloType CB\_UMI\_Simple --soloCellFilter EmptyDrops\_CR --soloFeatures Gene Velocityto GeneFull --soloMultiMappers EM --soloCBwhitelist barcodes\_whitelist”. For all samples except SRX7888067, we used the barcode whitelist from Cell Ranger chemistry V2; for sample SRX7888067, we used the barcode whitelist from chemistry V3. Filtered count matrices for all samples were imported into R (R\_Core\_Team, 2018) using Seurat v4.0.6.9900 (Hao et al., 2021) and cells were further removed from each matrix when the number of features was less than 500, number of counts less than 1000 and greater than 20,000, and percentage of mitochondrial genes greater than 3%. Matrices from all samples were normalized using the NormalizeData function, and variable features were identified using FindVariableFeatures with the following parameters “selection. Method = “vst”, nfeatures = 2000”. Additionally, we scaled the matrices and found principal components using the functions ScaleData, and RunPCA with the parameters “npcs = 100”. To generate the count matrix of all samples, we used the scRNA-seq integration approach from Seurat (Stuart et al., 2019). For that, we first identified integration features using the function

SelectIntegrationFeatures, then the integration anchors were identified using the function FindIntegrationAnchors with the following parameters “k.anchor = 20, dims = 1:78, anchor.features = features, reduction = ‘rpca’” and finally integrated the matrices using IntegrateData function. Then, the integrated matrix was scaled using the function ScaleData, and principal components were identified using the function runPCA with the following parameters “npcs = 100”. A final sparse matrix with 48,094 cells was obtained containing expression data for protein-coding genes, pseudogenes, and lncRNAs; and it was used for the following procedures.

## 2.2 Identification of lncRNA cell markers

To assign cell types to our new scRNA-seq data set, we projected the cell cluster annotation from Wendt et al. (2020) onto our re-analyzed scRNA-seq data set. For that, we retrieved the RDS object from the GEO project GSE146736 and imported it into R as a Seurat object using custom scripts. The Wendt et al. (2020) data set was used as the reference, and our new scRNA-seq data set was used as the query to identify cell anchors between both data sets with the function FindTransferAnchors with the following parameters “dims = 1:80, reference.reduction = ‘pca’”. Then we transferred the cell cluster annotation using the function TransferData with the following parameters “dims = 1:80”. Additionally, we transferred the uniform manifold approximation and projection (UMAP) plot embedding from the reference data set to our scRNA-seq data set. For that, we identified the first two UMAP embedding of the reference scRNA-seq data set with the function RunUMAP with the following parameters “return.model = TRUE, n.neighbors = 36, min.dist = 0.70”, then the embedding were transferred to our scRNA-seq data set using the function MapQuery with the following parameters “refdata = list (celltype = ‘cell\_types’), reference.reduction = ‘pca’, reduction.model = ‘umap’”.

After the cell annotation was transferred, we performed differential expression analysis among all clusters to identify cell type-specific markers. Normalization of read counts across different cells and different samples is particularly important when single-cell RNA-sequencing data is used for downstream analyses, such as differential expression testing, in which the results are confounded by cellular sequencing depth (Hafemeister and Satija, 2019). Moreover, because lncRNAs are known to be expressed at levels lower than those of protein-coding mRNAs, and because different groups of genes with different levels of expression cannot be normalized by the same constant factor (Hafemeister and Satija, 2019), at this step of the analysis we first performed scaled variance stabilization transformation (Hafemeister and Satija, 2019; Choudhary and Satija, 2022) in our scRNA-seq data set using the function SCTransform with the following parameters “method = ‘glmGamPoi’, vst.flavor = ‘v2’,

vars.to.regress = ‘percent.mt’”. Then, we set the transferred cell annotation as the active identity of the cells and ran the function FindAllMarkers with the following parameters “only.pos = TRUE, assay = ‘SCT’, min.pct = 0.25, logfc.threshold = 0.25, densify = TRUE, test.use = ‘bimod’”. To select the lncRNA markers, we considered as differentially expressed those genes with less than 0.05 corrected *p*-value in the Wilcox-test in each cluster, and removed differentially expressed genes with a median cluster expression of less than 1 SCT transformed counts compared to all cells of the data set using custom R scripts; this resulted in a final set of 74 lncRNAs identified as lncRNA markers, which were ranked by expression level within the cluster. The clusters where these 74 lncRNAs were identified as markers are shown with an UpSet intersection plot (Lex et al., 2014).

## 2.3 lncRNA markers selection for validation and primer design

To perform *in situ* hybridization experiments for lncRNA marker validation, sixteen lncRNAs were selected based on the clusters where they were identified as markers, on the existence of only one or a few transcript isoforms per gene in the locus, and on the ability to design a probe that only matched a single locus in the genome. To design primers that amplify sequences unique to each lncRNA, each lncRNA marker sequence was searched against the previously published *S. mansoni* transcriptome (Maciel et al., 2019) and only the lncRNA sequence segment that did not match any other transcript was selected for primer design and sequence amplification and cloning.

Information regarding the Gene\_ID, lncRNA Transcript\_ID and probe size for the 16 selected lncRNA markers is described at Supplementary Table S1. Notably, two different probes were designed for one lncRNA gene marker (G16045). One of the probes targets SmLINC129748, SmLNCA129749, SmLNCA129752, SmLNCA129753 and SmLNCA129758 transcript isoforms, while the other probe targets transcript isoforms SmLNCA129757 and SmLNCA129759. Pairs of primers to clone all 17 lncRNA marker probes were designed using PrimerQuest Tool provided by IDT Integrated DNA Technologies (<https://www.idtdna.com/PrimerQuest/>) and are shown in Supplementary Table S2. All cloned lncRNA marker sequences were confirmed with Sanger sequencing.

The sequences of interest were inserted into pJC53.2 (available from Addgene <https://www.addgene.org/26536/>) that had been previously digested with Eam1105I. The insert orientation was confirmed with Sanger sequencing using T3 or SP6 generical primers, and the *in situ* hybridization probes were synthesized accordingly, using T3 or SP6 RNA polymerase, as previously described (Collins et al., 2016; Wendt et al., 2020).

## 2.4 Whole *in situ* hybridization and imaging

Whole mount colorimetric and fluorescence *in situ* hybridization analyses were performed as previously described (Collins et al., 2016; Wendt et al., 2020). All lncRNA probes were used at 10 ng/ml in hybridization buffer, while probes of tissue/cell specific marker for double fluorescence were used at 50 ng/ml in hybridization buffer. All fluorescently labeled parasites were counterstained with DAPI (1 µg/ml) before being cleared in 80% glycerol, then mounted on slides with Vectashield (Vector Laboratories). Brightfield images were acquired on 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 Zeiss LSM900 Laser Scanning Confocal Microscope.

## 3 Results

### 3.1 lncRNAs identification in adult worm single-cells

To identify the lncRNAs signature of adult schistosome cell types we re-analyzed the publicly available single-cell RNA-sequencing (scRNA-seq) raw data obtained from adult *S. mansoni* by Wendt et al. (2020), as described in detail in the Methods. Briefly, scRNA-seq reads were mapped to the genome using a complete reference transcriptome, including 10,144 protein-coding (Smp) genes, 10,110 lncRNA genes, and 28 pseudogenes, and the numbers of reads mapped per gene locus (not per gene isoform) were counted. After normalization, a total of 17,429 genes were detected, of which 9388 SmPs (92.5% of all SmPs), 8023 lncRNAs (79.4% of all lncRNAs) and 18 pseudogenes. Our pipeline recovered 48,094 filtered cells, 10.2% more filtered cells than the 43,642 filtered cells recovered by Wendt et al. (2020). The mapping statistics including the number of reads mapped per sample and the number of cells recovered per sample are shown in Supplementary Table S3.

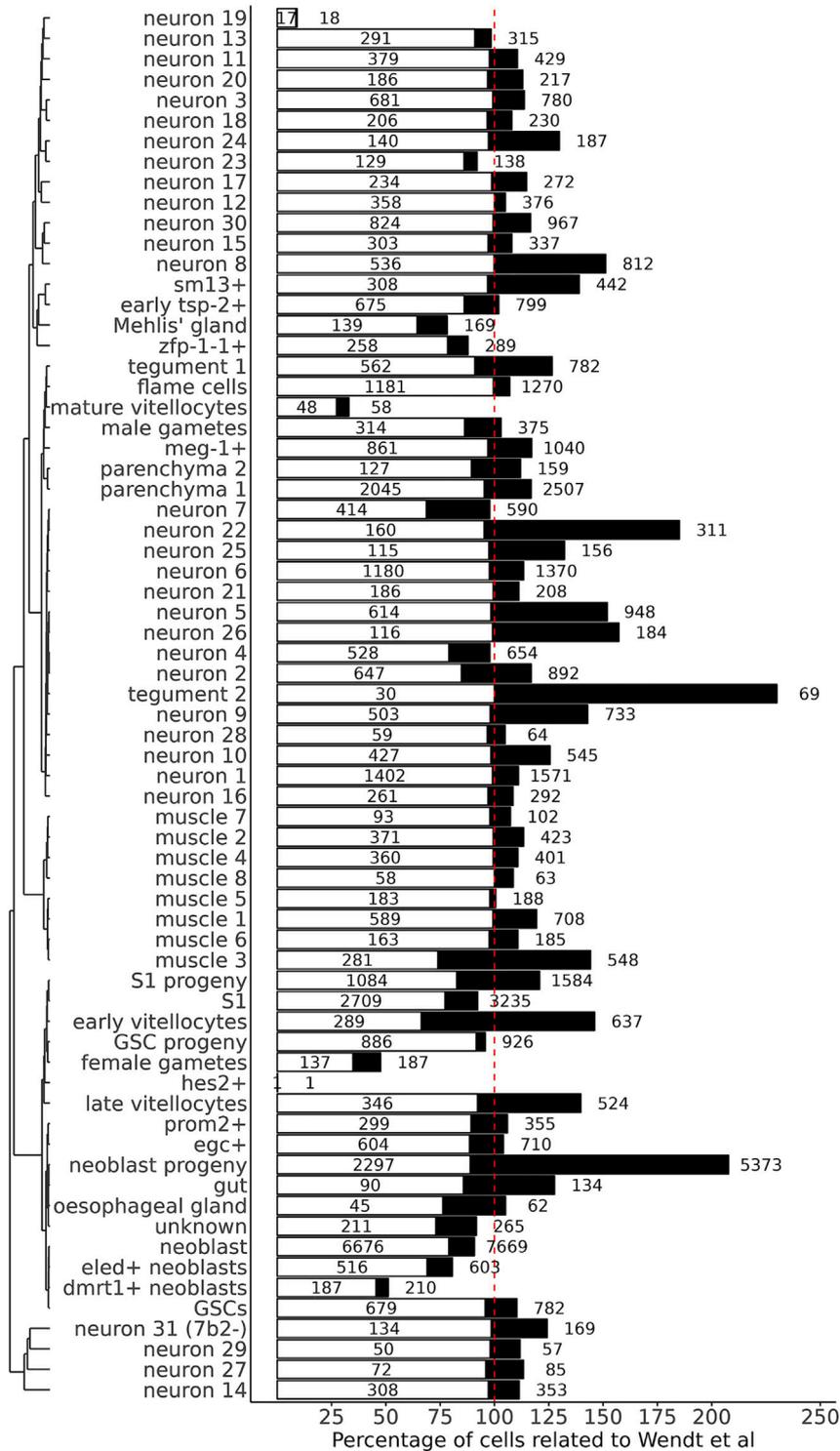
### 3.2 Transfer of cell cluster ID annotations

