UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA

Programa de Pós-Graduação em Ciências Biológicas (Bioquímica)

GILBERT DE OLIVEIRA SILVEIRA

Caracterização funcional de RNAs longos não-codificadores de proteínas em *Schistosoma mansoni*

Versão corrigida da Tese conforme Resolução CoPGr 5890
O original se encontra disponível na Secretaria de Pós-Graduação do IQ-USP

São Paulo

Data do Depósito na SPG:

04/10/2022

GILBERT DE OLIVEIRA SILVEIRA

Caracterização funcional de RNAs longos não-codificadores de proteínas em *Schistosoma mansoni*

Tese apresentada ao Instituto de Química da Universidade de São Paulo para a obtenção do Título de Doutor em Ciências (Bioquímica)

Orientador: Prof. Dr. Sergio Verjovski-Almeida

São Paulo

2022

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletronico, para fins de estudo e pesquisa, desde que citada a fonte.

Ficha Catalográfica elaborada eletronicamente pelo autor, utilizando o programa desenvolvido pela Seção Técnica de Informática do ICMC/USP e adaptado para a Divisão de Biblioteca e Documentação do Conjunto das Químicas da USP

Bibliotecária responsável pela orientação de catalogação da publicação:
Marlene Aparecida Vieira - CRB - 8/5562

Silveira, Gilbert de Oliveira
S587c Caracterização funcional de RNAs longos não-codificadores de proteínas em Schistosoma mansoni / Gilbert de Oliveira Silveira. - São Paulo, 2022.
190 p.

Tese (doutorado) - Instituto de Química da Universidade de São Paulo. Departamento de Bioquímica.
Orientador: Verjovski-Almeida, Sergio

1. Longos RNAs não-codificadores de proteínas. 2. Schistosoma mansoni. 3. Parasitas. 4. Pareamento. 5. Drogas. I. T. II. Verjovski-Almeida, Sergio, orientador.

Agradecimentos

Agradeço a vida por ter me proporcionado tantas oportunidades e me fazer uma pessoa melhor.

Agradeço a minha mãe, por todo amor, atenção, calma e dedicação. Foram muitos tombos e muitos contratempos, mas graças aos seus ensinamentos consegui superá-los. Muito obrigado por respeitar as minhas escolhas e acima de tudo apoiá-las.

Agradeço aos meus irmãos, em especial ao menor. Me desculpe por não ter presenciado seu crescimento, e muito obrigado por entender a distância. Muito obrigado pelo apoio, mesmo sem saber o que isso quer dizer.

Agradeço ao meu Psicanalista Michel por ter trabalhado tanto para eu não surtar mais, rs. A melhor coisa que pude fazer na vida é buscar um profissional para me ajudar a passar por toda esta barra que é lidar com a pressão e toxicidade do doutorado. Não foi fácil e será muito mais fácil daqui para frente, tendo em vista que consegui o tão sonhado emprego e finalmente consegui sair do âmbito acadêmico. Muito obrigado.

Agradeço ao Gabriel, por ser o marido que é. Muito obrigado por estar presente e por me auxiliar psicológico e emocionalmente durante todo este processo. O doutorado tende a ser uma fase bem difícil, mas se tornou menos pior com você do meu lado. Sei que não somos o casal de pote de margarina, mas to feliz estampando os potes de Nutella mesmo, rs. Te amo branco.

Agradeço às minhas duas filhas que nunca lerão isso, rs. Muito obrigado Blue e Maya por todo o suporte emocional e todo o carinho que me proporcionaram, mesmo sem saber que estavam fazendo isso tudo.

Agradeço a Karina por ser quem é. Muito obrigado por ser atenciosa, calma e tão amorosa. Desculpa por ser um ogro contigo. Irmãos, não são somente aqueles de sangue! Te amo! Agradeço ao Victor por sempre estar lá quando eu precisei e por sempre me apoiar em qualquer que seja a escolha que eu faça.

Agradeço a Adriana e Camilla por terem me ajudado a passar por toda essa barra que foi o Doutorado. Agradeço por todos ensinamentos, por me incentivar sempre a ser uma pessoa melhor e por me dar os puxões de orelha quando preciso rs. Agradeço por ter me suportado ao longo de tantos surtos, por serem amigas guerreiras e a Adriana por estar aguentando as buchas de ser minha amiga. Agradeço por me criarem oportunidades de boas memórias mesmo com tanto caos que foi esse doutorado. Amo vocês chatonildas.

Agradeço as minhas filhas de coração Helena e Giovanna. Eu nunca achei que fosse conseguir me relacionar de forma tão profunda com as minhas primeiras ICs oficiais. Helena, nosso começo foi turbulento por preconceito meu, mas hoje eu me orgulho de dizer que te conheci, aprendi muito contigo e sou muito grato pelo tempo que passamos juntos. Giovanna, você é um anjo caído do céu que veio pra minha vida para me atormentar e me dar muita alegria, rs. Muito obrigado por ser quem é e por ter me permitido ser seu co-co-orientador. A vocês duas, muito obrigado por terem sido tão excepcionais e por terem me deixado fazer parte da carreira de vocês que será maravilhosa, como já antecipei e antecipo constantemente. Amo vocês, pestes.

Agradeço a Ana Paula por ser a mãezona do lab e por me ajudar com tantas coisas pessoais e profissionais. Minha melhor vizinha rs. Agradeço a Ana Tahira por ser essa luz e

calmaria na minha vida. Eu realmente aprendi muita coisa contigo e não posso ser mais grato por toda a colaboração profissional e pessoal. Muito obrigado Tahira!!

Agradeço a Gabriela por ter participado de parte do meu percurso ao longo do doutoramento com essa energia boa, feliz e com apreciação por cafés e boas conversas. Por sempre me dar conselhos bons e por me dar o suporte quando precisei!

Agradeço a Daisy por ter sido a parceira de luta no doutorado. Apesar dos grandes solavancos que a vida nos deu, nossa história será sim contada com um sendo coadjuvante da história do outro. Muito obrigado por insistir que no final da tudo certo, mesmo eu sendo imediatista e não suportando a ideia de que eu não tenho controle rs. No final, deu tudo certo e eu sou o mais novo mensageiro da palavra da Daisy rs.

Agradeço ao Lucas Maciel, Vinícius, João, David, Leandro, João Victor (Victoria Beckham), Asaph, David e demais alunos que passaram pelo lab e fizeram da minha estadia mais agradável. Agradeço ao David Pires, Lucas Maciel, João e Vinícius por todo suporte com demandas de Bioinformática.

Agradeço a minha ex-orientadora Carla. Devo a ela todos os ensinamentos que a minha mãe dizia que a vida ia ensinar. Obrigado por todo o apoio, compreensão e ensinamentos. Obrigado por me aturar, entender, aceitar e ajudar por tanto e tanto tempo, rs. Obrigado pelas cobranças e pela orientação sem as quais não teria alcançado meu objetivo. Muito obrigado por tudo!

Agradeço ao Professor Collins por ter me recebido para um sanduíche curto de 1 mês e ter me mostrado que existe luz no fim do túnel. Que existem laboratórios saudáveis, com pessoas fazendo ciência de ponta e chefes preocupados com a saúde mental dos seus alunos/pos-docs.

Agradeço ao Murilo e ao Professor Sergio pelos ensinamentos profissionais e orientação.

Agradeço aos membros da banca por aceitarem prontamente o convite para participarem da banca de julgamento desta Tese de doutorado.

Agradeço ao CNPq e à FAPESP (2018/24015-0) pelo fundamental apoio financeiro.

Agradeço a todos que contribuíram de alguma forma para a conclusão deste trabalho.

RESUMO

SILVEIRA, G.O. **Caracterização funcional de RNAs longos não-codificadores de proteínas em *Schistosoma mansoni*.** 2022. (190p). Tese de Doutorado – Programa de Pós-graduação em Bioquímica. Instituto de Química, Universidade de São Paulo, São Paulo.

A esquistossomose é uma importante doença parasitária com alto impacto nas taxas de morbidade e mortalidade, afetando mais de 230 milhões de pessoas em 76 países. *Schistosoma mansoni* é a espécie prevalente na África e na América Latina, apresentando um ciclo de vida complexo com seis estágios diferentes: ovos, miracídios, esporocistos, cercárias, esquistossômulos, machos adultos e fêmeas adultas (dimorfismo sexual). A compreensão da biologia do esquistossoma em nível molecular pode sugerir novas alternativas terapêuticas. RNAs longos não-codificadores de proteínas (lncRNAs) são RNAs com mais de 200 nucleotídeos com baixo ou nenhum potencial de codificação de proteínas que, em humanos e muitas outras espécies, podem atuar como reguladores da expressão de genes codificadores de proteínas, manutenção de células-tronco e resistência a drogas. Devido à sua expressão tecido-específica e funções multifacetadas, os lncRNAs foram propostos como novos alvos terapêuticos em doenças humanas. Nesta Tese, produzimos uma revisão bibliográfica dos trabalhos que identificaram lncRNAs em protozoários, em *Schistosoma* e outros helmintos, e demonstramos pela primeira vez os impactos na fisiologia de *S. mansoni* causados pelo silenciamento *in vitro* de um lncRNA intergênico (Introdução). Em seguida, determinamos os genes de referência adequados para a normalização de dados de RT-qPCR de amostras dos seis diferentes estágios de desenvolvimento de *S. mansoni* (Capítulo 1). Mais adiante, mostramos por re-análises de dados públicos de RNA-Seq de parasitas fêmeas tratadas com 5-Aza-Citidina (uma droga epigenética inibidora da oviposição e desenvolvimento dos ovários das fêmeas), que centenas de lncRNAs eram diferencialmente expressos entre as condições controle e tratado. Muitos lncRNAs pertenciam a módulos de co-expressão relacionados com metabolismo em machos, e alguns foram validados por RT-qPCR (Capítulo 2). Posteriormente, reanalisando dados públicos de RNA-Seq de célula única (scRNA-seq) de *S. mansoni* adultos caracterizamos o perfil de expressão de lncRNAs nos 68 grupos de células identificados. Os grupos de células que continham a maioria dos lncRNAs marcadores eram gametas masculino e feminino e células progenitoras de tegumento. Identificamos lncRNAs marcadores específicos de células neurais. Por hibridização *in situ* de parasita inteiro, com marcação simples ou dupla, localizamos a expressão específica de 13 dos 16 lncRNAs marcadores selecionados (Capítulo 3). Por fim a re-análise de dados públicos de RNA-Seq de vermes adultos, recuperados de hamsters infectados com cercárias de um único sexo ou dos dois sexos, identificou milhares de lncRNAs diferencialmente expressos. Selecionamos doze lncRNAs e validamos seus níveis de expressão em um modelo similar de parasitas mantidos pareados ou não em cultivos *in vitro*. O silenciamento *in vitro* e *in vivo* de quatro dos lncRNAs selecionados mostrou que eles desempenham papéis fundamentais na proliferação celular nos vermes adultos e suas gônadas e são essenciais para a manutenção da vitelária das fêmeas, reprodução do verme adulto e desenvolvimento de ovos. Hibridização *in situ* mostrou que esses lncRNAs são expressos em tecidos que se correlacionam com os fenótipos observados no silenciamento *in vitro* (Capítulo 4). De modo geral, esses resultados mostram que os lncRNAs são componentes essenciais da biologia do *S. mansoni*, apresentando potencial significativo como novos candidatos a alvos terapêuticos.

Palavras-chave: RNA longos não-codificadores, parasitas, *Schistosoma mansoni*

ABSTRACT

SILVEIRA, G.O. **Functional characterization of long non-coding RNAs in *Schistosoma mansoni*.** 2022. (190p). Ph.D. Thesis – Graduate Program in Biochemistry. Instituto de Química, Universidade de São Paulo, São Paulo.

Schistosomiasis is an important parasitic disease with a high impact on morbidity and mortality rates, affecting more than 230 million people in 76 countries. *Schistosoma mansoni* is the most prevalent species in Africa and Latin America, presenting a complex life cycle with six different stages: eggs, miracidia, sporocysts, cercariae, schistosomula, adult males and adult females (sexual dimorphism). Understanding schistosome biology at the molecular level may suggest new therapeutic alternatives. Long non-protein-coding RNAs (lncRNAs) are RNAs of more than 200 nucleotides with low or no protein-coding potential that, in humans and many other species, can act as regulators of protein-coding gene expression, stem-cell maintenance and drug resistance. Due to their tissue-specific expression and multifaceted functions, lncRNAs have been proposed as new therapeutic targets in human diseases. In this Thesis, we produced a literature review of the works that identified lncRNAs in protozoa, *Schistosoma* and other helminths, and we demonstrated for the first time the impacts on *S. mansoni* physiology caused by the *in vitro* silencing of an intergenic lncRNA (Introduction). We then determined the appropriate reference genes for normalizing RT-qPCR data from samples from the six different developmental stages of *S. mansoni* (Chapter 1). Further, we showed by re-analyses of public RNA-Seq data from female parasites treated with 5-Aza-Cytidine (an epigenetic drug that inhibits female oviposition and ovarian development), that hundreds of lncRNAs were differentially expressed between control and treated conditions. Many lncRNAs belonged to metabolism-related co-expression modules in males, and some were validated by RT-qPCR (Chapter 2). Subsequently, by reanalyzing public data from single-cell RNA-Seq (scRNA-seq) from adult *S. mansoni*, we characterized the expression profile of lncRNAs in the 68 identified cell clusters. The cell clusters that contained the most lncRNA markers were male and female gametes and tegument progenitor cells. We identified neural cell-specific marker lncRNAs. By whole mount *in situ* hybridization, with single or double labeling, we localized the specific expression of 13 of the 16 selected marker lncRNAs (Chapter 3). Finally, re-analysis of public RNA-Seq data from adult worms, recovered from hamsters infected with single-sex or dual-sex cercariae, identified thousands of differentially expressed lncRNAs. We selected twelve lncRNAs and validated their expression levels in a model of *in vitro* cultures of paired or unpaired parasites that is similar to what performed before. *In vitro* and *in vivo* silencing of four of the selected lncRNAs showed that they play key roles in cell proliferation in adult worms and their gonads and are essential for female vitellaria maintenance, adult worm reproduction and egg development. *In situ* hybridization has shown that these four lncRNAs are expressed in tissues that correlate with the phenotypes observed in the *in vitro* silencing (Chapter 4). Overall, these results show that lncRNAs are essential components of *S. mansoni* biology, presenting significant potential as new candidates for therapeutic targets.

Keywords: long non-coding RNAs, parasites, *Schistosoma mansoni*

SUMÁRIO

1. INTRODUÇÃO

1.1. RNAs longos não-codificadores como possíveis alvos terapêuticos em protozoários, em *Schistosoma* e outros helmintos

PREÂMBULO

Contribuições do Doutorando Gilbert de Oliveira Silveira para o manuscrito apresentado nesta sessão:

Busca da literatura, análise de dados, realização dos ensaios de silenciamento *in vitro*, escrita do manuscrito e revisão do manuscrito.

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Sep 11, 2022

This Agreement between Prof. Sergio Verjovski-Almeida ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5386170716555
License date	Sep 11, 2022
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Parasitology Research
Licensed Content Title	Long non-coding RNAs as possible therapeutic targets in protozoa, and in Schistosoma and other helminths
Licensed Content Author	Gilbert O. Silveira et al
Licensed Content Date	Dec 3, 2021
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	full article/chapter
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	yes
Title	PhD Thesis of Gilbert O. Silveira
Institution name	Universidade de Sao Paulo - USP
Expected presentation date	Nov 2022
Requestor Location	Prof. Sergio Verjovski-Almeida Av. Prof. Lineu Prestes 748 Butantan Sao Paulo, SP 05508-000 Brazil Attn: Universidade de São Paulo
Billing Type	Invoice
Billing Address	Prof. Sergio Verjovski-Almeida Av. Prof. Lineu Prestes 748 Butantan Sao Paulo, Brazil 05508-000 Attn: Universidade de São Paulo
Total	0.00 USD
Terms and Conditions	

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. Grant of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.

1. 2. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

1. 3. If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Scope of Licence

2. 1. You may only use the Licensed Content in the manner and to the extent permitted by these Ts&Cs and any applicable laws.

2. 2. A separate licence may be required for any additional use of the Licensed Material, e.g. where a licence has been purchased for print only use, separate permission must be obtained for electronic re-use. Similarly, a licence is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence. Any content owned by third parties are expressly excluded from the licence.

2. 3. Similarly, rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

2. 4. Where permission has been granted **free of charge** for material in print, permission may also be granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

2. 5. An alternative scope of licence may apply to signatories of the STM Permissions Guidelines, as amended from time to time.

3. Duration of Licence

3. 1. A licence for is valid from the date of purchase ('Licence Date') at the end of the relevant period in the below table:

Scope of Licence	Duration of Licence
Post on a website	12 months
Presentations	12 months
Books and journals	Lifetime of the edition in the language purchased

3. Acknowledgement

4. 1. The Licensor's permission must be acknowledged next to the Licenced Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

5. Restrictions on use

5. 1. Use of the Licensed Material may be permitted for incidental promotional use and minor editing privileges e.g. minor adaptations of single figures, changes of format, colour and/or style where the adaptation is credited as set out in Appendix 1 below. Any other changes including but not limited to, cropping, adapting, omitting material that affect the meaning, intention or moral rights of the author are strictly prohibited.

5. 2. You must not use any Licensed Material as part of any design or trademark.

5. 3. Licensed Material may be used in Open Access Publications (OAP) before publication by Springer Nature, but any Licensed Material must be removed from OAP sites prior to final publication.

6. Ownership of Rights

6. 1. Licensed Material remains the property of either Licensor or the relevant third party and any rights not explicitly granted herein are expressly reserved.

7. Warranty

IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND
WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

8. Limitations

8. 1. ***BOOKS ONLY:*** Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

8. 2. For content reuse requests that qualify for permission under the STM Permissions Guidelines, which may be updated from time to time, the STM Permissions Guidelines supersede the terms and conditions contained in this licence.

9. Termination and Cancellation

9. 1. Licences will expire after the period shown in Clause 3 (above).

9. 2. Licensee reserves the right to terminate the Licence in the event that payment is not received in full or if there has been a breach of this agreement by you.

Appendix 1 — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT]** (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.3

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Parasitology Research (2022) 121:1091–1115
https://doi.org/10.1007/s00436-021-07384-5

HELMINTHOLOGY - REVIEW

Long non-coding RNAs as possible therapeutic targets in protozoa, and in *Schistosoma* and other helminths

Gilbert O. Silveira[1,2] · Helena S. Coelho[1] · Murilo S. Amaral[1] · Sergio Verjovski-Almeida[1,2]

Received: 2 June 2021 / Accepted: 14 November 2021 / Published online: 3 December 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Long non-coding RNAs (lncRNAs) emerged in the past 20 years due to massive amounts of scientific data regarding transcriptomic analyses. They have been implicated in a plethora of cellular processes in higher eukaryotes. However, little is known about lncRNA possible involvement in parasitic diseases, with most studies only detecting their presence in parasites of human medical importance. Here, we review the progress on lncRNA studies and their functions in protozoans and helminths. In addition, we show an example of knockdown of one lncRNA in *Schistosoma mansoni*, SmLINC156349, which led to in vitro parasite adhesion, motility, and pairing impairment, with a 20% decrease in parasite viability and 33% reduction in female oviposition. Other observed phenotypes were a decrease in the proliferation rate of both male and female worms and their gonads, and reduced female lipid and vitelline droplets that are markers for well-developed vitellaria. Impairment of female worms' vitellaria in SmLINC156349-silenced worms led to egg development deficiency. All those results demonstrate the great potential of the tools and methods to characterize lncRNAs as potential new therapeutic targets. Further, we discuss the challenges and limitations of current methods for studying lncRNAs in parasites and possible solutions to overcome them, and we highlight the future directions of this exciting field.

Keywords lncRNAs · Parasites · Protozoa · Helminths · Therapeutic targets · Novel technologies · Functional characterization

The rise of lncRNAs

The central dogma of molecular biology says that genetic information travels from DNA through RNA ending at protein synthesis (Crick 1958, 1970). In the past century, RNA characterization was mainly focused on the intermediaries in the pathway to protein production such as the housekeeping RNAs (tRNAs, rRNAs, and mRNAs) (Eddy 2001). Proteins were considered the paramount end pieces of genetic information, though their genes comprised less than 3% of the human genome (Rajic et al. 2005; Djebali et al. 2012; Piovesan et al. 2019).

At the end of the twentieth century, the first reports on RNAs that lack protein-coding potential from the nematode *Caenorhabditis elegans* were published, in which *lin-4* and *lin-7* were first described as conserved functional molecules required for the free-living roundworm (Lee et al. 1993; Reinhart et al. 2000). Later, with the advent of high-throughput RNA sequencing (RNA-Seq) technologies, the studies started focusing on RNAs that do not encode proteins (non-coding RNAs, ncRNAs) present in a broad range of lengths and functions (Wang et al. 2009). As for their lengths, ncRNAs are divided into small ncRNAs (miRNAs, piRNAs, siRNAs, crasiRNAs, telsRNAs, and others that are 20–50 nucleotides long), medium size ncRNAs (snoRNAs, tiRNAs, snRNAs, scRNAs, and others with a size between 50 and 200 nucleotides) and long ncRNAs (lncRNAs) that are longer than 200 nucleotides (Derrien et al. 2012; Cech and Steitz 2014; Anastasiadou et al. 2018; Dahariya et al. 2019). A glossary has been provided for guidance on acronyms identification.

Guest Editor: Anja Taubert

✉ Murilo S. Amaral
murilo.amaral@butantan.gov.br

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

[1] Laboratório de Parasitologia, Instituto Butantan, São Paulo, SP 05503-900, Brazil

[2] Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, SP 05508-900, Brazil

LncRNAs biogenesis is primarily similar to mRNAs; as such, they are transcribed by RNA polymerase II, have 5′-cap m^7guanosine, 3′-end poly(A) tails, and undergo splicing (Statello et al. 2021). Some particular classes of lncRNAs mapping to intragenic regions, such as intronic lncRNAs or antisense lncRNAs have characteristic turnover rates, with antisense lncRNAs (median t1/2 = 3.9 h) being on average significantly ($p < 0.0001$) more stable than mRNAs (median t1/2 = 3.2 h) (Ayupe et al. 2015). Many lncRNAs are localized in the nucleus, are less evolutionarily conserved, contain fewer and longer exons, and are much less abundantly expressed than mRNAs (Hezroni et al. 2015; Quinn and Chang 2016). However, some exceptions have been published recently in which the distinct transcription, processing, export, and lncRNA turnover are closely related to its function in the context of different cellular fates (Guo et al. 2020).

lncRNAs can be classified according to two major features: (1) location in the genome; and (2) functional mechanism of action (Ma et al. 2013; Quinn and Chang 2016; Ransohoff et al. 2018; Statello et al. 2021). Here, we will focus on classifying lncRNAs based on their location in the genome. Thus, a sense lncRNA (slncRNA) is transcribed from the sense strand of a protein-coding gene and contains exons from a protein-coding gene, overlapping with part of protein-coding genes or covering the entire sequence of the protein-coding gene through an intron (Fig. 1). An antisense lncRNA (alncRNA) is transcribed from the antisense strand of a protein-coding gene, overlapping with exonic or intronic regions or covering the entire protein-coding sequence through an intron. A bidirectional lncRNA (blncRNA) is transcribed from the opposite strand, in the opposite direction and within 1 kb of the promoter of a protein-coding gene. An intronic lncRNA (ilncRNA) is transcribed entirely from introns of protein-coding genes. Finally, intergenic lncRNAs (lincRNAs) can be transcribed from any DNA strand and must not be near another gene locus (Fig. 1). All of these lncRNAs are expected not to be translated into proteins or eventually to have short (< 300 nt) open reading frames (ORFs) encoding short peptides (< 100 amino acids) (Guttman et al. 2009; Cabili et al. 2011; Dahariya et al. 2019).

LncRNAs have been studied in diverse organisms (Fang and Fullwood 2016; Schmitt and Chang 2016; Novačić et al. 2020; Ren et al. 2021; Choudhary et al. 2021), although less than a hundred have been mechanistically described in detail (Wang and Chang 2011; Rinn and Chang 2012; Statello et al. 2021). Gene expression regulation by lncRNAs can be exploited on multiple levels. They can interact with DNA, RNA, and/or proteins and thus modulate chromatin structure and function by either *cis*- or *trans*-acting mechanisms (Quinn and Chang 2016; Statello et al. 2021). A common feature for many lncRNAs function is to promote gene silencing, and one of the most studied lncRNAs, *Xist* (X-inactive specific transcript), plays a central role in the induction of X-chromosome inactivation in female mammals (Loda and Heard 2019). lncRNAs can also have pivotal roles in scaffolding nuclear condensates as they participate in their assembly and function (Banani et al. 2017). A well-known example is the lncRNA *metastasis-associated lung adenocarcinoma transcript 1* (*MALAT1*), which is localized in nuclear speckles and acts in pre-mRNA splicing and transcription, regulating cancer progression and metastasis (Fei et al. 2017). Additional roles of lncRNAs in post-transcriptional regulation include mRNA splicing and translation interference, mRNA turnover regulation, modulation of signaling pathways, sponging miRNAs, and functioning as competitive endogenous RNAs (Statello et al. 2021). Lastly,

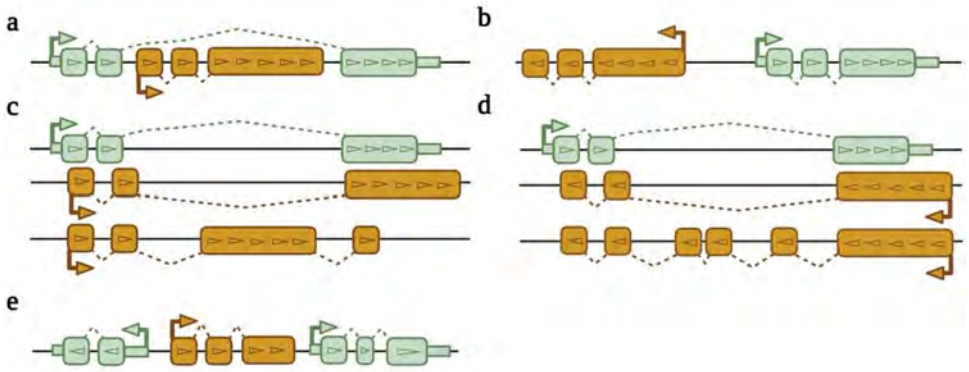


Fig. 1 Summary of different lncRNA classes. **a** Intronic long non-coding RNAs (ilncRNAs). **b** Bidirectional long non-coding RNAs (blncRNAs). Both transcriptional directions exemplified by the arrows. **c** Sense long non-coding RNAs (slncRNAs). **d** Antisense long non-coding RNAs (alncRNAs). **e** Long intergenic non-coding RNAs (lincRNAs). The brown boxes represent lncRNA exons, while the green boxes represent protein-coding exons with the UTR regions represented by the green rectangles. The dashed lines represent the splicing of intronic regions of all genes. The arrows within the exon boxes represent the transcription orientation of each gene. The arrows outside the boxes represent the transcription start site of each gene. Figure created with BioRender.com

lncRNAs have been described in specific organelles such as exosomes and mitochondria (Fatima and Nawaz 2017). For example, *SAMMSON*, the *survival associated mitochondrial melanoma-specific oncogenic non-coding RNA*, is a nuclear-encoded lncRNA that controls mitochondrial homeostasis by governing mitochondrial 16 s ribosomal RNA maturation and expression of polypeptides encoded by the mitochondrial DNA (Leucci et al. 2016; Vendramin et al. 2018). Granting all this, lncRNAs identification and function exploration are still commencing, thus warranting further attention.

lncRNAs as targets for treatment

The lncRNA characteristics make them very attractive for target-specific treatment of diseases. lncRNAs have a tissue-specific expression and are not conserved between distant species, compared with the protein-coding genes (Derrien et al. 2012; Ransohoff et al. 2018). Many studies have shown drugs that can regulate lncRNA expression in vertebrates and invertebrates (Wang et al. 2017; Smallegan and Rinn 2019; Amaral et al. 2020), which can indicate that targeting lncRNAs may be a novel therapy for human diseases (Jiang et al. 2019; Chen et al. 2021). As a matter of fact, RNA therapy is trending, where RNA sequences or RNA structures have been targeted by drugs or antisense oligonucleotides (ASO), helping the cure of diseases such as muscular dystrophy, cancer and hepatitis C infection (Young et al. 2010; Palacino et al. 2015; Haga et al. 2015; Sivaramakrishnan et al. 2017; Kim 2020; Wang et al. 2020a; Shao and Zhang 2020; Ferlini et al. 2021).

Parasites of human medical importance

Parasites refer to organisms that are physiologically dependent upon their host for survival (Baron 1996). The three main classes of human parasites are protozoa, helminths and ectoparasites. Most parasitic diseases occur in both tropical and subtropical areas. Out of all parasitic diseases, malaria is the one that kills the most, and over 400,000 people die every year worldwide (WHO Team: Global Malaria Programme 2020). Most parasitic diseases are among the Neglected Tropical Diseases (NTDs), which suffer from a lack of attention and funding by the public health systems. NTDs affect more than 2 billion people globally, mostly in rural areas of low-income countries (Stolk et al. 2016). Up to 500,000 deaths are attributed to NTDs annually. Despite being endemic mainly in low and middle-income countries, their occurrence has also been rising in high-income countries (Stolk et al. 2016).

Some of the parasitic diseases are treated by a single oral dose of an antiparasitic drug. Drugs such as Praziquantel have been used to treat entire populations to control schistosomiasis (Ross et al. 2015; Eisele 2019). However, this approach is costly and does not prevent both children and adults from re-infecting themselves in endemic areas (Rosenberg et al. 2016). Nowadays, a rising concern is the appearance of drug-resistant parasites due to mass drug administration (Wang et al. 2012; Zuber and Takala-Harrison 2018). Therefore, novel therapeutic targets are needed to control parasitic diseases (Lothstein and Gause 2021; Gupta et al. 2021) and understanding the lncRNAs' roles in these parasites is a matter of great relevance. In this review, lncRNAs from protozoan and helminth parasites will be addressed, paving the way for novel discoveries linking lncRNA and parasitosis treatment.

lncRNAs in protozoan parasites

Protozoan parasites are defined as the group of single-celled eukaryotes that live in a symbiotic relationship with their host (Baron 1996). The major groups of Protozoa include Mastigophora, Sarcodina, Sporozoa and Ciliophora, being the most common examples of each group *Giardia lambia*, *Entamoeba histolytica*, *Plasmodium falciparum*, and *Balantidium coli*, respectively. Several protozoan parasites infect human hosts, causing diseases such as Malaria, Amoebiasis, Giardiasis, Toxoplasmosis, Cryptosporidiosis, Trichomoniasis, Chagas disease, Leishmaniasis, African trypanosomiasis, Acanthamoeba keratitis, and Primary amoebic meningoencephalitis. The source of contamination goes from contaminated water and food exposure to arthropods' bite and sexual transmission. Some of these diseases, such as Amoebiasis, Giardiasis, and others, have an effective treatment for some of the parasite stages inside the human host (Zucca and Savoia 2011).