In the work of Wendt et al. (2020) the single-cells were grouped according to the expression profile of protein-coding genes into 68 different cell clusters, whose identities have been established by determining gene markers specifically expressed in each cluster. In addition, a thorough validation of the specific expression of a given marker at a given adult worm tissue had been obtained with whole mount *in situ* hybridization (WISH) and double fluorescence *in situ*

hybridization (dFISH) (Wendt et al., 2020). In order to transfer the cluster annotations to the re-analyzed single-cell set, we used the approach of Stuart et al. (2019) and queried the reference set of cells previously clustered by Wendt et al. (2020) with the newly obtained single-cell expression profile which included lncRNAs in addition to protein-coding genes. With this approach, correspondences between cells in the query and reference datasets can be identified, “anchors” can be used to harmonize datasets into a single reference, and reference labels and data can be projected onto the query dataset (Stuart et al., 2019). To give a visual sense of the cell cluster remapping efficiency we transferred the UMAP embedding from the reference data set to our reanalyzed scRNA-seq data set; Figure 1 shows the cells colored according to the clusters where they were remapped to, and the original cell cluster annotations are shown in light grey in the background. On a few clusters such as flame cells (at the bottom left), which gained approximately 7.5% more cells, or neuron 1, 6 and 30 (at the bottom center) which gained 10–12% more cells, the remapped cells (colored) have clustered more densely than in the original reference data set, leaving some light grey areas visible.

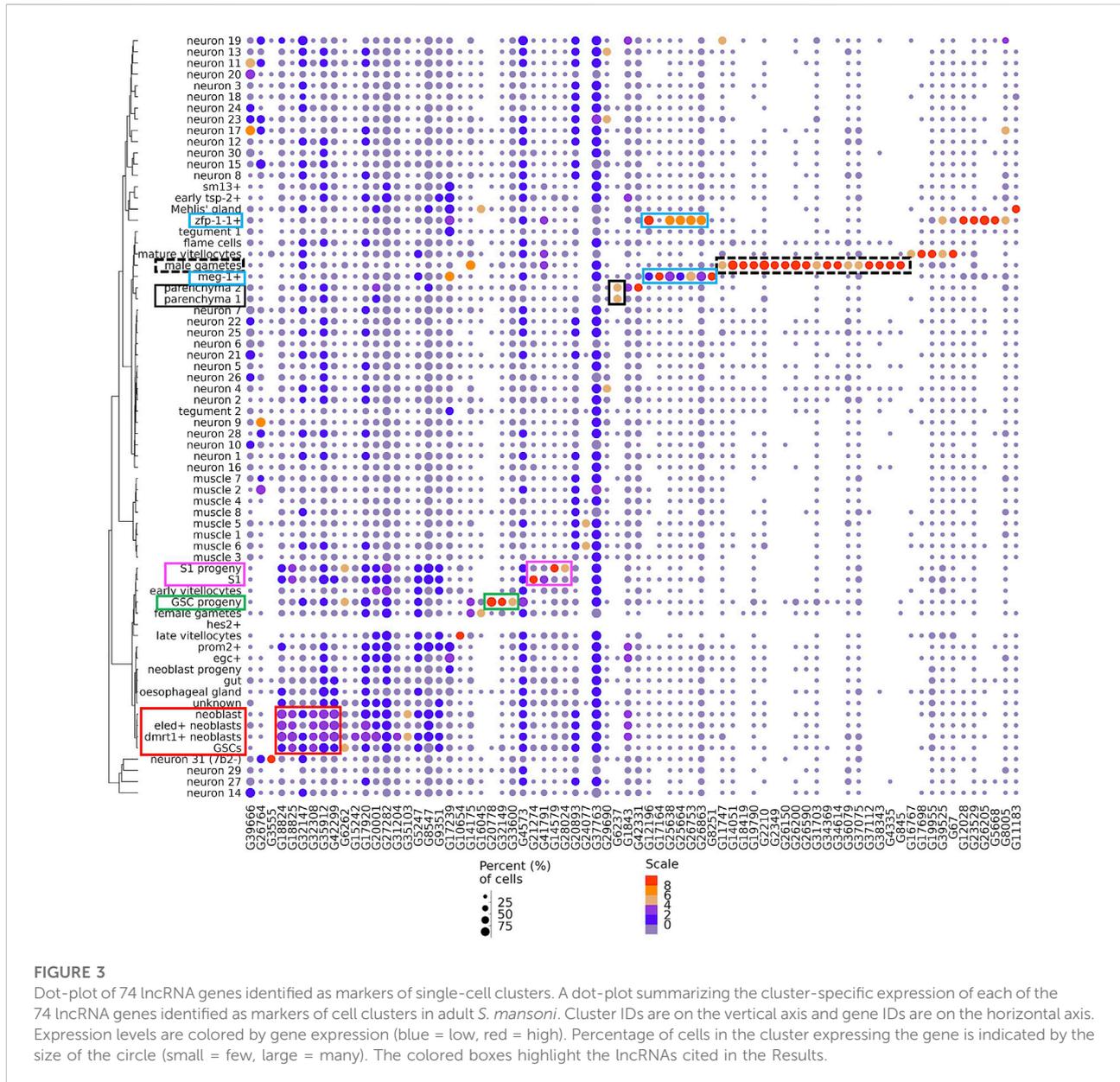
The percentage of cells mapped to each cluster in comparison to the number of cells in the original cluster annotation of Wendt et al. (2020) is shown in Figure 2. For most of the previously annotated clusters (62 out of 68, i.e. 91%) between 66 and 100% of the cells in the cluster were remapped to the same original clusters (Figure 2, see Supplementary Table S4). Note that 44 out of the 62 clusters (i.e. 71%) have between 90 and 100% of the cells coincidentally mapped to the same original clusters (Figure 2, see Supplementary Table S4). Only 6 clusters had less than 66% of the cells coincidentally mapped to the same original clusters; the cluster in which most of the cells were transferred to other clusters was the *hes2<sup>+</sup>*, where 531 out of its 561 cells (94.7%) were transferred to the neoblast progeny cluster and 13 cells (2.3%) to neoblast cluster (Supplementary Table S4), followed by the neuron 19 cluster, where 177 out of the 198 cells (89.4%) were transferred to the neuron 8 cluster. The other four clusters which lost a considerable fraction of the original cells were *dmrt1<sup>+</sup>* neoblasts, where 189 out of 409 cells (46.2%) were transferred to the neoblast cluster; female gametes, where 155 out of 388 cells (39.9%) were transferred to neoblast; mature vitellocytes, where 59 out of 154 cells (38.3%) were transferred to neoblast progeny; and Mehlis' gland, where 64 out of 214 cells (29.9%) were transferred to neoblast progeny and neoblast clusters (Supplementary Table S4). The median gene expression in the same set of cells, grouped by the original cell cluster annotation, were highly similar (Pearson correlation = 0.97–1.00) between the Wendt et al. (2020) reference matrix, which had only protein-coding genes, and the new query matrix, which includes lncRNAs (Supplementary Figure S1A), thus ruling out the possibility of





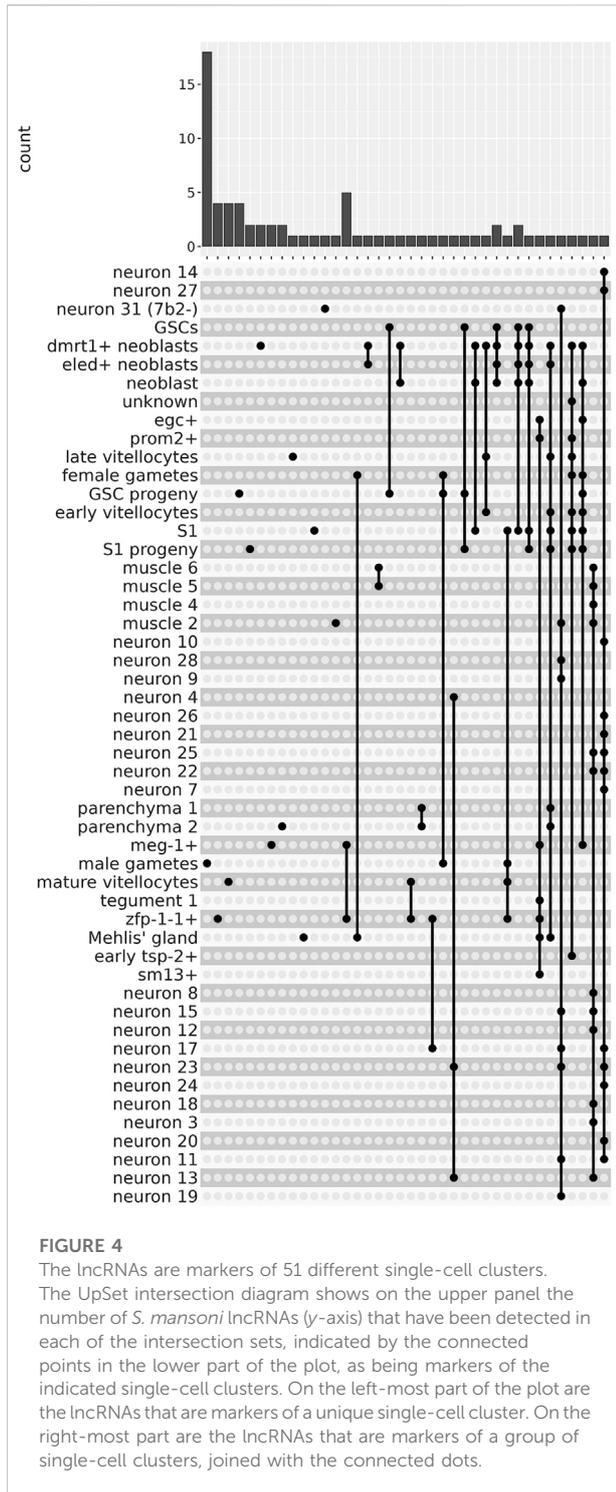
**FIGURE 2**

Percentage of cells mapped to each cluster in comparison to the number of cells in the original cluster annotation. The 68 clusters (indicated at left) are grouped according to the similarity of their gene expression patterns in the new, re-analyzed expression data set. The white bars indicate the percentage of cells that remained in the same cluster in the re-mapped data set, relative to the number of cells in the original cluster annotation. The black bars indicate the final percentage of cells in the cluster in the re-mapped data set, relative to the number of cells in the original cluster annotation. The numbers inside the white bars are the absolute numbers of cells that remained in the clusters after re-mapping, and on the right of the black bars are the absolute numbers of total cells in the clusters in the new, re-mapped data set. The vertical dotted red line goes through the 100% value in the x-axis.