On the other hand, other diseases still rely on diminishing their symptoms and not exterminating the parasite (Barrett et al. 2019). In addition, reports of parasite resistance to treatment have been published and the need to search for novel targets is recurrent (Gupta et al. 2021; Walters and Temesvari 2021). Up to today, little information about lncRNAs in protozoan parasites has been acquired, leaving room for further exploration. Previously published information regarding lncRNAs in human protozoan parasites is reviewed below, and a summary can be found in Table 1.

Apicomplexan

The Apicomplexan phylum is comprised of parasites from the protist clade. These parasites are responsible for many debilitating diseases such as Malaria, Toxoplasmosis,

Table 1 Summarized information on lncRNAs of protozoan parasites of medical importance

Organism	Reference	Year	Stages	Number of lncRNAs	Method of lncRNA validation*
Cryptosporidium parvum	Li et al	2021	0, 2, 24, and 48 h intracellular and extracellular oocyst	396	RT-qPCR
Entamoeba histolytica	Saha et al	2016	Trophozoites	No mention	RT-qPCR and Northern blotting
Leishmania braziliensis	Ruy et al	2019	Procyclic, metacyclic and axenic amastigote	7351	RT-qPCR and Northern blotting
Leishmania infantum	Dumas et al	2006	Amastigotes and promastigotes	No mention	Northern blotting and FISH
Leishmania major	Rastrojo et al	2013	Promastigote	1884	Not validated
Leishmania major	Pawar et al	2017	Promastigote	1876	LC–MS
Plasmodium falciparum	Yang et al	2021	All stages of parasites in an intraerythrocytic lifecycle	3623	Not validated
Trichomonas vaginalis	Woehle et al	2014	Trophozoites	2175	RT-PCR
Trypanosoma brucei	Kolev et al	2010	Procyclic	103	Not validated
Trypanosoma brucei	Chikne et al	2017	Procyclic	103	Northern blotting
Trypanosoma brucei	Guegan et al	2020	Procyclic and metacyclic	1428	RT-qPCR

**RT-qPCR*, reverse transcription quantitative PCR; *LC–MS*, liquid chromatography–mass spectrometry

Cryptosporidiosis and Cyclosporiasis. lncRNAs from the Apicomplexan phylum have already been revised elsewhere (Li et al. 2020). Recently, newly published work identified hundreds of lncRNAs by differential expression analysis across the life cycle of *Cryptosporidium parvum*, a parasite that promotes diarrhea as the most common symptom (Li et al. 2021). The authors identified 396 novel lncRNAs, most of them being alncRNAs. Approximately 86% of all lncRNAs were differentially expressed across the different life cycle stages analyzed and surprisingly, nearly 10% of the mRNAs have an antisense transcript (Li et al. 2021). Also, they found that most of the lncRNAs occur near the 3′ end of their corresponding mRNA and seem to be transcribed by bidirectional promoters. An interesting finding was the evolutionary conservation and expression of lncRNAs between different species of the *Cryptosporidium* genus (Li et al. 2021), which in principle would facilitate multi parasitic therapeutic strategies. Furthermore, Yang et al., 2021 combined RNA-Seq data using the PacBio platform with proteomics analyses to unravel 3623 novel lncRNAs out of 12,553 potential lncRNAs in the malaria parasite *Plasmodium falciparum* (Yang et al. 2021). Among the 3623 newly identified lncRNAs, 56% were classified as alncRNAs (2023), 15% as slncRNAs (529) and 29% as lincRNAs (1071). Interestingly, proteomics analysis identified 160 small proteins (< 100 amino acids) that are translated from 160 different lncRNA transcripts. That analysis has gathered a coverage of the predicted peptides ranging from 8.5 to 100%, with molecular weight mass from 1.1 to 11.8 kDa. Altogether, Yang et al., 2021 have paved the way for the discovery of small peptides arising from translated lncRNAs in *P. falciparum*, which may possess new biological functions as previously reported in other species (Kastenmayer 2006; Sberro et al. 2019; Fesenko et al. 2019; Martinez et al. 2020).

Entamoeba

Entamoeba histolytica is a protozoan intestinal parasite that causes amoebiasis in humans (Jeelani and Nozaki 2014). This disease affects 30–50 million people every year worldwide, with an annual mortality rate of 40,000–100,000 (Shirley et al. 2018; Ben Ayed and Sabbahi 2019). In some cases (approximately 10%), infected individuals can develop invasive intestinal and/or extraintestinal (liver, lungs, and brain) amoebiasis, developing mild to fatal symptoms that can include loose feces, stomach pain, and stomach cramping. Among parasitic diseases, *E. histolytica* is one of the leading causes of morbidity and mortality in developing countries (Shirley et al. 2018). Like all amoebas, their transmission occurs by ingestion of cysts in fecally contaminated food, water, or hands. Due to their protective walls, the cysts can remain viable for several weeks in external environments (Blessmann et al. 2003). The standard compound for treating amebiasis is metronidazole (MTZ); however, it causes adverse effects, such as diarrhea, metallic flavor, and nausea due to the doses and long-term treatment (Stanley 2003). In addition, MTZ can cross the blood–brain barrier causing cerebellar toxicity (Agarwal et al. 2016), and therefore, the development of an alternative treatment and the discovery of new therapeutic targets are necessary.

E. histolytica genome was published in 2005 (Loftus et al. 2005). A new assembly and reannotation were published in 2010 (Lorenzi et al. 2010), but still little is known about lncRNAs in this organism. In 2005, Bhattacharya et al. sequenced a 10 kb genome region of 14 different parasite

strains isolated at different years (Bhattacharya et al. 2005). They compared the Single Nucleotide Polymorphisms (SNPs) from coding and non-coding genomic regions and found that the occurrence of SNPs in each of these regions was significantly different (0.07% in coding and 0.37% in non-coding regions). Although the non-coding regions contemplated by this work do not necessarily encode transcribed lncRNAs, most of them are located at intronic and upstream regions of coding genes, indicating that non-coding regions may be better gauges to assess evolutionary trends in *E. histolytica* (Bhattacharya et al. 2005).

More recently, Saha et al. (2016) described EhslncRNA, a lncRNA that partially overlaps a B1 transmembrane kinase family protein-coding gene in *E. histolytica* (Saha et al. 2016). They found that EhslncRNA encodes a 2.6 kb transcript with no ORF longer than 150 bp. The lncRNA transcript is polyadenylated and mostly associated with monosomes in the cytoplasm of *E. histolytica* under serum starvation in vitro culturing. In contrast, in normal conditions, EhslncRNA is present mainly in the nucleus. Interestingly, the transcript expression level is upregulated 1.7 to 2.7 times when serum starvation, oxygen and heat stresses are applied to the cultured parasites. Their observations suggest that EhslncRNA is a long non-coding RNA that may help *E. histolytica* cells cope with stressful conditions inside their host (Saha et al. 2016), representing a novel potential target.

Trichomonas

Trichomonas vaginalis is the causative parasite of trichomoniasis, one of the most common sexually transmitted infections that affect over 220 million people worldwide (World Health Organization (WHO) 2019). Although most of the infected patients do not present any symptoms, they can include itching, burning, redness or soreness of the genitals with discomfort during urination. *T. vaginalis* is an anaerobic protist that can shift between its amoeboid and flagellated stages very quickly (Kusdian et al. 2013). This parasite has up to 60,000 protein-coding genes encoded on six chromosomes (Carlton et al. 2007), emerging as one of the organisms with the highest known coding capacity (Smith and Johnson 2011). Trichomoniasis treatment is based on metronidazole; however, evidence of *T. vaginalis* resistance to this drug has been extensively reported (Meri et al. 2000; Schwebke and Barrientes 2006; Kirkcaldy et al. 2012; Paulish-Miller et al. 2014).

LncRNAs have not been thoroughly investigated in *T. vaginalis* despite its larger genome size (Carlton et al. 2007) compared with other protists, which in principle could harbor space for lncRNA transcription. The only report on this subject is a large-scale study in which the authors found, based on previous transcriptomic data and newly generated 271 million RNA-Seq reads, that thousands of pseudogenes and lncRNAs are present in *T. vaginalis* (Woehle et al. 2014). Interestingly, about 20% of the transcripts mapped in non-protein-coding genomic loci and 2175 lncRNAs were described, from which 233 were classified as convergent, 434 as divergent, 329 as co-oriented and 334 as anti-oriented according to their neighboring protein-coding gene, totalizing 1330 lincRNAs. The remaining 845 lncRNAs were intragenic lncRNAs that were not classified according to their orientation from the gene in the same locus. Another interesting finding was that nearly half of the lncRNAs expressed are pseudogenes, which agrees with the dynamic nature of the *Trichomonas* genome, from which gene duplication events are frequently found in parts of the genome and gene families (Woehle et al. 2014).

Trypanosomatidae

Trypanosomatids are a group of kinetoplastidean parasites that are characterized by the corkscrew-like motion of their single flagellum (Kaufer et al. 2017). Another feature exclusive to these parasites is a periflagellar structure known as kinetoplast, the mitochondrial DNA in complex condensed DNA circles (Shapiro and Englund 1995). Most trypanosomatids infect insects, but some have life cycles involving a secondary host, which can be a vertebrate, invertebrate or plants (Podlipaev 2001). Several species cause significant diseases in humans, amongst them the African Trypanosomiasis (caused by *Trypanosoma brucei*), Chagas diseases (caused by *Trypanosoma cruzi*) and Leishmaniasis (caused by different species of the *Leishmania* genus) (Simpson et al. 2006). Chemotherapy is based on nitro compounds for Chagas disease (Molina et al. 2014; Moraes et al. 2015), melarsoprol and nifurtimox for African Trypanosomiasis (Steverding 2010) and pentavalent antimonials for Leishmaniasis (Sundar and Chakravarty 2015), but all compounds are costly and highly toxic (Menna-Barreto 2019). Furthermore, drug resistance has been already reported, emphasizing the need for novel treatments (Olliaro 2010; Barrett et al. 2011; Urbina 2015).

Tens of thousands of RNA-Seq data from Trypanosomatidae are currently deposited in the Sequence Read Archive (SRA) databank, but little is known about lncRNA functions in these organisms, and the papers that describe them are reviewed below.

The first report on a novel class of non-coding RNAs in *Leishmania* was published in 2006 by Dumas and collaborators (Dumas et al. 2006). In that work, the authors described a novel class of developmentally regulated non-coding RNAs in *L. infantum*, using as a screening probe radiolabeled cDNAs obtained from total RNA of promastigote and amastigote life cycle stages. This was the first attempt at showing a 274 nucleotide RNA sequence that is transcribed from *L. infantum* genomic clusters of tandem head-to-tail

repeats, primarily present in subtelomeric regions and having no homology with other eukaryotic non-coding RNA. Among the molecule properties, they found that the 272 nt RNA is transcribed by RNA Pol II, has a 3′-end processing by polyadenylation, is located at the parasite's cytosol and has a potential association with small ribonucleoprotein complexes (Dumas et al. 2006).

The first comprehensive transcriptomic analysis of a parasite from the genus *Leishmania* was performed by Rastrojo and collaborators, who analyzed Illumina RNA-Seq transcriptomic data from promastigote *Leishmania major* (Rastrojo et al. 2013). In that work, they described 10,285 transcripts from 14 million mapped reads, of which 1884 novel transcripts were found in genic regions lacking annotated genes. Hence, they were categorized as non-annotated genes (Rastrojo et al. 2013). Later, interestingly, Pawar and collaborators used previous peptide data from mass spectrometry of *L. major* (Pawar et al. 2017) to search against the three ncRNA translated frames database retrieved from the 1884 transcripts described by Rastrojo and collaborators (Rastrojo et al. 2013). They found that a tiny portion of those transcripts (only eight transcripts, 0.42%) had matches against their proteomic data, meaning that almost all of those intergenic new transcripts likely represent new lincRNAs (Pawar et al. 2017).

The same approach was applied to *Leishmania braziliensis*, for which an in-depth study on gene expression across the life cycle stages covering coding and non-coding RNAs was presented (Ruy et al. 2019). The authors have used RNA-Seq data from procyclic, metacyclic and axenic amastigote parasites gathering a total of 677 million reads, of which around 98% were mapped reads. These analyses have come to a total of 9269 protein-coding transcripts and 7351 long non-coding RNA transcripts. From the total number of lncRNAs, 4683 were assigned as lincRNAs, 2334 as slncRNAs and 334 as alncRNAs (Ruy et al. 2019). Most of the differentially expressed lncRNAs were found between the parasite's insect and mammalian proliferative stages, and 295 lncRNAs were differentially expressed in all three stages. Their expression profiles were different from their neighboring protein-coding genes, and amongst the 35 ncRNAs tested, 22 ncRNAs were validated by northern blotting. Overall, this work presented an overview of the *L. braziliensis* transcriptomic profile and its adjustments throughout development, including coding and non-coding genes (Ruy et al. 2019).

A single-nucleotide resolution genomic map of the *T. brucei* transcriptome was published (Kolev et al. 2010). The number of reads retrieved ranged from 7 to 12 million, of which 5.8–10.7 million reads were mapped. This was the second attempt at showing a better transcriptome sequencing and annotation of *T. brucei,* in which the authors were able to describe a new set of 1114 transcripts, of which 103 transcripts were described as non-coding RNAs. These 103 transcripts were confidently set as lncRNAs due to their transcript length ranging from 204 to 2229 nt, due to the absence of ORFs longer than 25 amino acids, and because the small peptides found from the predicted ORFs do not align with previously published proteomic data (Kolev et al. 2010). No further report on their function was mentioned, but this was the first time a set of lncRNAs was reported in *T. brucei* parasites (Kolev et al. 2010).

RNA poly (A) polymerases (PAP) are essential for adding tracts of adenosine residues to the 3′ end of precursor RNAs (Koch et al. 2016). Those proteins are essential for many organisms since their function denotes RNA stability and turnover (Di Giammartino and Manley 2014). Trypanosomatidae parasites possess two canonical PAPs in which one of them (PAP1) is in the nucleus and polyadenylates RNAs that undergo trans-splicing and polyadenylation. Most of the PAP1 substrates are small nucleolar RNAs and long non-coding RNAs (Chikne et al. 2017). 63 out of the 103 lncRNAs identified by Kolev and collaborators (Kolev et al. 2010) were 1.5 times fold upregulated when PAP1 gene silencing in *T. brucei* was performed, and surprisingly no lncRNAs were down-regulated (Chikne et al. 2017). Another interesting finding was that snoRNAs and lncRNAs are differentially affected by PAP1 gene silencing, compared with mRNAs, suggesting that PAP1 is responsible primarily for ncRNA processing (Chikne et al. 2017).

Another recent work, accessible as a pre-print, has used RNA-Seq to measure whether lncRNAs are involved in *T. brucei* differentiation (Guegan et al. 2020). The work used Illumina next-generation paired-end sequencing from procyclic and metacyclic stages and identifyied 1428 lncRNAs scattered through the 11 chromosomes of the *T. brucei* genome. Out of the 1428 lncRNAs, 357 were differentially expressed between the procyclic and metacyclic stages of *T. brucei*. At the same time, from the 9598 protein-coding genes, 1536 were differentially expressed (Guegan et al. 2020). Their main focus was to understand lncRNAs that participate in parasite density sensing pathways via the stumpy induction factor (SIF) and SIF signaling (Vassella et al. 1997). This pathway is responsible for gene expression, morphological, and metabolic changes associated with the stumpy form (a non-dividing quiescent form of the parasite) (Mony et al. 2014), which is mainly controlled by RNA binding proteins RBP7A and RBP7B. Since RBP7A/B are the core functioning of the SIF pathway, the authors first looked for lncRNAs in the neighboring genomic regions of those genes and found *grumpy*, a lincRNA located upstream of RBP7A and RBP7B (Guegan et al. 2020). *Grumpy* gene expression pattern followed its neighboring gene RBP7A/B. This lincRNA does not interact with *T. brucei* ribosomes, no peptide could be detected corresponding to its small ORF; its over-expression led to a premature parasite differentiation

to the quiescent stumpy form, which also impaired in vivo infection. The analyses pointed *grumpy* as the first lincRNA characterized in *T. brucei*, which modulates parasite-host interactions (Guegan et al. 2020).

lncRNAs in helminths

Helminth is a general term to describe worms. They are elongated invertebrates with flat or round bodies. Flatworms (or platyhelminths) include flukes and tapeworms, while roundworms (or nematodes) include hookworms, whipworms and filariae (Baron 1996). These organisms develop through egg, larval (juvenile), and adult stages, sometimes depending on different hosts for their development. They cause a wide range of diseases, including nematodiasis, cestodiasis and trematodiasis, being Ascariasis, Taeniasis, and Schistosomiasis the most common infections. The most usual mode of infection is through ingestion of contaminated vegetables, water, and raw or undercooked meat. While some of the diseases still lack treatment, most of them have at least one option (Hotez et al. 2008). However, many reports of helminths resistant to drugs have been published (Geerts and Gryseels 2000; Albonico et al. 2004; Prichard 2007; Prichard and Roulet 2007; Hotez et al. 2007) and the need to search for novel targets is of great importance (Lothstein and Gause 2021). Although worms englobe a wide range of parasites, little information about lncRNAs or even transcriptomic studies from those parasites have been performed. Parasites from the genus *Schistosoma* are the most studied regarding lncRNA identification, likely due to their spread occurrence and difficulties in eradicating schistosomiasis. Below, we provide detailed information on the studies previously published on lncRNAs in human helminth parasites, and a summary is shown in Table 2.

Echinococcus

Parasites from the genus *Echinococcus* are amongst the tapeworms that cause cystic and alveolar echinococcosis (Kumar et al. 2017). This disease may last for years and often has no symptoms; when present, they are usually associated with the cyst's location and size (Bharucha 2001). When the liver is affected, the symptoms can vary from abdominal pain, yellow-toned skin and weight loss, and when the lung is affected, it may cause chest pain, coughing and breath problems (Rinaldi 2014). This parasitosis is spread through egg-contaminated water or food that is ingested by the host. This disease occurs almost everywhere and is estimated to affect one million people (World Health Organization 2014). Most treatments rely on Albendazole after surgery for the removal of the parasite's cysts.

The first report on lncRNAs in *Echinococcus granulosus* has been recently published, with lncRNAs shown in

Table 2 Summarized information on lncRNAs of helminth parasites of medical importance*

Organism	Reference	Year	Stages	Number of lncRNAs	Method of lncRNA validation**
Echinococcus granulosus	Zhang et al	2020	Cyst	2662	Not validated
Schistosoma mansoni	Oliveira et al	2011	Eggs, miracidia, cercariae, schistosomula and adult worms	201	Not validated
Schistosoma mansoni	Vasconcelos et al	2017	Miracidia, sporocyst, cercariae, schistosomula, male and female adult worms and their tissues	7029	RT-qPCR
Schistosoma mansoni	Oliveira et al	2018	Adult worm couples	170	RT-qPCR
Schistosoma mansoni and Schistosoma japonicum	Liao et al	2018	Male and female adult worms	3247 for *S. mansoni* and 3033 for *S. japonicum*	Not validated
Schistosoma mansoni	Maciel et al	2019	Miracidia, sporocyst, cercariae, schistosomula, male and female adult worms and their tissues. Single-cell RNA-Seq	16,583	RT-qPCR
Schistosoma mansoni	Kim et al	2020	Sporocysts	4930	RT-qPCR
Schistosoma japonicum	Maciel et al	2020	Cercariae, sporocysts, schistosomula, young or mature adult males and females	12,291	Not validated
Schistosoma mansoni	Amaral et al	2020	Adult female worms	912	RT-qPCR
Schistosoma mansoni and Schistosoma haematobium	Sirekbasan et al	2021	Eggs, adult male and female	5132 for *S. mansoni* and 3589 for *S. haematobium*	Not validated

*No information on lncRNAs in nematodes of medical importance has been published (see Wang 2021);**_RT-qPCR_ reverse transcription-quantitative PCR

exosome-like vesicles (ELVs) (Zhang et al. 2020). ELVs are emerging as mediators of parasite and host interactions (Marcilla et al. 2014; Coakley et al. 2015), and parasite ELVs can transfer ncRNAs into host cells to control their gene expression (Garcia-Silva et al. 2014a, b; Buck et al. 2014). An RNA-Seq experiment was performed to analyze ncRNA expression (miRNAs, circRNAs and lncRNAs) in two different types of *E. granulosus* ELVs obtained from isolated protoscoleces (PSCs) and hydatid fluid (HF) of fertile sheep cysts (Zhang et al. 2020). 2361 lncRNAs were identified in the PSCs-ELVs, whereas 1254 lncRNAs were identified in the HF-ELVs, among which 1357 and 250 lncRNAs were uniquely found in PSC-ELVs and HF-ELVs, respectively. From the PSC-ELVs, nine lncRNAs were highly expressed, while 42 lncRNAs were considered differentially expressed compared to HF-ELVs (19 upregulated and 23 down-regulated). Another interesting finding was that 44 miRNA-lncRNA regulatory pairs were identified, including five miRNAs and 41 lncRNAs, demonstrating lncRNA-miRNA-mRNA intercommunications for the first time in this organism (Zhang et al. 2020). In conclusion, this was the first report on ncRNAs profiles in PSC–ELVs and HF–ELVs related to the host immunity and pathogenesis of *E. granulosus* (Zhang et al. 2020), providing numerous resources for further characterization of the regulatory potential of lncRNAs.

Schistosomatidae

Schistosomatidae is a group of parasitic trematodes with complex life cycles that include immature developmental stages in mollusk and adult stages occurring in vertebrates (Collins 2017). The parasites from the genus *Schistosoma* infect and cause schistosomiasis in humans (McManus et al. 2018). Schistosomatids are dioecious, which is an exceptional feature compared to other platyhelminthes, in which almost all species are hermaphroditic parasites that possess both male and female reproductive systems. The disease caused by schistosomes in humans is called schistosomiasis, with 252 million people infected and deaths ranging from 4400 to 200,000 per year worldwide (Gryseels et al. 2006; McManus et al. 2018). The symptoms include abdominal pain, diarrhea, bloody stool, and blood in the urine, leading to liver damage, kidney failure, infertility, and bladder cancer (McManus et al. 2018). The infection occurs when vertebrate hosts are exposed to contaminated water encountering larvae of the *Schistosoma* parasite. The only available drug for treatment of schistosomiasis is praziquantel (Bergquist et al. 2017). The search for novel therapeutics against *Schistosoma* parasites is needed because praziquantel is effective only against the adult stage of the parasite and because of reports on possible parasite resistance to treatment (Melman et al. 2009; Cioli et al. 2014; Vale et al. 2017).

The first study reporting ncRNAs in *Schistosoma mansoni* and *Schistosoma japonicum* performed a genome-wide annotation of housekeeping ncRNAs (tRNAs and rRNAs), snRNAs and miRNAs (Copeland et al. 2009). As for ncRNAs that are not housekeeping or miRNAs, the first work performed a re-analysis of public *S. mansoni* EST data, finding 303 antisense lncRNAs whose expression in *S. mansoni* was regulated in vitro by human TNF-α (Oliveira et al. 2011). The work found 104 lncRNAs differentially expressed across the eggs, miracidia, cercariae, seven-day-old in vitro-transformed schistosomula and adult stages (Oliveira et al. 2011).

Subsequently, Vasconcelos and collaborators gathered 88 public RNA-Seq data libraries from different tissues and life cycle stages of *S. mansoni* to identify lncRNAs in this parasite (Vasconcelos et al. 2017). This was the first attempt to categorize lncRNAs in *S. mansoni* using a rigorous pipeline that relied on different software for their adequate annotation. The authors identified 7029 lincRNA isoforms located on 2596 genomic loci (an average of 2.7 lincRNA isoforms per locus), and another set of 402 alncRNAs that are antisense to 268 protein-coding genes. Interestingly, evidence for lncRNA regulation was provided, with 40% of the lincRNAs presenting an H3K4me3 chromatin mark on their Transcription Start Site (TSS) surroundings, and 49% of the lincRNAs being conserved among *S. mansoni*, *S. japonicum* and *S. haematobium*. Also, they found that 181 lincRNAs are differentially expressed across five different developmental stages of the parasite, of which 16 lincRNAs were further validated by RT-qPCR (Vasconcelos et al. 2017). An expression correlation analysis of the lncRNAs to their protein-coding gene neighbors was performed, revealing that 2359 protein-coding genes were correlated to the 181 lincRNAs selected for this analysis, of which 8% exhibited a negative correlation while 92% exhibited a positive expression correlation. The Gene Ontology (GO) terms that were more significantly enriched among the protein-coding genes co-expressed with the lincRNAs were RNA-dependent DNA replication/RNA-directed DNA polymerase activity and G-protein coupled receptor (GPCR) signaling pathway/GPCR activity, both being essential for *S. mansoni* biology (Vasconcelos et al. 2017, 2018).

A work published in 2018 (Oliveira et al. 2018) revealed 170 novel lncRNAs in *S. mansoni* (that were not identified by the Vasconcelos et al. 2017 analysis). The analysis performed by Oliveira et al., 2018 used one single RNA-Seq library of 7-week-old mixed sex adult worms comprised of 10.5 million reads. Furthermore, no analysis of the lncRNA properties such as location in the genome and orientation has been provided (i.e., classification as lincRNA, alncRNA or slncRNA). These 170 new *S. mansoni* lncRNAs have no synteny with human and mouse lncRNAs, and the protein-coding genes located closest to these

lncRNAs were enriched in GO terms related to metabolism and biosynthesis (Oliveira et al. 2018). Out of the 170 lncRNAs, 15 were validated by RT-qPCR and showed different expression levels. Three lncRNAs were differentially expressed between male and female sensitive or resistant to praziquantel (Oliveira et al. 2018). Therefore, this has been the first attempt to show lncRNAs responsive to parasite drug treatment (Oliveira et al. 2018), paving the way for novel studies.

Liao et al., 2018 applied a computational pipeline that had been previously used to identify potential lncRNAs in *P. falciparum,* to describe lncRNAs in *S. mansoni* and *S. japonicum* using RNA-Seq data (Liao et al. 2014). With these analyses, the authors found 3247 lncRNAs in *S. mansoni* and 3033 lncRNAs in *S. japonicum* (Liao et al. 2018). They analyzed H3K4me3 Chromatin Immuno-precipitation Sequencing (ChIP-Seq) libraries and found that in 12% of the cases, the H3K4me3 chromatin mark was enriched in regions proximal to lncRNA TSSs (Liao et al. 2018), a number four times lower than the one previously reported by Vasconcelos et al. (2017).

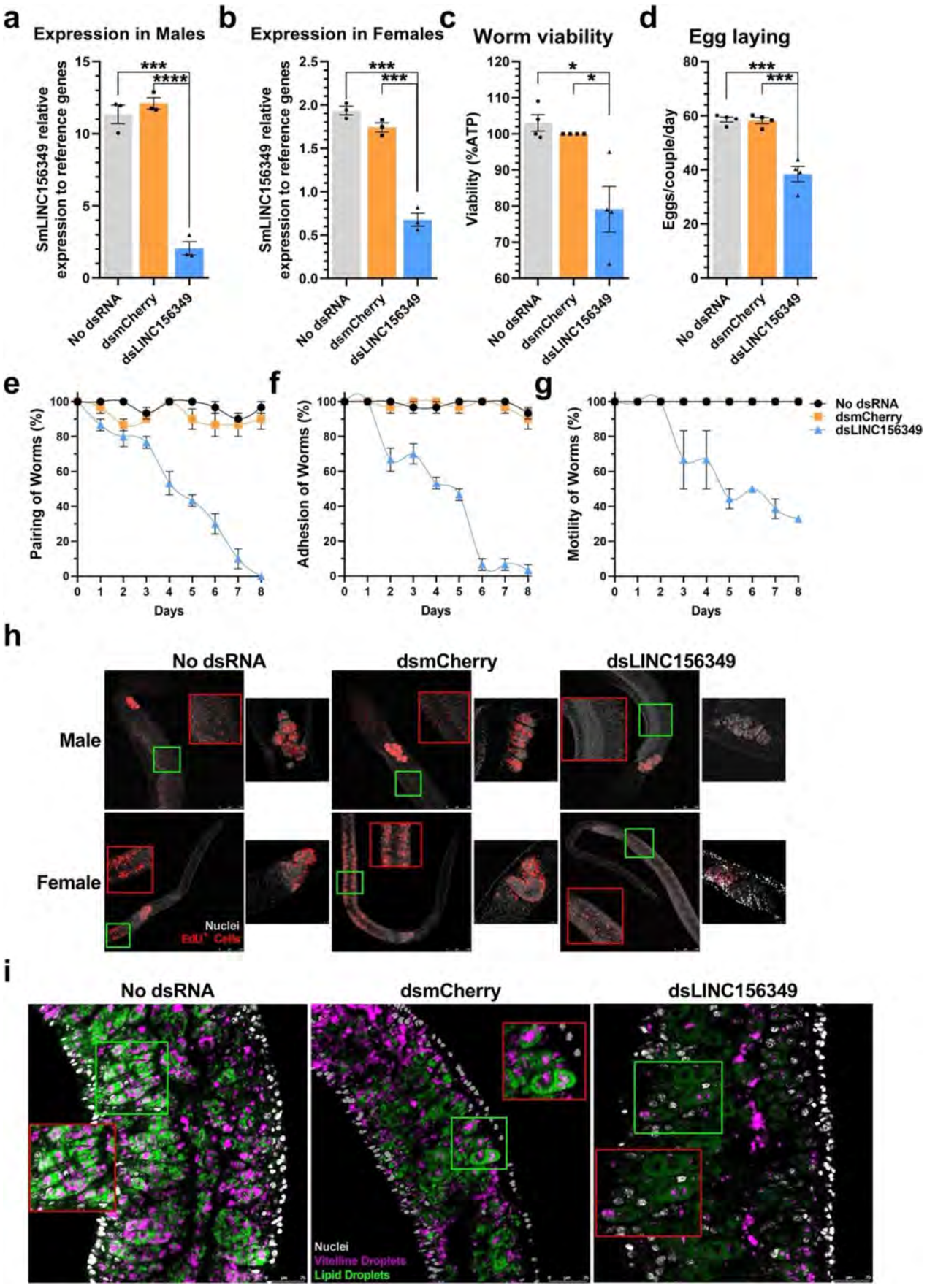
An improved lncRNA annotation and classification in *S. mansoni* was recently implemented (Maciel et al. 2019). The authors described 16,583 lncRNA transcripts originating from 10,024 *loci* (Maciel et al. 2019), of which 11,022 were novel *S. mansoni* lncRNA transcripts. In contrast, the remaining 5561 transcripts comprised 120 lncRNAs identical to and 5441 lncRNAs with transcript overlap with *S. mansoni* lncRNAs already reported in previous works (Vasconcelos et al. 2017; Liao et al. 2018; Oliveira et al. 2018). The analysis (Maciel et al. 2019) used 633 RNA-Seq libraries available from different *S. mansoni* life cycle stages (121), isolated tissues (24), cell populations (81) and single-cell experiments (407). The 16,583 lncRNA transcripts were divided into 7954 lincRNAs, 1191 slncRNAs and 7438 alncRNAs. One of the most important discoveries from Maciel et al. analysis was identifying and removing a set of 4293 lncRNA transcripts that were previously reported as lncRNAs, but are likely partially processed mRNAs with intron retention (Maciel et al. 2019). The majority of the confirmed lncRNAs (67%) were retrieved from RNA-Seq libraries from the parasite's life cycle stages (about 19% of the RNA-Seq libraries used). The lncRNAs described in that work have a putative ORF sizes distribution that is drastically different from mRNAs, a feature that can be used to document the likelihood that they represent *bona-fide* lncRNAs. Another interesting finding was that H3K4me3 and H3K27me3 chromatin marks are present around the TSSs of both protein-coding and lncRNA genes throughout the cercariae, schistosomula and adult life cycle stages of the parasite. Of note, lncRNAs have a lower abundance of histone marks than protein-coding genes (Maciel et al. 2019), which has also been reported for humans (Sati et al. 2012).