The 3 lncRNAs at the left-most end of the image (Figure 3) were expressed in a number of different neuron clusters, and the next 6 lncRNAs to the right of those were more highly expressed in germline stem cells (GSCs) and neoblasts (Figure 3, red box), clusters of progenitor cells for gametes and somatic tissues, respectively. Three lncRNAs (G20778, G32149 and G33600) were highly expressed only in the GSC progeny cluster (Figure 3, green box), while four lncRNAs were markers of S1 progeny and S1 (Figure 3, magenta box), two of them (lncRNAs G14579 and G28024) in the S1 progeny cluster and two (G21274 and G41791) in the S1 cluster. Interestingly, lncRNA G6237 was detected as expressed only in the parenchyma 1 and parenchyma 2 clusters (Figure 3, black box).

There were 13 clusters in which we were able to identify sets of lncRNAs that were significantly more highly expressed exclusively in a single cluster (Figure 4, left-most side). For example, male gametes cluster had 18 exclusive lncRNA markers (Figure 4 top panel, see Supplementary Table S5 for lncRNA gene names); *zfp-1-1+*, mature vitellocytes and GSC progeny clusters had 4 exclusive lncRNA markers each, and another 9 clusters had one or two exclusive lncRNA markers each (Figure 4 top panel; see Supplementary Table S5). Besides those lncRNAs exclusively more expressed in a single cluster, we identified lncRNA markers that were shared by two or more cluster groups (Figure 4); one such interesting example is lncRNA



G39666 that was a marker of 13 different neuron clusters (Figure 4, right-most end; see Figure 3, left-most lncRNA). These results are in accordance with the findings in human cell lines (Encode\_Project\_Consortium et al., 2007), in which the vast majority of intronic and

intergenic long non-coding RNA transcripts are expressed only in 1 cell line (out of the 11 cell lines tested), while the majority of protein-coding transcripts are expressed in one up to 7 cell lines (out of the 11) (Encode\_Project\_Consortium et al., 2007).

### 3.4 Validation of lncRNA neuron markers

We have selected 16 out of the 74 lncRNA gene markers (Supplementary Table S1) to visualize their sites of expression in adult male and female worm tissues with WISH, based on the clusters where they were identified as markers, on the existence of only one or a few transcript isoforms per gene in the locus, and on the ability to design a probe that only matched a single locus in the genome. A total of 13 out of the 16 selected lncRNAs (81%) were validated, as they were found localized in tissues that are consistent with the cell clusters of which they are markers; of these, 9 were also detected at some other tissues. Three, lncRNAs G38343, G20001, and G17920 were detected in tissues not consistent with the scRNA-seq analysis (see Supplementary Table S1). All of them are described below.

The lncRNA marker of 13 different neuron clusters, lncRNA G39666 (SmLINC173882) (Figure 5A, see Supplementary Table S5), was detected by WISH in the head sides and body of males (Figure 5B, left) as well as in the head and in proximity to the vitellaria of females (Figure 5C, left) with a pattern that is very similar to that of neuroendocrine protein 7b2 gene (Wendt et al., 2020), a general neuron marker. This result supports our previous study, in which G39666 (SmLINC173882) was present in the turquoise gene co-expression network module involved in generation of neurons, synapse, locomotory behavior and axon guidance (Maciel et al., 2019). In fact, dFISH showed that G39666 and the neuroendocrine protein 7b2 messages colocalized in the head side cells of males and in the trunk (Figure 5B, middle, right), as well as in the head of females and in a few neuron cells near the vitellaria (Figure 5C, middle, right).

lncRNA G26764 (SmLNCA149530/1) was identified with scRNA-seq as a marker of 8 different neuron clusters and of the muscle 2 cluster (Figure 6A, see Supplementary Table S5). G26764 lncRNA was detected by WISH as dispersed throughout the bodies of males and females (Figure 6B). dFISH showed that the G26764 lncRNA was detected in the head and trunk cells in well-defined spots near the nuclei of cells expressing the neuroprotein 7b2 mRNA (Figures 6C–F). Consistent with the scRNA-seq data, G26764 was detected also in cells expressing the general muscle marker tropomyosin 2 (*tpm2*) (Figures 6G,H).

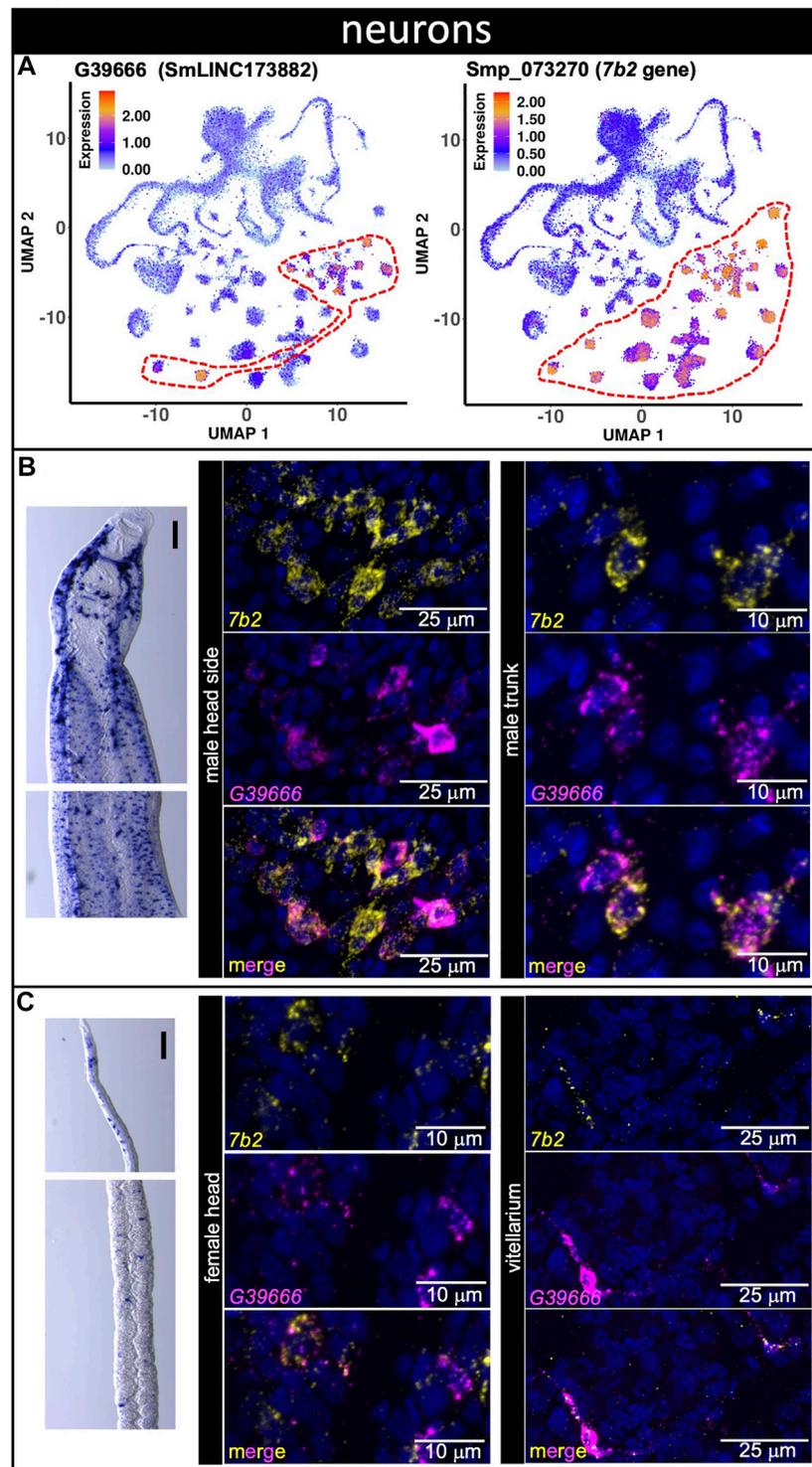
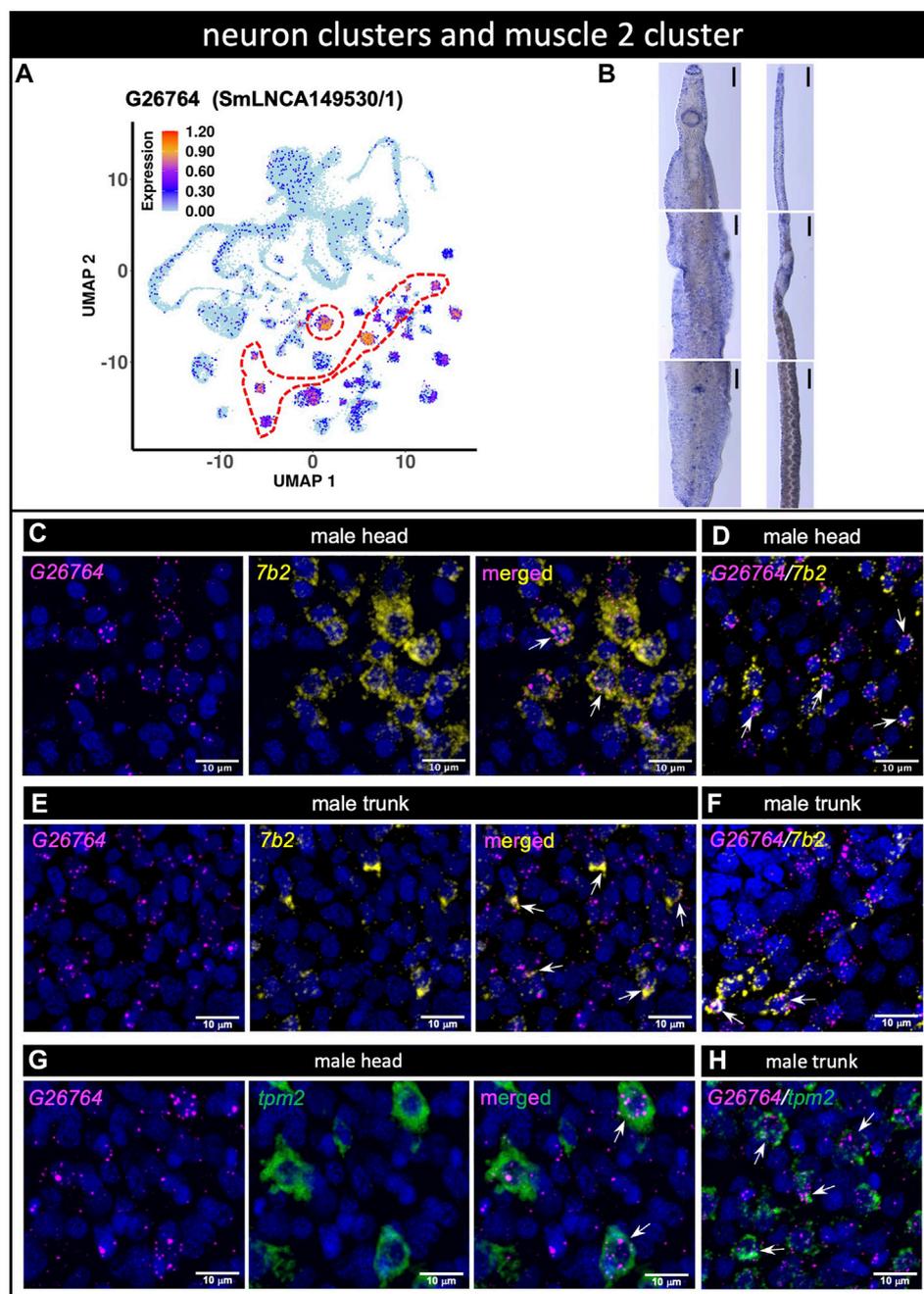
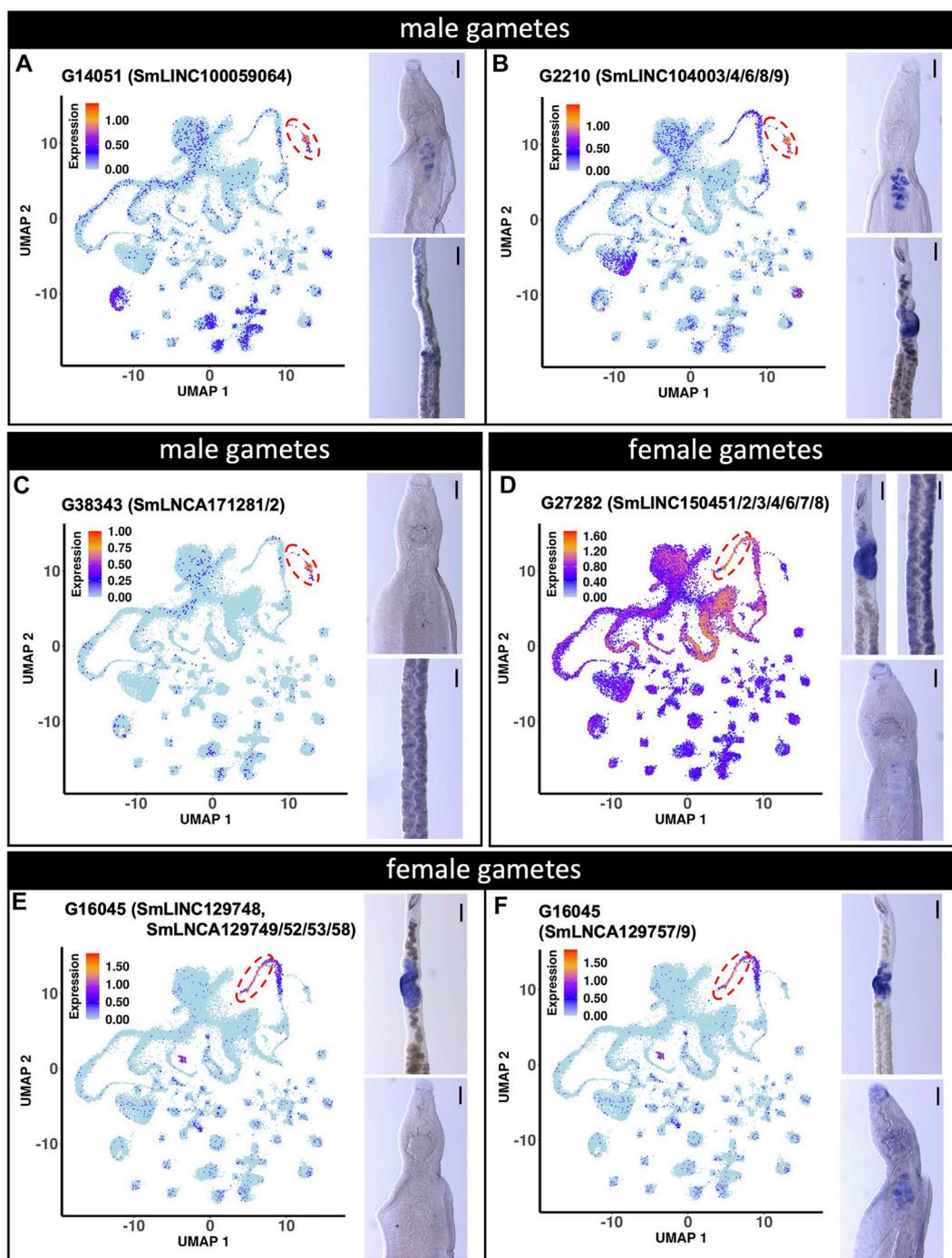


FIGURE 5

A lncRNA marker of 13 different neuron clusters co-localizes with neuroendocrine protein *7b2* message. (A) UMAP plot of lncRNA neurons cluster marker G39666 (left) and of general neuronal marker *7b2* (right). UMAP plots are colored by gene expression (blue = low, red = high) and the scale represents  $\log_{10}(\text{UMIs}+1)$ . The regions enclosed by red dashed lines indicate the location of the relevant neuron clusters on the UMAP plots. (B,C) Whole-mount *in situ* hybridization (WISH) of lncRNA gene G39666 in the head (left, top) and body (left, bottom) of a male (B) or a mature female (C). Scale bars are 100 μm. Double FISH with G39666 lncRNA and *7b2* of male (B) head sides (middle) and trunk (right), and of female (C) head (middle) and vitellarium (right). Nuclei: blue.

**FIGURE 6**

(A) IncRNA marker of 8 different neuron clusters and of muscle 2 cluster co-localizes with neuroendocrine protein *7b2* and muscle tropomyosin genes. (A) UMAP plot of IncRNA cluster marker *G26764*. UMAP plot is colored by gene expression (blue = low, red = high) and the scale represents  $\log_{10}(\text{UMIs}+1)$ . The region enclosed by the red dashed line indicates the location of the relevant neuron cluster on the UMAP plot. (B) WISH of IncRNA *G26764* in the head (top), trunk (middle) and tail (bottom) of a male [left] or a mature female [right]. Scale bars are 100  $\mu$ m. (C–F) Double FISH with IncRNA *G26764* and *7b2* gene in male head (C,D) and trunk (E,F); panels D and F show the dFISH images of other worm sections different from (C) and (E) (G,H) Double FISH with IncRNA *G26764* and the general muscle marker gene *tpm2* tropomyosin in the male head (G) and trunk (H). Nuclei: blue.



**FIGURE 7**

lncRNA markers of male and female gametes single-cell clusters are localized in testis and ovary. (A to C) UMAP plot (left) of the indicated lncRNA marker of male gametes cluster. WISH with the indicated lncRNA in a male head [right, top] and the ovary region of a female [right, bottom]. (D to F) UMAP plot (left) of the indicated lncRNA marker of female gametes cluster. WISH with the indicated lncRNA in a female region of the ovary and vitellarium [D, right, top] and male head [D, right, bottom]. WISH of the ovary region of a female [E,F, right, top] and male head [E,F, right, bottom]. UMAP plots are colored by gene expression (blue = low, red = high) and the scale represents  $\log_{10}(\text{UMIs}+1)$ . The regions enclosed by the red dashed lines indicate the location of the relevant male or female gametes cluster on the UMAP plots. WISH scale bars are 100  $\mu\text{m}$ .

### 3.5 LncRNAs as markers of reproductive tissues

The expression of three male gamete-enriched lncRNA markers were also evaluated (Figure 7). G14051 (SmLINC100059064) and G2210 (SmLINC104003/4/6/8/9) lncRNAs were confirmed by WISH to be expressed in the male testis (Figures 7A,B). Surprisingly, G14051 expression was also detected by WISH in mature vitellocytes (Figure 7A), despite not being detected in these cells by scRNA-seq. Nevertheless, a re-analysis of publicly-available RNA-seq data ((Silveira et al., 2021), <https://verjolab.shinyapps.io/Reference-genes/>) showed evidence of G14051 (SmLINC100059064) transcription in female worms, albeit at a level 3 to 6 times lower than in males. Perhaps not surprisingly given its modest level of expression in female germ cells, G2210 was also detected by WISH in the female ovary (Figure 7B). Despite its high-level of expression in male germ cells by scRNA-seq, by WISH G38343 (SmLNCA171281/2) lncRNA showed little expression in male worms but was weakly expressed in cells in or near the vitellaria of female schistosomes (Figure 7C).

We also examined the expression of two female gametes lncRNA markers (Figures 7D–F). G27282 (SmLINC150451/52/53/54/56/57/58) lncRNA expression was confirmed by WISH to be expressed in the ovary (female gametes) (Figure 7D). Interestingly, as suggested by our scRNA-seq analysis, G27282 was also detected by WISH to be expressed in the vitellarium (Figure 7D). The other lncRNA, G16045 had 12 different transcript isoforms detected in the scRNA-seq dataset ([http://verjolab.usp.br:8081/cluster\\_view/G16045](http://verjolab.usp.br:8081/cluster_view/G16045)). Two different probes were designed for the transcripts in the G16045 gene locus, each representing one of two groups of transcripts. Each group has a common last exon, which is different in the two groups of transcripts in the G16045 gene locus (see this locus in the genome browser, along with mapping of the pairs of primers that were used to generate the two probes). The transcripts in the group SmLINC129748/SmLNCA129749/52/53/58 (Figure 7E) were confirmed by WISH to be expressed in the female ovary and absent from male testis, in accordance with our previously published work, in which SmLINC129748/SmLNCA129752 was expressed in females and almost absent in male worms ((Silveira et al., 2021), <https://verjolab.shinyapps.io/Reference-genes/>). Interestingly, the other group of transcript isoforms representing SmLNCA129757/9 (Figure 7F) was detected by WISH as expressed in the ovary, and also in male testis, illustrating that a different lncRNA transcript isoform from a single locus can have their expression differentially regulated in different adult worm tissues.