In addition, a weighted gene co-expression network analysis was performed (Maciel et al. 2019). The authors identified 15 different gene co-expression modules associated with the different life cycle stages or the adult worm gonads, which comprised hundreds to thousands of lncRNA-mRNA interaction pairs. A careful investigation of the most connected gene in the modules' networks has pointed lncRNAs as potential hub genes at the different life-cycle stages, adding another layer of relevance to these lncRNA molecules, which should be further considered for functional characterization (Maciel et al. 2019).

Kim and collaborators have analyzed by RNA-Seq the lncRNA expression of in vivo mature sporocyst samples of *S. mansoni* (Kim et al. 2020). Across development, sporocysts must integrate into the snail hepatopancreas, making it hard to distinguish and perform RNA-Seq experiments because of the snail and sporocyst RNA mixture. Thus, the authors were able to identify lncRNAs in this parasite stage using a high-fidelity pipeline that filters out snail host transcripts from the sporocysts and classifies lncRNAs and protein-coding genes. The authors found 4930 novel lncRNA transcripts from 3687 lncRNA genes, an average of 1.34 isoform per lncRNA (Kim et al. 2020). Among the 4930 lncRNA transcripts identified, 64% (3157) were expressed in adult worm stages (male and female), while 711 (14.4%) lncRNAs were exclusively expressed at the mature sporocyst stage (Kim et al. 2020).

When combining the Kim et al. 2020 dataset with the one previously reported (Maciel et al. 2019), the authors found that from the total 21,512 lncRNA transcripts, 4779 transcripts were sporocyst specific lncRNAs, of which only 14.9% (711 lncRNAs) were novel lncRNAs identified from in vivo sporocysts (Kim et al. 2020). Hundreds to thousands of lncRNAs were differentially expressed between the in vivo sporocysts, cercariae and adult worm stages. Some of these lncRNA expression levels correlated to protein-coding (or another lncRNA) expression. Thus, lncRNA importance has been highlighted, as those genes are expressed in a stage-specific manner and have co-expression patterns (Maciel et al. 2019; Kim et al. 2020) that suggest a role in the parasite's biology.

A re-analysis of *S. japonicum* RNA-Seq data has identified 12,291 lncRNAs transcripts from 7960 genes in *S. japonicum* (Maciel et al. 2020). The authors used 66 publicly available libraries, from different life-cycle stages (Maciel et al. 2020) and applied the same pipelines (Maciel and Verjovski-Almeida 2020) previously used for *S. mansoni*. An average of 1.5 isoform per gene was identified, and from the 12,291 lncRNAs, 6593 were classified as lincRNAs, 4694 as alncRNAs and 1004 as slncRNAs (Maciel et al. 2020). Sequence similarity search and syntenic conservation among *S. mansoni* and *S. japonicum* lncRNAs revealed that 14% of the lincRNAs were syntenic between the two species.


a
Expression in Males
SmLINC156349 relative expression to reference genes
No dsRNA
dsmCherry
dsLINC156349
b
Expression in Females
c
Worm viability
Viability (%ATP)
d
Egg laying
Eggs/couple/day
e
Pairing of Worms (%)
Days
f
Adhesion of Worms (%)
g
Motility of Worms (%)
h
Male
Female
Nuclei
EdU+ Cells
i
Nuclei
Vitelline Droplets
Lipid Droplets

◂**Fig. 2** *Schistosoma mansoni* adult worm phenotypes upon SmLINC156349 silencing. Ten adult worm pairs retrieved from perfusion of 42 days-infected Syrian hamsters were cultivated in 6-well plates with 5 mL of ABC media for eight days without dsRNA, with unrelated dsRNA targeting mCherry (control dsRNA) or with dsRNA targeting SmLINC156349, as indicated in each panel. SmLINC156349 expression knockdown was measured by RT-qPCR in **a** males or **b** females using cDNA generated from 4 µg of RNA extracted from parasites after eight days of dsRNA treatment. RT-qPCR primers for SmLINC156349 were: Fwd (5′-ACTGGGAAT CGTCGTTTGGA-3′) and Rev (5′-ACAGCCAGTTCGTTACCC AG-3′). Gene expression is relative to reference genes Smp_099690.1 and Smp_023150.1 that were selected among different reference genes. **c** Worm viability was evaluated by macerating ten worm couples and measuring the levels of ATP with the ATP-Glo Assay kit (Promega). The viability luminescence values were normalized against the dsmCherry control. **d** Female egg-laying was measured on the 8th day of in vitro cultivation by counting the total number of eggs and normalizing to the total worm couples in the plate well and to the days of cultivation. **e–g** Pairing, adhesion to the plate and motility of worms were observed every day and counted. Motility was evaluated based on a score (Horiuchi et al. 2005). **h** Adult worm bodies (larger images) and adult worm gonads (smaller images) are shown for males (upper row) and females (lower row), cultured for 8 days under the three different conditions indicated at the top of each panel, namely, control (no dsRNA), unrelated dsRNA control (dsmCherry), and dsRNA targeting SmLINC156349 (dsLINC156349). Cell proliferation was assayed by labeling with thymidine analog 5-ethynyl-2′-deoxyuridine (EdU), which was added to the cultures on the 7th day of cultivation at a final concentration of 10 µM and incubating for 24 h. EdU detection was performed as previously described (Collins III et al. 2013). DAPI (nuclei) cells are stained in gray and EdU + cells (proliferating cells) are stained in red. Scale bars: 250 µm for the larger, adult worm images, and 25 µm for the smaller, adult worm gonad images. **i** Female vitellaria staining with Fast Blue BB (pink) and BODIPY (green) labeling of vitelline and lipid droplets in the vitellaria, respectively, as previously described (Wang et al. 2019). DAPI staining is represented in gray. Scale bars: 25 µm. Mean ± SEM from at least three biological replicate experiments is shown. Unpaired two-tailed Student *t*-test with Welch's correction was used to calculate the statistical significance (*$p < 0.05$; ***$p < 0.001$; ****$p < 0.0001$). Representative images from 3 experiments with $n > 10$ parasites. The red borders define zoomed-in insets of interest that correspond to the regions within green borders

Co-expression analyses of lncRNAs and protein-coding genes revealed a dynamic relationship between the two RNA classes, being related to sexual maturation processes in both male and female *S. japonicum* adult worms (Maciel et al. 2020). The co-expression patterns indicate that lncRNAs may regulate these processes in *S. japonicum,* representing the first attempt to identify and perform a basic function prediction for lncRNAs in this parasite (Maciel et al. 2020).

LncRNAs have been associated for a long time with cancer epigenetics (Beckedorff et al. 2013). In *S. mansoni*, Amaral et al. evaluated the effect of epigenetic anticancer drug 5-Azacytidine (5-AzaC) (Stresemann and Lyko 2008) on the lncRNA gene expression by analyzing RNA-Seq samples of female adult worms treated or not with the drug (Amaral et al. 2020), a known ribonucleoside analogue compound that prevents *S. mansoni* DNA methylation (Geyer et al. 2011). In that work, the authors found 912 lncRNAs that were differentially expressed between female worms treated with 5-AzaC compared with the control (Amaral et al. 2020). Among the lncRNAs differentially expressed, 522 were lincRNAs (353 upregulated and 169 downregulated), 358 were alncRNAs (183 upregulated and 175 downregulated) and 32 were slncRNAs (16 upregulated and 16 downregulated). Another interesting finding was that most of the lncRNAs differentially expressed in female worms treated with 5-AzaC were associated with male metabolism co-expression modules (Amaral et al. 2020). In contrast, upregulated protein-coding genes were primarily associated with the juvenile co-expression module and the downregulated protein-coding genes with male metabolism (Amaral et al. 2020). Nearly half of the lncRNAs differentially expressed have at least one histone modification mark (H3K4me3 or H4K27me3) within 1 kb from their TSSs. Finally, some of those lncRNAs were differentially expressed among different life cycle stages of the parasite (Amaral et al. 2020).

The latest work on lncRNAs in *S. mansoni* and *S. haematobium* was published in February 2021. The authors used RNA-Seq data from eggs and adult males and females to unravel lncRNA transcriptomics (Sirekbasan and Gurkok Tan 2021). In *S. mansoni* the authors found 5132 expressed lncRNAs, while in *S. haematobium* 3589 lncRNAs were expressed (Sirekbasan and Gurkok Tan 2021). Unfortunately, the authors used the older reference transcriptome of Vasconcelos et al. 2017, instead of the updated version by Maciel et al. 2019, finding that from the 5132 lncRNAs detected in *S. mansoni*, they uniquely identified 32% (Sirekbasan and Gurkok Tan 2021); thus, a comparison between the lncRNAs identified by Sirekbasan and Gurkok Tan 2021 and the currently known set of *S. mansoni* lncRNAs (Maciel et al. 2019) is still lacking. Using a simple pipeline for the homology-based identification of lncRNAs, the authors identified 694 lncRNAs shared between *S. mansoni* and *S. haematobium* (Sirekbasan and Gurkok Tan 2021).

A knock-down approach for the characterization of a lincRNA function in *S. mansoni*

A long intergenic non-coding RNA, SmLINC156349, found to be upregulated upon 5-AzaC treatment in *S. mansoni* females but not in males (Amaral et al. 2020), has been selected in the present work to exemplify the initial steps towards its functional characterization. Because 5-AzaC is an epigenetic drug that inhibits *S. mansoni* oviposition and ovarian development (Geyer et al. 2011), we assumed that SmLINC156349 expression being affected by 5-AzaC treatment (Amaral et al. 2020), would make it a good lncRNA

candidate to be tested for its possible involvement with oviposition and ovarian development. Thus, we used the RNAi technique by soaking the adult *S. mansoni* couples in vitro for eight days with a double-stranded RNA (dsRNA) targeting SmLINC156349 (293 bp long, mapping to chromosome 7 at bases SM_V7_7:2,735,166–2,735,458, see genome browser at http://schistosoma.usp.br), and we then analyzed the phenotypes associated with the lncRNA knockdown.

Adult worms retrieved from infected Syrian hamsters were cultivated for eight days in ABC media (Wang et al. 2019) supplemented or not with the dsRNA targeting SmLINC156349 or mCherry (a control dsRNA that will activate the RNAi pathway but will not target any parasite gene). While dsRNA was added every day to the culture (to a final concentration of 30 µg/mL), the medium was exchanged every two days (70% of medium exchange). Worm pairing, adhesion, and motility were observed every day, while worm gene silencing, worm viability, worm proliferative cell status, female worm vitellaria status and egg-related phenotypes were observed only after the eight-day treatment.

The effective knockdown of SmLINC156349 in *S. mansoni* adult worms is demonstrated by RT-qPCR in Fig. 2, in which a reduction of 83% and 61% of SmLINC156349 expression was found in males and females, respectively, compared with the appropriate dsmCherry control (Fig. 2a and 2b). Note that a comparison of dsmCherry control with another control, namely the parasites with no dsRNA added to the medium, demonstrated that the dsmCherry control treatment did not disturb SmLINC156349 expression (Fig. 2a and 2b) and did not affect parasite viability, as shown by the phenotypic assays described below.

A reduction of 21% on adult worm viability measured by ATP quantification using the ATP-Glo-Assay Kit (Promega) was detected when the parasites were exposed to the dsRNA targeting SmLINC156349 (Fig. 2c). Female egg-laying was reduced by 34% when SmLINC156249 was silenced (Fig. 2d). As for the other phenotypes, we observed that pairing and adhesion to the plate were almost completely abolished in the parasites after eight days of SmLINC156349 dsRNA treatment, while the parasite's motility decreased 77% (Fig. 2e, f, and g). Adult worm stem cell proliferation and their gonad's cell proliferation were detected using a thymidine analog 5-ethynyl-2′-deoxyuridine (EdU) incorporation into newly replicated cells. The detection proceeded as previously reported (Collins III et al. 2013). As shown in Fig. 2h, male and female stem cell proliferation status was diminished once dsRNA targeting SmLINC156349 was added to the media. As for their gonads, the most proliferating organ in schistosomes, we detected a slight decrease in EdU labelling (Fig. 2h). Interestingly, the decrease in stem cell proliferation observed was more evident in male than in female adult worms, coherent with the basal expression levels of the studied lincRNA. Female vitellaria proliferation status was detected by double staining of vitelline and lipid droplets within the vitellaria, as previously reported (Wang et al. 2019). One can see that upon SmLINC156349 silencing the female vitellaria has been impacted mainly at the vitelline droplets production (pink staining) (Fig. 2i), corroborating with the egg-laying results (Fig. 2d). Both stem cell proliferation and vitellaria status assays demonstrated that dsmCherry control treatment did not disturb either of the biological processes.

We observed egg phenotypes upon SmLINC156349 silencing (Fig. 3). The eggs laid across the eight days of treatment were retrieved, and the area of eggs laid by females submitted to dsRNA targeting SmLINC156349 was significantly smaller than that of the eggs from control parasites (Fig. 3a). However, no difference in egg autofluorescence (i.e., the integrity of the eggshell) was observed when the dsRNA was added to the media (Fig. 3b). The egg cell number is inferred by DAPI staining, and a more developed egg is comprised of more cells stained with DAPI. As shown in Figs. 3c and 3d, the number of cells per egg and the $EdU^+/DAPI^+$ stained eggs ratio decreased significantly upon SmLINC156349 silencing. We observed a decrease in egg hatching in the eggs retrieved from the worm couples treated with dsRNA targeting SmLINC156349 for eight days (Fig. 3e), directly associated with decreased egg viability. A representative egg, each from a different field, obtained at each condition tested is shown in Fig. 3f, where different eggs used for measuring the area, autofluorescence (AUF), the number of cells per egg ($DAPI^+$) and the ratio of proliferating egg cells over total egg cells ($EdU^+/DAPI^+$) can be seen. This is the first evidence in *S. mansoni* of a lncRNA being essential for critical biological events of the parasite such as cell proliferation, female vitellaria development and female reproduction.

Interestingly, the expression level of SmLINC156349 in males is 6.7 times higher than in females (compare dsmCherry bars in Figs. 1a and 1b). Furthermore, by looking at previous gene co-expression network data provided by Maciel et al. (2019), we found that SmLINC156349 was assigned to the Turquoise module that characterizes the group of protein-coding and lncRNA genes expressed mainly in adult male worms compared with the other life cycle stages. Therefore, we hypothesize that the reduction in SmLINC156349 expression in males may have led to the phenotypes observed in females due to a secondary effect caused by the early unpairing that reached 50% upon 4 days of silencing. However, all phenotypes were measured on the 8th day of treatment, when pairing was entirely abolished, and most females presented abnormalities in their vitellaria and egg-laying. Thus, we can say that SmLINC156349 silencing impacts cell proliferation in male worm gonads,

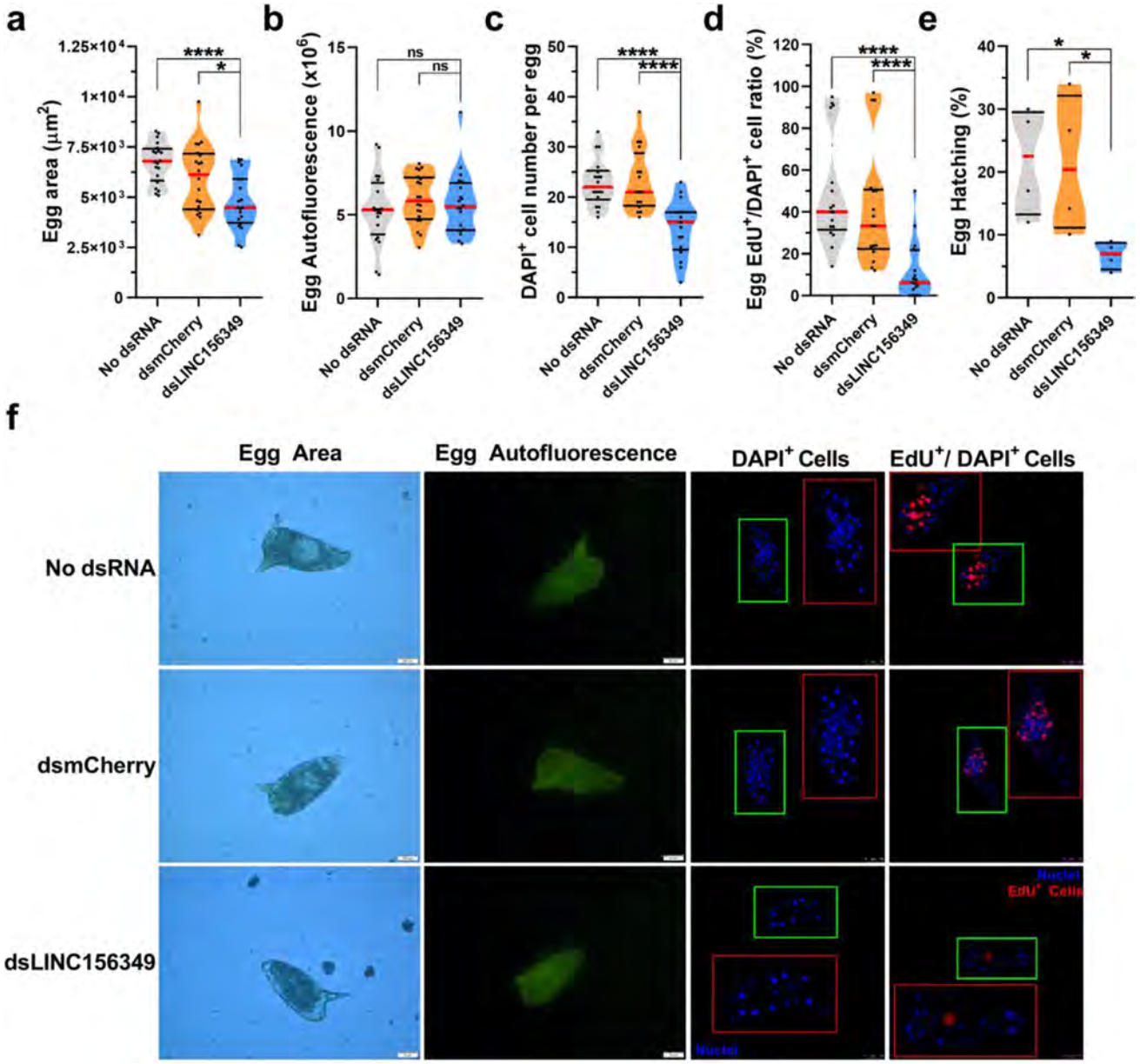


Fig. 3 *Schistosoma mansoni* egg phenotypes upon SmLINC156349 silencing. Adult worms retrieved from perfusion of 42 days-infected Syrian hamsters were cultivated in 6-well plates with 5 mL of ABC media for eight days without dsRNA, with dsRNA targeting mCherry (control dsRNA) or with dsRNA targeting SmLINC156349, as indicated in the graphs. Eight days after treatment, the eggs were counted and collected for further characterization. **a** Egg area, **b** autofluorescence (AUF), **c** total cell number, and **d** proliferating cells/total cell number ratio, were measured by using ImageJ. Egg autofluorescence is present because of a high concentration of phenolic proteins that form the eggshell (Ashton et al. 2001). For determining total cell number in **c**, the collected eggs were stained with DAPI (blue) for 4 h as previously described (Wang et al. 2019). For the EdU and DAPI labeling in **d**, EdU was added to the collected eggs and incubated in ABC medium for another day. After that, the eggs were processed and EdU (red) was measured as previously described (Collins III et al. 2013). **e** For eggs hatching, collected eggs were kept in ABC media for another seven days for the correct development of embryos. After that, the eggs were exposed to light at room temperature for 1 h and egg-hatching measured. In **f** representative images from 1 out of 3 experiments with $n > 20$ eggs. Note that a different egg in the field is shown at each of the different assays that are presented. Scale bars: 20 μm for egg area and autofluorescence (AUF) and 25 μm for DAPI + and EdU⁺/DAPI⁺ eggs. **a–e** Violin-plots with medium smoothing are represented, showing all points, median (red line) and quartiles (black lines) from three biological replicate experiments. Unpaired two-tailed Student *t*-test with Welch's correction was used to calculate the statistical significance ($*p < 0.05$; $****p < 0.0001$; $ns = p > 0.05$). Comparison of no dsRNA and dsmCherry showed no difference. The red borders define zoomed-in insets of interest that correspond to the regions within green borders

as evidenced by the EdU experiments, but a more robust male phenotype characterization could be further explored.

Box: new technologies for lncRNA functional characterization

The first step for lncRNA identification is creating a reference lncRNAs transcriptome for the studied parasite. For this, high-throughput RNA-Seq data must be processed, mapped to the genome and annotated. Standard pipelines for the accurate classification of lncRNAs should comprise steps for removing transcripts shorter than 200 nucleotides, mono-exonic transcripts, transcripts with exon-exon overlap with protein-coding genes from the same genomic strand, transcripts with coding potential, and finally transcripts with ORFs that match any other predicted protein. Maciel and Verjovski-Almeida 2020, provided an easy-to-follow step-by-step pipeline that might help biologists to start the discovery of new lncRNAs. The next step will be determining if these lncRNAs have biological functions. Most experimental approaches are challenging, especially for high-throughput studies, due to lncRNAs' diverse functions. A list of the technologies that can be used for the lncRNA functional characterization is provided below along with a summary figure of the basic principle for each method (Fig. 4), as an attempt to guide readers for novel discoveries on lncRNA function:

a *Basic lncRNA functional characterization*:

Single-cell RNA-Seq (scRNA-Seq): A technique that sequence individual cells from a multicellular organism providing evidence of cell-specific expression of lncRNAs (Nawy 2014).

lncRNA co-expression networks: This is the most used functional data for lncRNA. An approach that constructs networks of genes with a tendency to co-express in a group of samples, thus being possibly related (Horvath 2011).

Co-expression syntenies: Synteny is a genetic concept that dictates gene or genomic blocks in the genome being preserved across different species. Syntenic co-expression of genomic blocks suggests functional preservation across different species (Lee 2003).

In situ *hybridization (ISH)*: This approach relies on hybridization between a probe and the cellular lncRNA, documenting the tissue-specific or subcellular expression. Fluorescent (F) probes can be used for subcellular expression patterns, thus being called FISH. Techniques for whole-mount in situ hybridization (WISH) have been developed for some parasites (Young et al. 2020).

Ribosome profiling: A technique used as a global snapshot of all ribosomes actively translating RNAs in a cell. Having a translatable lncRNA is an indication that the studied lncRNA might not be an ncRNA and could be translated into small peptides; it should be noted that not all ribosome-bound RNAs are translated (Ingolia 2014).

b In-depth mechanistic exploration of lncRNAs based on protein-centric methods:

RNA immunoprecipitation (RIP): An antibody-based technique used to map in vivo RNA–protein interactions. The RNA binding protein (RBP) of interest is immunoprecipitated together with its associated RNA to identify bound transcripts by RT-qPCR, microarrays, or sequencing (Townley-Tilson 2006).

UV cross-linking and immunoprecipitation (CLIP): Similar to RIP, this approach has several variants based on an antibody, where the RBP is cross-linked to RNA/DNA with UV radiation. Importantly, covalent cross-linking permits stringent washing of immuno-precipitates and thus reduces background noise (Jensen and Darnell 2008).

In-depth mechanistic exploration of lncRNAs based on RNA-centricmethods:

ChIRP, CHART, RAP: All technologies are based on labeled oligonucleotides complementary to the lncRNA of interest as a way to pull down lncRNA-associated proteins and chromatin DNA. The differences among them rely on the oligo synthesis and size (Chu et al. 2011; Simon 2013; Engreitz et al. 2013).

RNA pull-down assay: Labeled lncRNA is synthesized and incubated with lysate or recombinant protein to form a specific lncRNA–protein complex. The protein complex is pulled down because of probe and resin interaction (biotin/streptavidin, for example) and identified with Western blotting or mass spectrometry (Torres et al. 2018).

MARGI, GRID-Seq, RADICL-Seq, and ChAR-Seq: All approaches aim to identify RNA–DNA interactions, differing from each other at the adapter ligation and sequencing steps (Sridhar et al. 2017; Li et al. 2017; Bell et al. 2018; Bonetti et al. 2020).

MARIO, LIGR-Seq, CLASH, SHAPE, SPLASH, PARIS: These approaches are all destined to detect RNA-RNA interactions between the same RNA, thus providing evidence of RNA secondary structure, or with another RNA like miRNAs, other lncRNAs or other ligands. The main difference between them is the fragmentation, proximity ligation and sequencing method (Calvet and Pederson

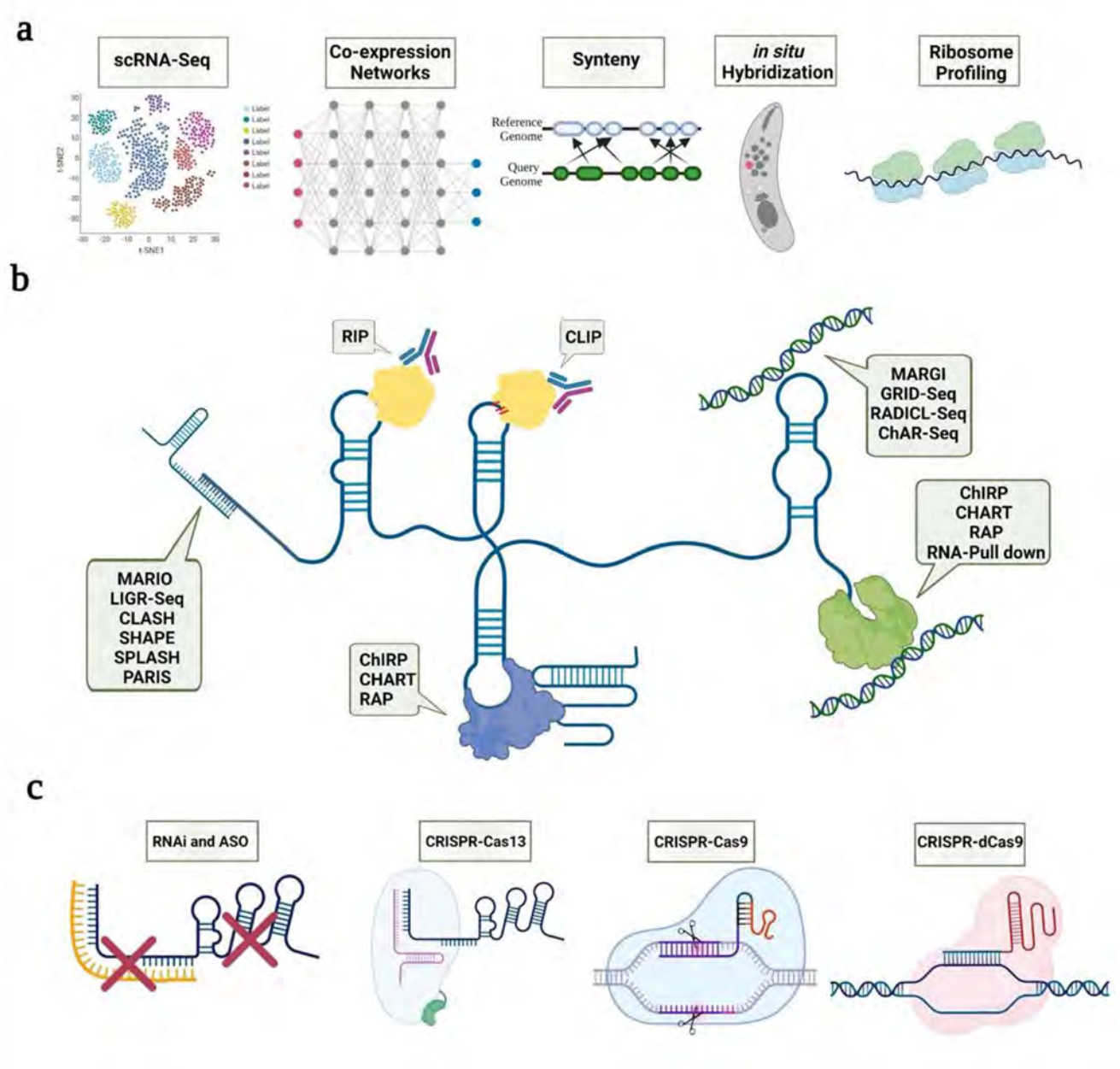


Fig. 4 Summary of different methods for lncRNA characterization. The different methods are evidenced within the light green call-out balloons. **a** The different basic functional characterization methods for lncRNA studies include single-cell RNA-Seq (scRNA-Seq), co-expression networks, synteny analyses, in situ hybridization and ribosome profiling. **b** The different in-depth mechanistic exploration methods for lncRNA function are separated based on the interactive nature that the method is detecting. RNA-RNA interactions are detected by MARIO, LIGR-Seq, CLASH, SHAPE, SPLASH, PARIS, and others. RNA–DNA interactions are detected by MARGI, GRID-Seq, RADICL-Seq, ChAR-Seq, and others. RNA–Protein-DNA interactions are detected by ChIRP, CHART, RAP, and RNA-Pull down. ChIRP, CHART, and RAP detect RNA–Protein-RNA interactions. RNA–protein interactions are detected by RIP and CLIP, differing from each other only because of the cross-linking (red lines) between RNA–protein in the CLIP method. The functional elements within the lncRNA (blue RNA molecule in the center) are represented, e.g., other RNA structures in blue, RNA-binding proteins in yellow, purple and green, DNA molecules in a double strand of blue and green colors. **c** The different biological exploration methods for lncRNA function prediction. The yellow single-stranded molecule represents siRNA or ASO; the lncRNA is represented by the blue single-stranded molecule with secondary structures. The red X represents RNA degradation. The double-stranded DNA is represented within the CRISPR-Cas complexes. Figure created with BioRender.com

1979; Loughrey et al. 2014; Helwak and Tollervey 2016; Aw et al. 2016; Sharma et al. 2016; Nguyen et al. 2016).

c. Biological exploration of lncRNAs:

RNA interference (RNAi): A non-strand specific approach used to down-regulate gene expression by using siRNAs or dsRNAs. This technique relies on the organism having RNAi machinery for the correct processing and down-regulation of the gene (Fire et al. 1998).
Antisense oligonucleotides (ASO): A strand-specific approach used to diminish gene expression levels. This approach targets RNAs that are active in the nucleus and do not require RNAi machinery to find their complementary gene. This method blocks ribosomes or other factors to bind the RNA and recruits RNase H to degrade the targeted RNA (Stephenson and Zamecnik 1978; Zamecnik and Stephenson 1978).
Over-expression of lncRNA: An approach not commonly used in multicellular organisms. It relies on cloning the lncRNA into a plasmid controlled by an endogenous promoter and integrating that plasmid into the desired cell or organism. Over-expression of the lncRNA should show a reversed phenotype compared with that of silencing the lncRNA.
CRISPR-Cas13: A CRISPR guided RNA silencing based on Cas13 and guide RNA recognition of cytoplasmic lncRNA (Abudayyeh et al. 2017).
CRISPR-Cas9: Genomic deletion of lncRNA gene body or promoter sequences. Since the deletion of gene expression is expected, harsher phenotypes could be noted. It is an evolving technology for multicellular organisms with different life cycle stages as is the case of most parasites (Jinek et al. 2012).
CRISPR-dCas9: A defective Cas9 is used for transcription initiation and elongation blocking on the desired lncRNA by using guide RNAs that target the promoter or gene body (Myers et al. 2018).