A marker detected with scRNA-seq in the mature vitellocytes cluster, lncRNA G17698 (SmLINC132934, SmLNCA132935/36/37/38/40/41) (Supplementary Figure S2, top) was confirmed by WISH to be well expressed in mature vitellocytes throughout the female bodies (Supplementary Figure S2, bottom). Of note,

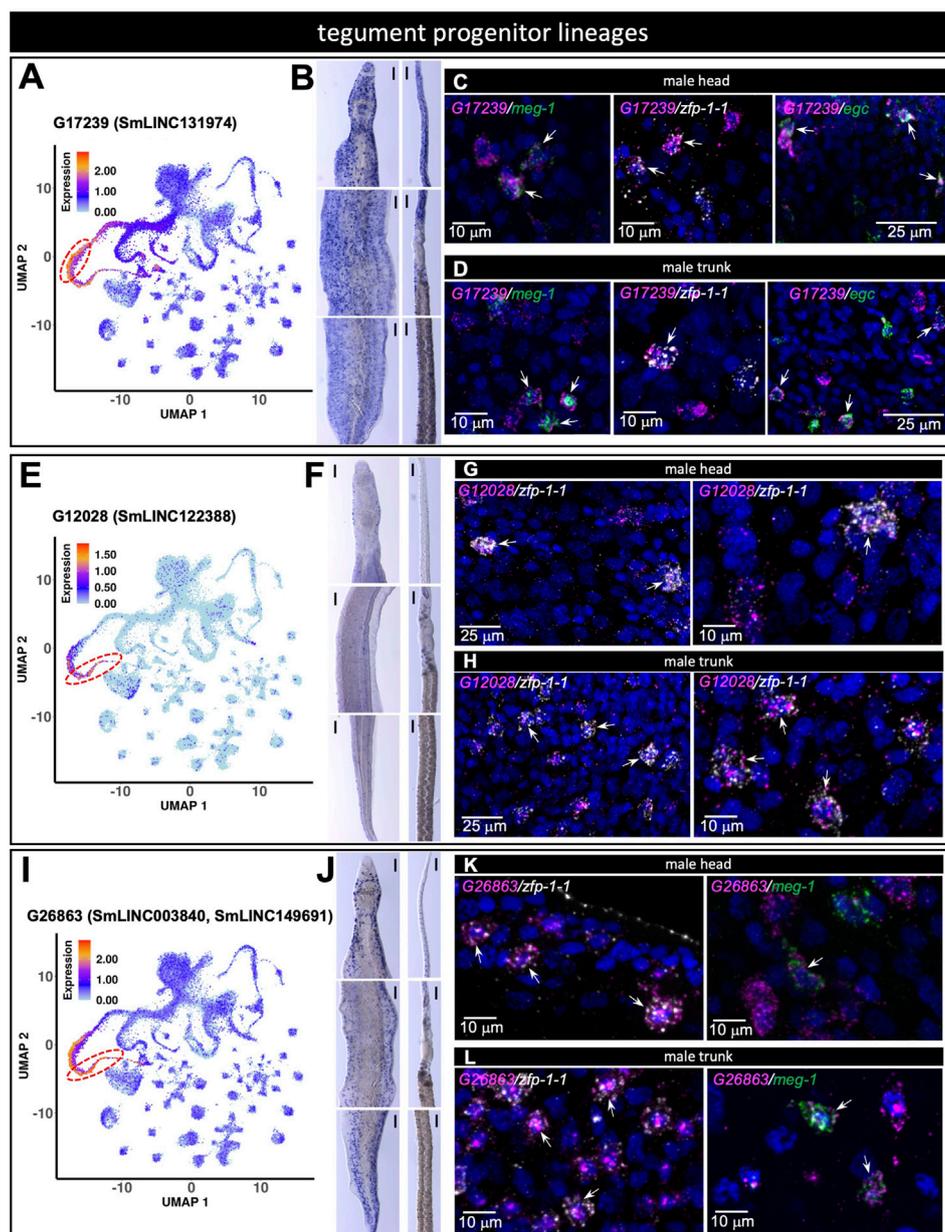
transcripts in the lncRNA G17698 locus were detected as belonging to the female pink gene co-expression module related to endoplasmic reticulum, protein and glycoprotein biosynthetic processes (Maciel et al., 2019), which are functions important for egg production.

### 3.6 Validation of lncRNA tegument markers

Tegument progenitor cluster markers were assayed with three probes representing different marker lncRNAs. G17239 (SmLINC131974) was identified with scRNA-seq as a marker of *meg-1*<sup>+</sup> cells (Figure 8A) and was detected by WISH throughout the head and body of both male and female worms (Figure 8B). dFISH analysis confirmed the co-localization of G17239 with *meg-1* transcripts in the male head and trunk (Figures 8C,D). We observed similar co-localization patterns of G17239 with *zfp-1-1* and *egc* messages, corroborating the detection with scRNA-seq of G17239 in *zfp-1-1*<sup>+</sup> and *egc*<sup>+</sup> clusters (Figure 8A). The second tegument progenitor marker, G12028 (SmLINC122388) was identified with scRNA-seq as a marker of *zfp-1-1*<sup>+</sup> cluster (Figure 8E), and was detected in small numbers of cells by WISH in the head and body of both male and female worms (Figure 8F); dFISH confirmed that G12028 and *zfp-1-1* messages were co-localized in the head and trunk of males (Figures 8G,H). The third marker, G26863 (SmLINC003840, SmLINC149691) was identified by scRNA-seq as a marker of the *zfp-1-1*<sup>+</sup> cluster (Figure 8I) with expression also in the *meg-1*<sup>+</sup> cluster. WISH showed expression of G26863 both in the head and trunk of male and female schistosomes (Figure 8J) and dFISH confirmed G26863 co-expression with *zfp-1-1* and with *meg-1* in male head and trunk (Figures 8K, L).

Three additional lncRNA probes were assayed as tegument progenitor *zfp-1-1*<sup>+</sup> cluster markers, namely G25638 (SmLINC147486/7/8), G12196 (SmLINC002017/25, SmLINC122706/7) and G26205 (SmLINC148606) (Supplementary Figure S3A–L). The three were similarly detected by WISH throughout the tegumental lineage, as described above, and were confirmed with dFISH to co-localize with *zfp-1-1* message in the male head and trunk, except for G26205, which was detected with dFISH only in the trunk (Supplementary Figure S3A–L). A weighted gene co-expression analysis had previously identified that expression of all these tegument lncRNA markers is correlated with the brown or turquoise co-expression modules involved in focal adhesion, actin cytoskeleton, cell and adherens junctions or contractile fibers (Maciel et al., 2019), which are cellular components consistent with their finding in the tegument.

Finally, Supplementary Figure S4 shows two markers whose localizations were not confirmed. lncRNA G20001 (SmLINC137107), a parenchyma 1 marker identified with scRNA-seq (Supplementary Figure S4A), was detected by



**FIGURE 8**

lncRNA markers of tegument progenitor lineages co-localize with protein-coding genes known to mark those tegument progenitors. (A,E,I) UMAP plot of the indicated lncRNA marker of tegument progenitor lineages. (B,F,J) WISH with the indicated lncRNA in a male [left] and a female [right] head [top], trunk [middle] and tail [bottom]. (C,G,K) Double FISH in male head with the indicated lncRNA and the general tegument progenitor marker genes *meg-1*, *zfp-1-1*, *egc* (C), *zfp-1-1* (G), *zfp-1-1*, *meg-1* (K). (D,H,L) Double FISH in male trunk with the indicated lncRNA and the general tegument progenitor marker genes *meg-1*, *zfp-1-1*, *egc* (D), *zfp-1-1* (H), *zfp-1-1*, *meg-1* (L). UMAP plots are colored by gene expression (blue = low, red = high) and the scale represents  $\log_{10}(\text{UMIs}+1)$ . The regions enclosed by the red dashed lines indicate the location of the relevant *meg-1*<sup>+</sup> (A) or *zfp-1-1*<sup>+</sup> (E,I) clusters on the UMAP plots. WISH scale bars are 100  $\mu\text{m}$ .

WISH in a pattern (Supplementary Figure S4B) that is similar to the pattern of labeling of parenchyma 1 determined by Wendt et al. (2020), however dFISH did not show colocalization with parenchyma marker *tgfb1* (Supplementary Figures S4C-H). lncRNA G17920 (SmLINC133371), was identified by scRNA-

seq as a marker of the *dmrt*<sup>+</sup> and *eled*<sup>+</sup> neoblasts (see Supplementary Table S5). However, we detected no expression consistent with neoblast expression (Supplementary Figure S4H). Instead, by WISH we detected G17920 most highly in the germ cells of male and female worms, which is not unexpected given

the detection of G17920 expression in these cell types by scRNA-seq (Figure S4G). The WISH approach depends on the *in situ* accessibility of the probe to the target transcript, which may be more tightly associated with different protein or DNA partners in different tissues, and eventually not available to base-pair with the probe.