Challenges and limitations to the study of lncRNAs in parasites of human medical importance

Challenges and limitations are always present when studying molecules of unknown function. Particularly, studying lncRNA can present numerous challenges that novel developing technologies should overcome. The first challenge for lncRNA studies is lncRNA identification. Distinguishing lncRNA expression from transcriptional noise is a current and very discussed topic in the lncRNA field (Ponjavic et al. 2007; Churchman 2017; Yang and Meng 2019). Most lncRNA identification pipelines use a two-step process, in which first the transcripts are assembled and then the lncRNAs are annotated. Still, most parasites that cause diseases in humans lack information on lncRNA discovery, despite the vast RNA-Seq information already deposited in public databanks (Table 3). Therefore, using a rigorous and defined pipeline could help overcome the main challenge for lncRNA studies, preventing incorrect identification. A step-by-step bioinformatics protocol for discovering and analyzing lncRNAs in *S. mansoni* has been published to make it more accessible to biologists with no bioinformatics training (Maciel and Verjovski-Almeida 2020) and can be adapted to any organism of interest. The guide includes how to (1) download RNA-Seq libraries from the public database, (2) process and map reads against the genome, (3) reconstruct transcripts, (4) identify novel lncRNAs, (5) determine transcript expression level, and (6) identify differentially expressed lncRNAs (Maciel and Verjovski-Almeida 2020). This could be the starting point for researchers to look at lncRNA expression in RNA-Seq data, promoting a faster development of the field.

Another important feature that might help the correct discovery of lncRNAs is the adequate read length available in the different RNA-Seq datasets. While Illumina technology can sequence up to 300 bp-reads, PacBio and Oxford Nanopore can sequence an average of 10 kb per read. Although long reads provide more comprehensive information about the transcript's splicing annotation, it lacks sequence accuracy due to its high error rates (Pearman et al. 2020), and it still lacks larger sequencing depths. For adequate discovery and quantification of lowly expressed lncRNAs, between 10 and 40 million sequences per sample are required. So, combining longer reads for accurate transcript annotation with shorter reads for correct sequence quantification of reads should be the best approach for characterizing novel transcripts in any organism. Another fundamental limitation for lncRNA identification is the generally low expression level of those molecules, on average tenfold lower compared with the level of expression of protein-coding genes in a given cell, tissue or organism, making it hard for their discovery when using lower RNA-Seq coverages. The more sequencing depth in the experiment, the more information is acquired, so the development of cheaper RNA-Seq methods could help to overcome the barrier from which a deeper RNA-Seq is imposed.

The biggest challenge in lncRNA studies is function prediction. The most used method for initial lncRNA function prediction is the computational inference of a lncRNA being co-expressed with other genes of known biological functions. However, this can be a challenging effort for parasites with higher rates of non-annotated protein-coding genes. Sequence conservation among known lncRNAs may help to infer functionality for novel lncRNAs (Quinn et al. 2016;

Table 3 Protozoan and helminth RNA-Seq data available in the public domain to be explored for identification of novel lncRNAs. Numbers of RNA-Seq datasets available at the Sequence Read Archive (SRA) database at NCBI (https://trace.ncbi.nlm.nih.gov/Traces/sra/) for each indicated species are shown. Bio-Sample includes descriptive information about the biological source materials related to the experimental assays, while Bio-Project is a collection of biological data for a single initiative originating from a single organization or from a consortium. (a) Protozoan RNA-Seq information and (b) Helminths RNA-Seq information. All numbers reflect searches up to May 17th, 2021

a

Species	Number of Bio-projects	Number of Bio-samples
Acanthamoeba castellanii	7	92
Acanthamoeba pyriformis	1	6
Babesia bovis	8	36
Babesia divergens	4	22
Babesia microti	18	118
Criptosporidium parvum	28	133
Cryptosporidium baileyi	1	1
Cryptosporidium hominis	1	4
Dientamoeba fragilis	1	1
Entamoeba gingivalis	1	12
Entamoeba histolytica	33	226
Entamoeba invadens	6	28
Giardia intestinalis	20	262
Giardia lamblia	1	6
Leishmania amazonensis	5	25
Leishmania braziliensis	20	377
Leishmania donovani	53	1367
Leishmania guyanensis	5	18
Leishmania infantum	39	633
Leishmania major	67	240
Leishmania mexicana	12	178
Leishmania panamensis	7	40
Leishmania tarentolae	7	45
Leishmania tropica	44	293
Plasmodium berghei	69	7438
Plasmodium chabaudi	21	650
Plasmodium cynomolgi	5	26
Plasmodium falciparum	457	71,822
Plasmodium knowlesi	17	549
Plasmodium vinckei	6	40
Plasmodium vivax	219	7509
Plasmodium yoelii	22	111
Toxoplasma gondii	137	1093
Trichomonas tenax	1	4
Trichomonas vaginalis	13	189
Trypanosoma brucei	244	2756
Trypanosoma cruzi	64	612

b

Species	Number of Bio-projects	Number of Bio-samples
Ancylostoma caninum	5	90
Ancylostoma ceylanicum	3	97
Ascaris suum	13	336
Brugia malayi	29	309
Echinococcus granulosus	18	79
Echinococcus multilocularis	10	155
Fasciola gigantica	4	29
Fasciola hepatica	14	58
Hymenolepis diminuta	8	79
Hymenolepis microstoma	4	26
Necator americanus	7	117
Onchocerca colculus	10	269
Schistosoma haematobium	9	141
Schistosoma japonicum	28	433
Schistosoma mansoni	79	2715
Schistosoma mekongi	2	3
Schistosoma turkestanicum	3	3
Strongyloides papillosus	6	64
Strongyloides ratti	9	328
Strongyloides stercoralis	10	128
Strongyloides venezuelensis	5	14
Taenia asiatica	1	10
Taenia multiceps	4	4
Taenia pisiformis	2	9
Taenia saginata	2	3
Taenia solium	1	3
Trichuris muris	9	112
Trichuris suis	4	24
Trichuris trichiura	1	1

Rivas et al. 2017; Wang et al. 2020b). Nevertheless, lncRNA sequence divergency has been reported, being this, the most relevant characteristic that makes lncRNA so promising for target-specific treatments (Amaral et al. 2018; Tavares et al. 2019). Alternatively, lncRNA folding conservation may provide reliable information on lncRNA function (Diederichs 2014). Still, accurate prediction of lncRNA folding structures is challenging due to the high free energy that

those molecules have, allowing them to fold into millions of different secondary and tertiary structures under different environmental conditions (Blythe et al. 2016; Zampetaki et al. 2018; Chillón and Marcia 2020). A promising feature that has been studied and developed lately involves methodologies of differential labeling for characterizing lncRNA structures such as PARIS, LIGR-Seq and others (see Box) (Calvet and Pederson 1979; Sharma et al. 2016).

Biological function prediction is mainly based on knocking down or knocking out the lncRNA from the studied organism. The development of inducible RNAi or knocked-out organisms for unicellular parasites is an easier task than for multicellular parasites, a field which still needs further methodological progress. LncRNA location is another layer of complexity in knocking-down genes since most RNAi-based technologies work mainly on cytoplasmic transcripts. Therefore, novel technologies for biological function characterization such as CRISPR-based methods should help overcome these pitfalls.

Although lncRNA biogenesis mostly relies on RNA-Pol II transcription and RNA processing in the same way as the mRNAs, some reports on lncRNAs that differ from the canonical ones have been reported in higher eukaryotes (Dhanoa et al. 2018). Thus, identifying a given lncRNA biogenesis and subcellular expression pattern might help predicting the function of a lncRNA of interest. A challenge for lncRNA prediction is the access to a poorly annotated genome. The subcellular expression patterns can be observed by in situ hybridization experiments that depend on acquiring adequately labeled fluorescent riboprobes.

Traditionally, an ORF contains codons for at least 100 amino acids in eukaryotes and 50 amino acids in bacteria, ending with a stop codon that promotes ribosome disassembly and protein synthesis termination (Harrow et al. 2012). However, aplying these arbitrary in silico criteria to decide if an RNA is translated or not into a protein has led to the misannotation of some putative ncRNAs, which in fact might be translated into small proteins or micro peptides (Guttman et al. 2013; Fesenko et al. 2019; Martinez et al. 2020). Most lncRNAs are in the nucleus, while all translatable lncRNAs must be exported to the cytoplasm if they are to act in a dual function manner (Choi et al. 2019). When in the nucleus, it could be acting at transcription control, and when in the cytoplasm, it could be translated into a micro peptide that is crucial for some processes (Xing et al. 2021). Some techniques should help to overcome the doubt of whether the lncRNA is translated or not, such as ribosome profiling, mass spectrometry, proteomics, proteogenomics (Hartford and Lal 2020).

The future is pointing to single-cell RNA-Seq (scRNA-Seq) and CRISPR methodologies for the study of lncRNAs. scRNA-Seq aims to sequence individual cells isolated from a tissue, organ, or whole body. While some parasites have already had their protein-coding gene expression explored with scRNA-Seq (Howick et al. 2019; Wendt et al. 2020; Li et al. 2020; Diaz Soria et al. 2020; Nanes Sarfati et al. 2021), most of them still have plenty of room for the lncRNAs to be explored. In addition, CRISPR-based methodologies are bound to change the panoramic view of gene function due to their several uses that go from activation to inactivation of transcription, to complete deletion of gene function. A few approaches have been used in parasites of human health importance, representing just the tip of the iceberg of those strategies (McVeigh and Maule 2019; Bryant et al. 2019; Du et al. 2021). At last, the use of lncRNA as therapeutic targets for the treatment of parasitic diseases can be viewed as a much-desired aim, such as it has been proposed for other human diseases like cancer (Prabhakar et al. 2017; Matsui and Corey 2017; Blokhin et al. 2018; Wang et al. 2018; Elsayed et al. 2020; Yuan et al. 2020; Fathi Dizaji 2020; Chen et al. 2021).

Concluding remarks

lncRNAs play pivotal roles in humans and parasites, participating in developmental processes at the different life cycle stages. Little is known about lncRNA's roles in parasites, which points to the need of their identification and functional characterization. The starting point could be the use of available RNA-Seq data for lncRNA discovery. Whenever possible, a careful RNA-Seq design should be considered, bearing in mind that read length and depth are critical features for lncRNA discovery. Furthermore, customized data analyses should be applied to accurately identify new lncRNAs. Further genetic manipulation, targeting lncRNAs and their possible partners, is a crucial tool to decode the plentiful biological roles they can have.

Glossary

5-AzaC	5-azacytidine
alncRNAs	antisense lncRNAs
ASO	antisense oligonucleotides
AUF	autofluorescence
blncRNAs	bidirectional lncRNAs
ChAR-Seq	chromatin-associated RNA sequencing
CHART	capture hybridization analysis of RNA targets
ChIRP	chromatin isolation by RNA purification.
circRNAs	circular RNAs
CLASH	crosslinking, ligation, and sequencing of hybrids
CLIP	UV cross-linking and immunoprecipitation.
crasiRNAs	centromere repeat-associated short interacting RNAs

CRISPR-Cas clustered regularly interspaced short palindromic repeats – CRISPR associated protein
dsRNA double-stranded RNA
EdU 5-ethynyl-2′-deoxyuridine
ELVs exosome-like vesicles
ESTs expressed sequence tags
FISH fluorescence in situ hybridization.
GO gene ontology
GPCR G-protein coupled receptor
GRID-Seq global RNA interactions with DNA identified by deep sequencing
HF hydatid fluid
ilncRNAs intronic lncRNAs
ISH in situ hybridization
LC-MS liquid chromatography–mass spectrometry
LIGR-Seq ligation of interacting RNA followed by deep sequencing
lincRNAs intergenic lncRNAs
lncRNAs long non-coding RNAs
MARGI mapping RNA-genome interactions
MARIO mapping RNA interactome in vivo
miRNAs micro-RNAs
MTZ metronidazole
ncRNA non-coding RNA
NTDs neglected tropical diseases
ORF open reading frame
PAP RNA poly (A) polymerases
PARIS psoralen analysis of RNA interactions and structures
piRNAs piwi-interacting RNAs
PSCs protoscoleces
RADICL-Seq RNA and DNA-interacting complexes ligated identified by deep sequencing
RAP RNA antisense purification
RBP RNA binding protein
RIP RNA immunoprecipitation
RNAi RNA interference
RNA-Seq RNA-sequencing
rRNAs ribosomal RNAs
RT-qPCR reverse transcription quantitative PCR
scRNAs small conditional RNAs
scRNA-Seq single-cell RNA-Seq
SHAPE selective 2′-hydroxyl acylation analyzed by primer extension
SIF stumpy induction factor
siRNAs small interfering RNAs
slncRNAs sense lncRNAs
snoRNAs small nucleolar RNAs
SNPs single nucleotide polymorphisms
snRNAs small nuclear RNAs
SPLASH sequencing of psoralen-crosslinked, ligated, and selected hybrids
SRA sequence read archive
telsRNAs telomere-specific small RNAs
tiRNAs tRNA-derived stress-induced RNAs
tRNAs transfer RNAs
TSS transcription start site
WISH whole-mount in situ hybridization
Xist X-inactive specific transcript

Acknowledgements We acknowledge Patricia Aoki Miyasato and Dr. Eliana Nakano, Laboratorio de Malacologia, Instituto Butantan, for maintaining the *S. mansoni* life cycle and providing the infected hamsters. We would like to thank the Laboratório de Biologia Celular from Instituto Butantan and the Confocal Lab Technician Alexsander Seixas de Souza for the services provided on the Confocal Microscope Leica TCS SP8, purchased through Projeto 175 FINEP—IBUINFRA grant 01.12.0175.00 to Dr. Carlos Jared.