### 3.7 lncRNAs detected as expressed exclusively in one cluster

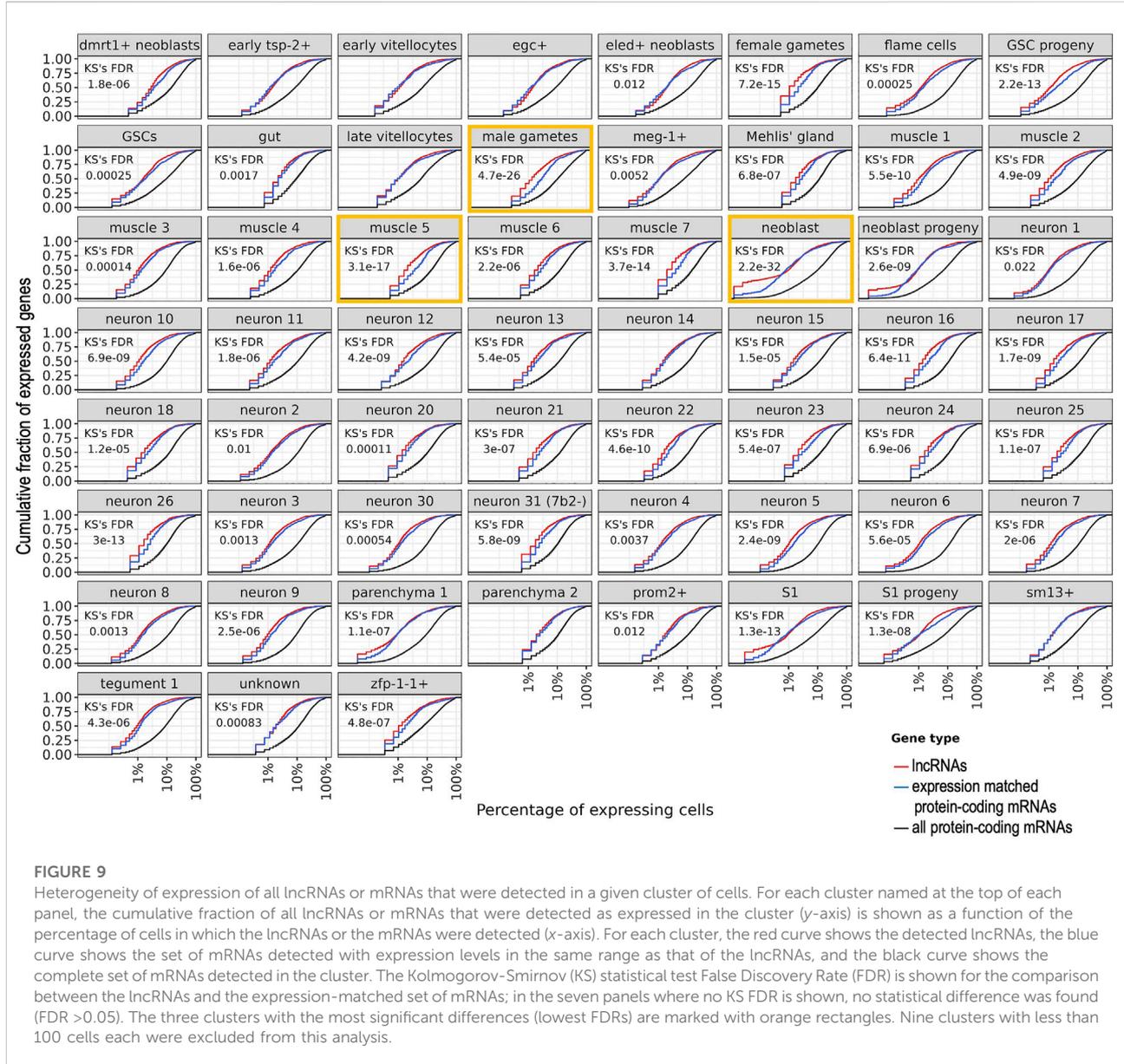
It is known in certain human cell types that the expression of thousands of lncRNAs is more heterogeneous than the expression of mRNAs (Lv et al., 2016; Yunusov et al., 2016), and it has been proposed that averaging transcriptomes over thousands of cells masks the presence of rare cells with high lncRNAs expression (Mercer et al., 2008; Yan et al., 2013). In fact, an analysis of single-cell RNA-seq data from each of five different human cell types (Yunusov et al., 2016) showed that, when comparing on a cell-to-cell basis the lncRNAs to the protein-coding mRNAs that are expressed at similar low levels in a given cell type, there is a statistically significant higher heterogeneity of expression of lncRNAs (Yunusov et al., 2016), possibly reflecting the specific roles played by lncRNAs on different individual cells that are not synchronized among them in a given tissue. Therefore, we postulated that another way of identifying interesting cluster-enriched lncRNAs, possibly important for function in *S. mansoni* adult worms, was to look for lncRNAs that were detected as expressed in only one cluster, and in at least 10% of the cells of that given cluster. This stands as a complementary way to look for cluster-enriched lncRNAs, besides finding lncRNA markers, which are the lncRNAs significantly more highly expressed in a given cluster compared with the median expression in all other clusters.

We found 204 lncRNAs that were detected as expressed in at least 10% of cells exclusively in one cluster among the 68 clusters analyzed in this work, with no other cluster expressing the indicated lncRNA (Supplementary Table S6). Interestingly, male gametes cluster has 55 such exclusive lncRNAs, with G38343 (SmLNCA171281/2), the most frequent one, being expressed in 74% of the male gamete cells. In fact, there are 10 lncRNAs expressed in more than 50% of male gamete cells (Supplementary Table S6), with the remaining 45 lncRNAs being expressed in the range of 45 to 10% of the cells. Female gametes cluster has 10 lncRNAs exclusively expressed in more than 10% of cells, the most frequent one, G29240 (SmLINC154048) being expressed in 32% of the female gamete cells (Supplementary Table S6). Late vitellocytes has 16, and mature vitellocytes has 9 lncRNAs exclusively expressed in more than 10% of cells; the most frequent in late vitellocytes is G25294 (SmLNCA146832), expressed in 42% of the cells, and in mature vitellocytes the most frequent is G17698 (SmLINC132934 to SmLINC132941), expressed in 57% of the mature vitellocyte cells (Supplementary Table S6).

Noteworthy, the tegument progenitor zfp-1-1<sup>+</sup> cluster has 14 lncRNAs, and the tegument 1 cluster has only one lncRNA, G29145 (SmLINC153880/1/3) exclusively expressed in more than 10% of cells (Supplementary Table S6). Interestingly, G29145 was found to be expressed in 25% of cells in the tegument 1 cluster, and in no other cluster it was expressed in at least 10% of cells. When observed in the sub-sets of scRNA-seq data ([http://verjolab.usp.br:8081/cluster\\_view/G29145](http://verjolab.usp.br:8081/cluster_view/G29145)) G29145 showed a sex-specific expression, being detected only in the tegument 1 cluster of immature and mature females, with no expression detected in males.

We observed that for each cluster of cells analyzed in this work, when looking at all expressed genes, not necessarily exclusively detected in any cluster, there was a correlation between the level of expression of the genes and the fraction of cells from the cluster in which the genes were detected, both for lncRNAs and mRNAs (Supplementary Figure S5A, B), which indicates that the depth of RNA-sequencing might impact the frequency of detection of lowly expressed genes. Nevertheless, we observed many conspicuous outliers that were expressed at high levels and yet were detected in only 10–30% of the cells (Supplementary Figure S5), suggesting that they were genes that could play a specific role in a fraction of cells in that cluster.

To evaluate the lncRNAs expression heterogeneity across cells, we then computed the cumulative fraction of all lncRNAs or mRNAs that were detected in one cluster as a function of the percentage of cells in which the lncRNAs or the mRNAs were detected (Figure 9). In all but seven clusters, a statistically significant lower percentage of cells were detected as expressing lncRNAs compared to cells expressing the set of mRNAs of similar expression levels (FDR = 0.022 to  $2.2 \times 10^{-32}$ , Kolmogorov-Smirnov KS test). The top three clusters with higher heterogeneity of expression of lncRNAs compared to the set of expression-matched mRNAs were neoblasts, male gametes, and muscle 5 (Figure 9, orange rectangles). Of note, half of the lncRNAs expressed in one given cluster were detected in up to 1–3% of cells (Figure 9, red curves), whereas the mRNAs of similar expression levels were detected in a significantly higher percentage of cells (Figure 9, blue curves are significantly shifted to the right compared to the red curves), and analyzing the complete set of mRNAs expressed in one given cluster, half of the mRNAs were detected in up to 10–20% of cells (Figure 9, black curves). Because a considerable number of lncRNAs were detected in 1–3% of cells, we again searched for lncRNAs exclusively expressed in a single cluster, this time in at least 1% of cells, using a stringent requirement of exclusive expression, namely that the lncRNA was not expressed in more than 1% of cells in any other cluster (Supplementary Figure S6). Again, male gametes cluster has the highest number of exclusive lncRNAs, followed by late vitellocytes and female gametes (Supplementary Figure S6); the list of all lncRNAs and protein-coding mRNAs expressed in at least 1% of cells exclusively in one cluster is given



in [Supplementary Table S7](#). Interestingly, the exclusive protein-coding mRNAs detected in the gut cluster include Cathepsins B, L and S, Saposin B domain-containing protein, Prosaposin, Phospholipase A, and Sphingomyelin phosphodiesterase 2, whereas the exclusive protein-coding mRNAs detected in the oesophageal gland cluster include MEGs 4, 9, 11 and 32.2, Annexin, Cystatin, and Natterin-4 ([Supplementary Table S7](#)). These lncRNAs and mRNAs might play specific roles in a fraction of cells in those clusters.

Confirming that lncRNAs expression is heterogeneous across cells, we observed that only 626 different lncRNAs (8%) were non-exclusively detected in at least 10% of cells in any cluster, when compared with all 8023 different lncRNAs

detected with scRNA-seq across all *S. mansoni* cells. For comparison, 7563 Smp protein-coding messages (81%) were detected in at least 10% of cells in any cluster, compared with all 9388 different Smpls detected with scRNA-seq across all *S. mansoni* cells.

[Supplementary Table S8](#) shows the number of lncRNAs and of Smp protein-coding genes not exclusively expressed in any cluster and detected as expressed in 10, 25, 50, 75 or 95% of cells in each cluster. The top 5 clusters with the highest numbers of lncRNAs (in at least 10% of the cells) were neuron 19, dmrt1<sup>+</sup> neoblasts, male gametes, eled<sup>+</sup> neoblasts, GSC progeny; note that the same lncRNA is counted multiple times, when it is expressed in multiple clusters.

## 4 Discussion

In this work we have re-analyzed publicly available scRNA-seq data obtained from *S. mansoni* adult male and immature and mature adult female, showing that lncRNAs are differentially expressed across different single-cell clusters. Whole-mount *in situ* hybridization and double fluorescence *in situ* hybridization confirmed the localization of most of the single-cell cluster lncRNA markers in specific adult *S. mansoni* tissues. Our re-analyses detected the expression of 8023 lncRNAs, 79.4% of all 10,110 lncRNAs known in *S. mansoni* (Maciel et al., 2019). Further scRNA-seq with deeper coverage and including other life-cycle stages may reveal a more detailed pattern of expression of lncRNAs possibly involved in parasite development and homeostasis.

Here, our approach was to use the set of protein-coding gene markers of single-cell clusters that had been extensively characterized in adult *S. mansoni* (Wendt et al., 2020) to probe lncRNAs tissue co-localization. For this, we relied on the well-documented strategy of transfer of cell clusters ID annotations from one set of scRNA-seq data used as reference, to another query set (Stuart et al., 2019). We found that in 91% of the previously annotated clusters, the great majority of cells (66–100%) were re-mapped to the same original clusters, attesting to the robustness of the transfer method (Stuart et al., 2019). In fact, only two original clusters were depleted by over 50% of their cells, and the cells were transferred to closely related clusters, namely neuron 19, where 89.4% of the cells were transferred to neuron 8 cluster, and *hes2*<sup>+</sup>, where 94.7% of the cells were transferred to the neoblast progeny cluster. Two factors may have played a role in the transfer of cells to different clusters. First, the expression level of genes may have changed because the raw scRNA-seq reads have been re-mapped to the genome and counted with the STARsolo tool, rather than with CellRanger used in the original paper (Wendt et al., 2020); STARsolo uses a different algorithm to quantify gene expression, which results in a higher number of recovered cells compared with CellRanger (Kaminow et al., 2021). We checked the similarity of the two sets of original and re-mapped expression data of protein-coding genes and found that they are highly correlated (0.972–1.0, Pearson correlation), thus ruling out a major impact of gene expression counting on the discrepant re-mapping of cells. A second factor could be the similarity between the overall expression profiles among certain clusters, which may affect the identification of proper anchor cells that are used for guiding the assignment of cells to clusters (Stuart et al., 2019). Despite these factors, cells were mostly transferred to related clusters, thus not impairing the ability to use the established protein-coding gene markers as a tool to determine co-localization of lncRNAs in the parasite tissues.

An average of 1.65 lncRNA isoforms per lncRNA gene in *S. mansoni* was identified in our previous work using RNA-seq libraries from whole worms at different stages, from isolated tissues, from cell populations, and from single-cells (Maciel et al., 2019). Among the different isoforms in a gene locus, there are alternate transcription start sites (TSSs), alternate use of exons, and exon skipping; exon splice sites identified for all these lncRNA isoforms have canonical GU/AG splicing acceptor/donor pairs (Maciel et al., 2019), thus making them bona-fide alternatively spliced messages in *S. mansoni*, in analogy to the large number of lncRNA isoforms in animals (Ulitsky, 2016); of note, no systematic analysis of the functional impact of lncRNA isoforms on *S. mansoni* biology has been documented. Here, we have selected one female gametes lncRNA marker, G16045 with 12 transcript isoforms, and we used two probes that encompass two groups of isoforms, with two different TSSs and two different last exons. We observed that one group of isoforms was expressed only in the ovary, while the other was expressed in the ovary and in the testis, giving support for a tissue-specific use of different lncRNA isoforms in *S. mansoni*. Further exploration of the wide occurrence of lncRNAs alternative splicing is warranted.

In *S. mansoni*, lncRNA knock-down with dsRNA caused important phenotypic changes such as a decreased worm viability and impaired oviposition (Silveira et al., 2022). We propose that the lncRNAs identified here as single-cell cluster markers might be good candidates to be targeted and possibly interfere with adult schistosomes homeostasis, especially those of tegument and gametes lineages. We corroborate with findings from the literature, which show that gamete lineages in animals and plants are rich in lncRNAs expression (Golicz et al., 2018; Guo et al., 2018). Four gametes marker lncRNAs were confirmed with WISH, two in male gametes (out of the three tested) and two in female gametes (Figure 7). Given their conspicuous expression in the gametes, these lncRNAs might be important for fertilized egg production and can be candidate targets to be silenced and potentially disrupt the completion of the parasite life cycle.

We found that in *S. mansoni*, the tegument progenitor lineages express a high number of lncRNAs compared with other clusters. Tegument interface protects the parasite from host (Skelly and Wilson, 2006; Wendt et al., 2018). We found a female-specific lncRNA in the tegument (G29145), and 7 lncRNA markers that are expressed in *meg-1*<sup>+</sup> and *zfp-1-1*<sup>+</sup> clusters, two clusters that belong to the tegument lineage. With dFISH, localization of 6 lncRNA markers in the *meg-1*<sup>+</sup>, *zfp-1-1*<sup>+</sup> and *egc*<sup>+</sup> clusters was confirmed. They might be good candidate targets to interfere with tegumental development, thus breaking the parasite-host barrier.

lncRNAs are known in other organisms to act in the nucleus (as enhancers, histone modification modulators, and activators/inhibitors of transcription) and in the cytoplasm (by inhibition of translation) (Statello et al., 2021). Here we observed that the

marker of neuron clusters and muscle 2 cluster, lncRNA G26764 (SmLNCA149530/1) is a good example of nuclear localization, in cells where the messages of neuropeptide *7b2* gene neuron marker and of *tpm2* gene muscle marker are predominantly localized in the cytoplasm (Figures 6C,G).

Interestingly, lncRNAs expression distribution across cells in a given cluster was significantly more heterogeneous than that of protein-coding mRNAs expressed at levels similar to those of the lncRNAs, for all but seven clusters, among the 59 clusters analyzed here (Figure 9). While only 626 out of 8023 lncRNAs (8%) are expressed in at least 10% of cells from a given cluster, with a median of 126 lncRNAs per cluster (1.6%), a median of 4060 mRNAs per cluster out of 9388 protein-coding mRNAs (43%) are expressed in at least 10% of cells from a given cluster (Supplementary Table S8). lncRNAs' cell-to-cell expression heterogeneity seems to epitomize one of the fundamental properties of the lncRNA expression patterns (Lv et al., 2016; Yunusov et al., 2016). Nevertheless, one cannot rule out the possibility that the low levels of expression of a set of lncRNAs might impose difficulties in detecting them, thus resulting into some overestimation of the extent of expression heterogeneity. The significantly higher cell-to-cell expression heterogeneity of lncRNAs compared to mRNAs might be related to the fact that while proteins are expected to play basal roles that are shared between different cells in a given tissue, lncRNAs are expressed with considerably higher tissue-specificity, developmental stage-specificity, and cell-subtype specificity (Liu et al., 2016; Yunusov et al., 2016). In fact, it has recently been shown that lncRNAs are expressed with higher cell-to-cell variability than mRNAs across a wide range of expression levels in mouse fibroblasts, in mouse embryonic stem cells and in human HEK293 cells, highlighting lncRNAs with cell state-specific functions involved in cell cycle progression and apoptosis (Johnsson et al., 2022). There are intrinsic differences in transcriptional bursting kinetics between lncRNAs and mRNAs, with lncRNAs having a fourfold lower burst frequency compared to mRNAs and only a twofold decrease in burst size (Johnsson et al., 2022). Thus, the decreased expression of lncRNAs is mainly achieved through a longer duration between transcriptional bursts of expression, which accounts for a high cell-to-cell heterogeneity of lncRNAs expression (Johnsson et al., 2022). In this regard, the half-life of a class of lncRNAs has been shown in humans to be shorter than that of mRNAs (Ayupe et al., 2015). Altogether, our data is compatible with a transient, desynchronized expression of lncRNAs in a diverse population of cells from the same tissue, which calls the attention to the fact that lncRNAs with low population-level abundance might instead be expressed at high levels in a subset of individual cells within that population, where they may have important functions.

lncRNAs may act *in cis*, regulating the neighbor protein-coding genes (Rinn and Chang, 2020). Localization of a lncRNA in the genome, and identification of protein-coding gene

neighbors can give clues to possible mechanisms of action. Curiously, lncRNA G39666 (SmLINC173882) neuron marker is located in an intergenic region ([http://genome.verjolab.usp.br/cgi-bin/hgTracks?hgS\\_doLoadUrl=submit&hgS\\_loadUrlName=http://genome.verjolab.usp.br/folders/geneNetwork/schMan3/tracks/genes/lincRNAs/htmlPage/Morales-VicenteG39666publicLocus.txt](http://genome.verjolab.usp.br/cgi-bin/hgTracks?hgS_doLoadUrl=submit&hgS_loadUrlName=http://genome.verjolab.usp.br/folders/geneNetwork/schMan3/tracks/genes/lincRNAs/htmlPage/Morales-VicenteG39666publicLocus.txt)), between neural-cadherin (Smp\_306,450.1) and an uncharacterized protein (Smp\_084010.1) conserved in helminths. Because expression of the latter was strongly detected in almost all neuron clusters ([http://verjolab.usp.br:8081/cluster\\_view/Smp-084010](http://verjolab.usp.br:8081/cluster_view/Smp-084010)), further studies could elucidate a possible regulatory function of the lncRNA on the expression of this protein coding gene.

In conclusion, in this study we provide a comprehensive view of the expression of lncRNAs in the different cell types of adult *S. mansoni*, paving the way for functional studies of lncRNAs as potential regulators of the parasite homeostasis.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

## Author contributions

MA and SV-A conceived the project. DM-V performed the *in silico* analyses. LZ performed the hybridization experiments and analyses. GS selected the genes for validation, designed primers, cloned probes and prepared labeled probes for the hybridization experiments. MA and SV-A wrote the paper. AT contributed with informatics resources. DM-V, LZ, GS, MA, JC, and SV-A analyzed and interpreted the data. All authors read and reviewed the manuscript.

## Funding

This work was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) grant number 2018/23.693-5 to SVA and by a grant from the NIH (R01AI121037) to JJC. GOS received a FAPESP fellowship (number 2018/24.015-0) and DM-V received a fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Finance Code 001). SVA laboratory was also supported by institutional funds from Fundação Butantan and received an established investigator fellowship award from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brasil.

## Acknowledgments

We thank Vinicius C. Mesel (<http://vmesel.com>) for providing services for the construction of the website tool. We thank Michael Reese, University of Texas, Southwestern Medical Center, for advice on the website design.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- Ayupe, A. C., Tahira, A. C., Camargo, L., Beckedorff, F. C., Verjovski-Almeida, S., and Reis, E. M. (2015). Global analysis of biogenesis, stability and sub-cellular localization of lncRNAs mapping to intragenic regions of the human genome. *RNA Biol.* 12, 877–892. doi:10.1080/15476286.2015.1062960
- Bergquist, R., Utzinger, J., and Keiser, J. (2017a). Controlling schistosomiasis with praziquantel: How much longer without a viable alternative? *Infect. Dis. Poverty* 6, 74. doi:10.1186/s40249-017-0286-2
- Bergquist, R., Zhou, X. N., Rollinson, D., Reinhard-Rupp, J., and Klohe, K. (2017b). Elimination of schistosomiasis: The tools required. *Infect. Dis. Poverty* 6, 158. doi:10.1186/s40249-017-0370-7
- Chen, J., Wang, Y., Wang, C., Hu, J. F., and Li, W. (2020). LncRNA functions as a new emerging epigenetic factor in determining the fate of stem cells. *Front. Genet.* 11, 277. doi:10.3389/fgene.2020.00277
- Choudhary, S., and Satija, R. (2022). Comparison and evaluation of statistical error models for scRNA-seq. *Genome Biol.* 23, 27. doi:10.1186/s13059-021-02584-9
- Colley, D. G., Bustinduy, A. L., Secor, W. E., and King, C. H. (2014). Human schistosomiasis. *Lancet* 383, 2253–2264. doi:10.1016/S0140-6736(13)61949-2
- Collins, J. J., 3rd, Wendt, G. R., Iyer, H., and Newmark, P. A. (2016). Stem cell progeny contribute to the schistosome host-parasite interface. *Elife* 5, e12473. doi:10.7554/eLife.12473
- Diaz Soria, C. L., Lee, J., Chong, T., Coghlan, A., Tracey, A., Young, M. D., et al. (2020). Single-cell atlas of the first intra-mammalian developmental stage of the human parasite *Schistosoma mansoni*. *Nat. Commun.* 11, 6411. doi:10.1038/s41467-020-20092-5
- Encode\_Project\_Consortium/Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigo, R., Gingeras, T. R., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816. doi:10.1038/nature05874
- Golicz, A. A., Bhalla, P. L., and Singh, M. B. (2018). lncRNAs in plant and animal sexual reproduction. *Trends Plant Sci.* 23, 195–205. doi:10.1016/j.tplants.2017.12.009
- Guo, J., Grow, E. J., Mlcochova, H., Maher, G. J., Lindskog, C., Nie, X., et al. (2018). The adult human testis transcriptional cell atlas. *Cell Res.* 28, 1141–1157. doi:10.1038/s41422-018-0099-2
- Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* 20, 296. doi:10.1186/s13059-019-1874-1
- Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Iii, Zheng, S., Butler, A., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* 184, 3573–3587.e29. doi:10.1016/j.cell.2021.04.048
- Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P., and Berriman, M. (2017). WormBase ParaSite - a comprehensive resource for helminth genomics. *Mol. Biochem. Parasitol.* 215, 2–10. doi:10.1016/j.molbiopara.2016.11.005
- Jandura, A., and Krause, H. M. (2017). The new RNA world: Growing evidence for long noncoding RNA functionality. *Trends Genet.* 33, 665–676. doi:10.1016/j.tig.2017.08.002
- Jiang, W., Qu, Y., Yang, Q., Ma, X., Meng, Q., Xu, J., et al. (2019). D-Lnc: A comprehensive database and analytical platform to dissect the modification of drugs on lncRNA expression. *RNA Biol.* 16, 1586–1591. doi:10.1080/15476286.2019.1649584
- Johnsson, P., Ziegenhain, C., Hartmanis, L., Hendriks, G. J., Hagemann-Jensen, M., Reinius, B., et al. (2022). Transcriptional kinetics and molecular functions of long noncoding RNAs. *Nat. Genet.* 54, 306–317. doi:10.1038/s41588-022-01014-1
- Kaminow, B., Yunusov, D., and Dobin, A. (2021). STARsolo: Accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. *bioRxiv* 2005, 442755. doi:10.1101/2021.05.05.442755
- Kittur, N., King, C. H., Campbell, C. H., Kinung'hi, S., Mwinzi, P. N. M., Karanja, D. M. S., et al. (2019). Persistent hotspots in schistosomiasis consortium for operational research and evaluation studies for gaining and sustaining control of schistosomiasis after four years of mass drug administration of praziquantel. *Am. J. Trop. Med. Hyg.* 101, 617–627. doi:10.4269/ajtmh.19-0193
- Lex, A., Gehlenborg, N., Strobel, H., Vuillemot, R., and Pfister, H. (2014). UpSet: Visualization of intersecting sets. *IEEE Trans. Vis. Comput. Graph.* 20, 1983–1992. doi:10.1109/TVCG.2014.2346248
- Li, P., Nanes Sarfati, D., Xue, Y., Yu, X., Tarashansky, A. J., Quake, S. R., et al. (2021). Single-cell analysis of *Schistosoma mansoni* identifies a conserved genetic program controlling germline stem cell fate. *Nat. Commun.* 12, 485. doi:10.1038/s41467-020-20794-w
- Liu, S. J., Nowakowski, T. J., Pollen, A. A., Lui, J. H., Horlbeck, M. A., Attenello, F. J., et al. (2016). Single-cell analysis of long non-coding RNAs in the developing human neocortex. *Genome Biol.* 17, 67. doi:10.1186/s13059-016-0932-1
- Lv, D., Wang, X., Dong, J., Zhuang, Y., Huang, S., Ma, B., et al. (2016). Systematic characterization of lncRNAs' cell-to-cell expression heterogeneity in glioblastoma cells. *Oncotarget* 7, 18403–18414. doi:10.18632/oncotarget.7580
- Maciell, L. F., Morales-Vicente, D. A., Silveira, G. O., Ribeiro, R. O., Olberg, G. G. O., Pires, D. S., et al. (2019). Weighted gene Co-expression analyses point to long non-coding RNA hub genes at different *Schistosoma mansoni* life-cycle stages. *Front. Genet.* 10, 823. doi:10.3389/fgene.2019.00823
- Mcmanus, D. P., Dunne, D. W., Sacko, M., Utzinger, J., Vennervald, B. J., and Zhou, X. N. (2018). Schistosomiasis. *Nat. Rev. Dis. Prim.* 4, 13. doi:10.1038/s41572-018-0013-8
- Mercer, T. R., Dinger, M. E., Sunkin, S. M., Mehler, M. F., and Mattick, J. S. (2008). Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl. Acad. Sci. U. S. A.* 105, 716–721. doi:10.1073/pnas.0706729105
- Molehin, A. J. (2020). Schistosomiasis vaccine development: Update on human clinical trials. *J. Biomed. Sci.* 27, 28. doi:10.1186/s12929-020-0621-y
- Nath, A., Lau, E. Y. T., Lee, A. M., Geeleher, P., Cho, W. C. S., and Huang, R. S. (2019). Discovering long noncoding RNA predictors of anticancer drug sensitivity beyond protein-coding genes. *Proc. Natl. Acad. Sci. U. S. A.* 116, 22020–22029. doi:10.1073/pnas.1909998116
- R\_Core\_Team (2018). R: A language and environment for statistical computing. [Online]. Available at: <https://www.R-project.org/> (Accessed).
- Ransohoff, J. D., Wei, Y., and Khavari, P. A. (2018). The functions and unique features of long intergenic non-coding RNA. *Nat. Rev. Mol. Cell Biol.* 19, 143–157. doi:10.1038/nrm.2017.104

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.924877/full#supplementary-material>

- Rinn, J. L., and Chang, H. Y. (2020). Long noncoding RNAs: Molecular modalities to organismal functions. *Annu. Rev. Biochem.* 89, 283–308. doi:10.1146/annurev-biochem-062917-012708
- Silveira, G. O., Amaral, M. S., Coelho, H. S., Maciel, L. F., Pereira, A. S. A., Olberg, G. G. O., et al. (2021). Assessment of reference genes at six different developmental stages of *Schistosoma mansoni* for quantitative RT-PCR. *Sci. Rep.* 11, 16816. doi:10.1038/s41598-021-96055-7
- Silveira, G. O., Coelho, H. S., Amaral, M. S., and Verjovski-Almeida, S. (2022). Long non-coding RNAs as possible therapeutic targets in protozoa, and in *Schistosoma* and other helminths. *Parasitol. Res.* 121, 1091–1115. doi:10.1007/s00436-021-07384-5
- Skelly, P. J., and Wilson, R. A. (2006). Making sense of the schistosome surface. *Adv. Parasitol.* 63, 185–284. doi:10.1016/S0065-308X(06)63003-0
- Stattolo, L., Guo, C. J., Chen, L. L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* 22, 96–118. doi:10.1038/s41580-020-00315-9
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., 3rd, et al. (2019). Comprehensive integration of single-cell data. *Cell* 177, 1888–1902. doi:10.1016/j.cell.2019.05.031
- Tarashansky, A. J., Xue, Y., Li, P., Quake, S. R., and Wang, B. (2019). Self-assembling manifolds in single-cell RNA sequencing data. *Elife* 8, e48994. doi:10.7554/eLife.48994
- Tebeje, B. M., Harvie, M., You, H., Loukas, A., and Mcmanus, D. P. (2016). Schistosomiasis vaccines: Where do we stand? *Parasit. Vectors* 9, 528. doi:10.1186/s13071-016-1799-4
- Ulitsky, I. (2016). Evolution to the rescue: Using comparative genomics to understand long non-coding RNAs. *Nat. Rev. Genet.* 17, 601–614. doi:10.1038/nrg.2016.85
- Vale, N., Gouveia, M. J., Rinaldi, G., Brindley, P. J., Gartner, F., and Correia Da Costa, J. M. (2017). Praziquantel for schistosomiasis: Single-drug metabolism revisited, mode of action, and resistance. *Antimicrob. Agents Chemother.* 61, e02582–02516. doi:10.1128/AAC.02582-16
- Wang, B., Lee, J., Li, P., Saberi, A., Yang, H., Liu, C., et al. (2018). Stem cell heterogeneity drives the parasitic life cycle of *Schistosoma mansoni*. *Elife* 7, e35449. doi:10.7554/eLife.35449
- Wendt, G. R., Collins, J. N., Pei, J., Pearson, M. S., Bennett, H. M., Loukas, A., et al. (2018). Flatworm-specific transcriptional regulators promote the specification of tegumental progenitors in *Schistosoma mansoni*. *Elife* 7, e33221. doi:10.7554/eLife.33221
- Wendt, G., Zhao, L., Chen, R., Liu, C., O'donoghue, A. J., Caffrey, C. R., et al. (2020). A single-cell RNA-seq atlas of *Schistosoma mansoni* identifies a key regulator of blood feeding. *Science* 369, 1644–1649. doi:10.1126/science.abb7709
- Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., et al. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* 20, 1131–1139. doi:10.1038/nsmb.2660
- Yunusov, D., Anderson, L., Dasilva, L. F., Wysocka, J., Ezashi, T., Roberts, R. M., et al. (2016). HIPSTR and thousands of lncRNAs are heterogeneously expressed in human embryos, primordial germ cells and stable cell lines. *Sci. Rep.* 6, 32753. doi:10.1038/srep32753
- Zhou, J., Xu, J., Zhang, L., Liu, S., Ma, Y., Wen, X., et al. (2019). Combined single-cell profiling of lncRNAs and functional screening reveals that H19 is pivotal for embryonic hematopoietic stem cell development. *Cell Stem Cell* 24, 285–298. doi:10.1016/j.stem.2018.11.023