Author contribution Conceived the review: S.V.-A. and M.S.A.; literature search: G.O.S. and M.S.A; data analysis: G.O.S.; performed the in vitro experiments: G.O.S and H.S.C.; wrote the manuscript: G.O.S.; revised the manuscript: G.O.S., M.S.A., S.V.-A.; all authors reviewed the manuscript and approved the submitted version.

Funding This work was supported by a grant from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (Thematic grant number 2018/23693–5 to SV-A). G.O.S. received a fellowship from FAPESP (18/24015–0); H.S.C. received a fellowship from Conselho Nacional do Desenvolvimento Científico e Tecnológico (CNPq) (116733/2019–5); S.V-A laboratory was also supported by institutional funds from Fundação Butantan and S.V-A received an established investigator fellowship award from CNPq, Brasil.

Declarations

Ethics approval Housing conditions of the hamsters and experimental procedures used in this study were in strict accordance with the Ethical Principles in Animal Research adopted by the CONCEA and the experimental protocol was approved by the Ethics Committee for Animal Experimentation of Butantan Institute (CEUAIB n° 8859090919).

Conflict of interest The authors declare no competing interests.

References

Abudayyeh OO, Gootenberg JS, Essletzbichler P et al (2017) RNA targeting with CRISPR–Cas13. Nature 550:280–284. https://doi.org/10.1038/nature24049

Agarwal A, Kanekar S, Sabat S, Thamburaj K (2016) Metronidazole-induced cerebellar toxicity. Neurol Int 8:6365. https://doi.org/10.4081/ni.2016.6365

Albonico M, Engels D, Savioli L (2004) Monitoring drug efficacy and early detection of drug resistance in human soil-transmitted nematodes: a pressing public health agenda for helminth control. Int J Parasitol 34:1205–1210. https://doi.org/10.1016/j.ijpara.2004.08.001

Amaral MS, Maciel LF, Silveira GO et al (2020) Long non-coding RNA levels can be modulated by 5-azacytidine in *Schistosoma mansoni*. Sci Rep 10:21565. https://doi.org/10.1038/s41598-020-78669-5

Amaral PP, Leonardi T, Han N et al (2018) Genomic positional conservation identifies topological anchor point RNAs linked to developmental loci. Genome Biol 19:32. https://doi.org/10.1186/s13059-018-1405-5

Anastasiadou E, Jacob LS, Slack FJ (2018) Non-coding RNA networks in cancer. Nat Rev Cancer 18:5–18. https://doi.org/10.1038/nrc.2017.99

Ashton PD, Harrop R, Shah B, Wilson RA (2001) The schistosome egg: development and secretions. Parasitology 122:329–338. https://doi.org/10.1017/s0031182001007351

Aw JGA, Shen Y, Wilm A et al (2016) In vivo mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. Mol Cell 62:603–617. https://doi.org/10.1016/j.molcel.2016.04.028

Ayupe AC, Tahira AC, Camargo L et al (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. https://doi.org/10.1080/15476286.2015.1062960

Banani SF, Lee HO, Hyman AA, Rosen MK (2017) Biomolecular condensates: organizers of cellular biochemistry. Nat Rev Mol Cell Biol 18:285–298. https://doi.org/10.1038/nrm.2017.7

Baron S (1996) Medical microbiology, 4th edition, 4th edn. Galveston (TX)

Barrett MP, Kyle DE, Sibley LD et al (2019) Protozoan persister-like cells and drug treatment failure. Nat Rev Microbiol 17:607–620. https://doi.org/10.1038/s41579-019-0238-x

Barrett MP, Vincent IM, Burchmore RJ et al (2011) Drug resistance in human African trypanosomiasis. Future Microbiol 6:1037–1047. https://doi.org/10.2217/fmb.11.88

Beckedorff FC, Amaral MS, Deocesano-Pereira C, Verjovski-Almeida S (2013) Long non-coding RNAs and their implications in cancer epigenetics. Biosci Rep 33:667–675. https://doi.org/10.1042/BSR20130054

Bell JC, Jukam D, Teran NA et al (2018) Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNA-to-DNA contacts. Elife 7:e27024. https://doi.org/10.7554/eLife.27024

Ben Ayed L, Sabbahi S (2019) *Entamoeba histolytica*. In: Fayer R, Jakubowski W (eds) Water and Sanitation for the 21st Century: Health and Microbiological Aspects of Excreta and Wastewater Management (Global Water Pathogen Project). Michigan State University. https://doi.org/10.14321/waterpathogens.34

Bergquist R, Utzinger J, Keiser J (2017) Controlling schistosomiasis with praziquantel: how much longer without a viable alternative? Infect Dis Poverty 6:74. https://doi.org/10.1186/s40249-017-0286-2

Bharucha H (2001) Diagnostic pathology of parasitic infections with clinical correlations (2nd edn). Yezid Gutierrez, MD, MPH&TM, PhD. Oxford University Press, New York, 2000. J Pathol 193:277–277. https://doi.org/10.1002/1096-9896(200102)193:2%3c277::AID-PATH806%3e3.0.CO;2-P

Bhattacharya D, Haque R, Singh U (2005) Coding and noncoding genomic regions of *Entamoeba histolytica* have significantly different rates of sequence polymorphisms: implications for epidemiological studies. J Clin Microbiol 43:4815–4819. https://doi.org/10.1128/JCM.43.9.4815-4819.2005

Blessmann J, Ali IKM, Ton Nu PA et al (2003) Longitudinal study of intestinal *Entamoeba histolytica* infections in asymptomatic adult carriers. J Clin Microbiol 41:4745–4750. https://doi.org/10.1128/JCM.41.10.4745-4750.2003

Blokhin I, Khorkova O, Hsiao J, Wahlestedt C (2018) Developments in lncRNA drug discovery: where are we heading? Expert Opin Drug Discov 13:837–849. https://doi.org/10.1080/17460441.2018.1501024

Blythe AJ, Fox AH, Bond CS (2016) The ins and outs of lncRNA structure: how, why and what comes next? Biochim Biophys Acta Gene Regul Mech 1859:46–58. https://doi.org/10.1016/j.bbagrm.2015.08.009

Bonetti A, Agostini F, Suzuki AM et al (2020) RADICL-seq identifies general and cell type–specific principles of genome-wide RNA-chromatin interactions. Nat Commun 11:1018. https://doi.org/10.1038/s41467-020-14337-6

Bryant JM, Baumgarten S, Glover L et al (2019) CRISPR in parasitology: not exactly cut and dried! Trends Parasitol 35:409–422. https://doi.org/10.1016/j.pt.2019.03.004

Buck AH, Coakley G, Simbari F et al (2014) Exosomes secreted by nematode parasites transfer small RNAs to mammalian cells and modulate innate immunity. Nat Commun 5:5488. https://doi.org/10.1038/ncomms6488

Cabili MN, Trapnell C, Goff L et al (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 25:1915–1927. https://doi.org/10.1101/gad.17446611

Calvet JP, Pederson T (1979) Heterogeneous nuclear RNA double-stranded regions probed in living HeLa cells by crosslinking with the psoralen derivative aminomethyltrioxsalen. Proc Natl Acad Sci 76:755–759. https://doi.org/10.1073/pnas.76.2.755

Carlton JM, Hirt RP, Silva JC et al (2007) Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. Science (80-) 315:207–212. https://doi.org/10.1126/science.1132894

Cech TR, Steitz JA (2014) The noncoding RNA revolution—trashing old rules to forge new ones. Cell 157:77–94. https://doi.org/10.1016/j.cell.2014.03.008

Chen Y, Li Z, Chen X, Zhang S (2021) Long non-coding RNAs: from disease code to drug role. Acta Pharm Sin B 11:340–354. https://doi.org/10.1016/j.apsb.2020.10.001

Chikne V, Gupta SK, Doniger T et al (2017) The canonical poly (A) polymerase PAP1 polyadenylates non-coding RNAs and is essential for snoRNA biogenesis in *Trypanosoma brucei*. J Mol Biol 429:3301–3318. https://doi.org/10.1016/j.jmb.2017.04.015

Chillón I, Marcia M (2020) The molecular structure of long non-coding RNAs: emerging patterns and functional implications. Crit Rev Biochem Mol Biol 55:662–690. https://doi.org/10.1080/10409238.2020.1828259

Choi S-W, Kim H-W, Nam J-W (2019) The small peptide world in long noncoding RNAs. Brief Bioinform 20:1853–1864. https://doi.org/10.1093/bib/bby055

Choudhary C, Sharma S, Meghwanshi KK et al (2021) Long non-coding RNAs in insects. Animals 11:1118. https://doi.org/10.3390/ani11041118

Chu C, Qu K, Zhong FL et al (2011) Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell 44:667–678. https://doi.org/10.1016/j.molcel.2011.08.027

Churchman LS (2017) Not just noise: genomics and genetics bring long noncoding RNAs into focus. Mol Cell 65:1–2. https://doi.org/10.1016/j.molcel.2016.12.017

Cioli D, Pica-Mattoccia L, Basso A, Guidi A (2014) Schistosomiasis control: praziquantel forever? Mol Biochem Parasitol 195:23–29. https://doi.org/10.1016/j.molbiopara.2014.06.002

Coakley G, Maizels RM, Buck AH (2015) Exosomes and other extracellular vesicles: the new communicators in parasite infections. Trends Parasitol 31:477–489. https://doi.org/10.1016/j.pt.2015.06.009

Collins JJ III, Wang B, Lambrus BG et al (2013) Adult somatic stem cells in the human parasite *Schistosoma mansoni*. Nature 494:476–479. https://doi.org/10.1038/nature11924

Collins JJ (2017) Platyhelminthes. Curr Biol 27:R252–R256. https://doi.org/10.1016/j.cub.2017.02.016

Copeland CC, Marz M, Rose D et al (2009) Homology-based annotation of non-coding RNAs in the genomes of *Schistosoma*

mansoni and *Schistosoma japonicum*. BMC Genomics 10:464. https://doi.org/10.1186/1471-2164-10-464
Crick F (1970) Central dogma of molecular biology. Nature 227:561–563. https://doi.org/10.1038/227561a0
Crick FH (1958) On protein synthesis. Symp Soc Exp Biol 12:138–163
Dahariya S, Paddibhatla I, Kumar S et al (2019) Long non-coding RNA: classification, biogenesis and functions in blood cells. Mol Immunol 112:82–92. https://doi.org/10.1016/j.molimm.2019.04.011
Derrien T, Johnson R, Bussotti G et al (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 22:1775–1789. https://doi.org/10.1101/gr.132159.111
Dhanoa JK, Sethi RS, Verma R et al (2018) Long non-coding RNA: its evolutionary relics and biological implications in mammals: a review. J Anim Sci Technol 60:25. https://doi.org/10.1186/s40781-018-0183-7
Di Giammartino DC, Manley JL (2014) New links between mRNA polyadenylation and diverse nuclear pathways. Mol Cells 37:644–649. https://doi.org/10.14348/molcells.2014.0177
Diaz Soria CL, Lee J, Chong T et al (2020) Single-cell atlas of the first intra-mammalian developmental stage of the human parasite *Schistosoma mansoni*. Nat Commun 11:6411. https://doi.org/10.1038/s41467-020-20092-5
Diederichs S (2014) The four dimensions of noncoding RNA conservation. Trends Genet 30:121–123. https://doi.org/10.1016/j.tig.2014.01.004
Djebali S, Davis CA, Merkel A et al (2012) Landscape of transcription in human cells. Nature 489:101–108. https://doi.org/10.1038/nature11233
Du X, McManus DP, French JD et al (2021) CRISPR/Cas9: a new tool for the study and control of helminth parasites. BioEssays 43:2000185. https://doi.org/10.1002/bies.202000185
Dumas C, Chow C, Müller M, Papadopoulou B (2006) A novel class of developmentally regulated noncoding RNAs in *Leishmania*. Eukaryot Cell 5:2033–2046. https://doi.org/10.1128/EC.00147-06
Eddy SR (2001) Non–coding RNA genes and the modern RNA world. Nat Rev Genet 2:919–929. https://doi.org/10.1038/35103511
Eisele TP (2019) Mass drug administration can be a valuable addition to the malaria elimination toolbox. Malar J 18:281. https://doi.org/10.1186/s12936-019-2906-8
Elsayed AM, Amero P, Salama SA et al (2020) Back to the future: rethinking the great potential of lncRNAS for optimizing chemotherapeutic response in ovarian cancer. Cancers (Basel) 12:2406. https://doi.org/10.3390/cancers12092406
Engreitz JM, Pandya-Jones A, McDonel P et al (2013) The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science (80) 341:1237973. https://doi.org/10.1126/science.1237973
Fang Y, Fullwood MJ (2016) Roles, functions, and mechanisms of long non-coding RNAs in cancer. Genomics Proteomics Bioinformatics 14:42–54. https://doi.org/10.1016/j.gpb.2015.09.006
Fathi Dizaji B (2020) Strategies to target long non-coding RNAs in cancer treatment: progress and challenges. Egypt J Med Hum Genet 21:41. https://doi.org/10.1186/s43042-020-00074-4
Fatima F, Nawaz M (2017) Vesiculated long non-coding RNAs: offshore packages deciphering trans-regulation between cells, cancer progression and resistance to therapies. Non-Coding RNA 3:10. https://doi.org/10.3390/ncrna3010010
Fei J, Jadaliha M, Harmon TS et al (2017) Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution. J Cell Sci 130:4180–4192. https://doi.org/10.1242/jcs.206854
Ferlini A, Goyenvalle A, Muntoni F (2021) RNA-targeted drugs for neuromuscular diseases. Science 80(371):29–31. https://doi.org/10.1126/science.aba4515
Fesenko I, Kirov I, Kniazev A et al (2019) Distinct types of short open reading frames are translated in plant cells. Genome Res 29:1464–1477. https://doi.org/10.1101/gr.253302.119
Fire A, Xu S, Montgomery MK et al (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391:806–811. https://doi.org/10.1038/35888
Garcia-Silva MR, Cabrera-Cabrera F, Cura das Neves FR et al (2014a) Gene expression changes induced by *Trypanosoma cruzi* shed microvesicles in mammalian host cells: relevance of tRNA-derived halves. Biomed Res Int 2014:1–11. https://doi.org/10.1155/2014/305239
Garcia-Silva MR, Cura das Neves RF, Cabrera-Cabrera F et al (2014b) Extracellular vesicles shed by *Trypanosoma cruzi* are linked to small RNA pathways, life cycle regulation, and susceptibility to infection of mammalian cells. Parasitol Res 113:285–304. https://doi.org/10.1007/s00436-013-3655-1
Geerts S, Gryseels B (2000) Drug resistance in human helminths: current situation and lessons from livestock. Clin Microbiol Rev 13:207–222. https://doi.org/10.1128/CMR.13.2.207-222.2000
Geyer KK, Rodríguez López CM, Chalmers IW et al (2011) Cytosine methylation regulates oviposition in the pathogenic blood fluke *Schistosoma mansoni*. Nat Commun 2:424. https://doi.org/10.1038/ncomms1433
Gryseels B, Polman K, Clerinx J, Kestens L (2006) Human schistosomiasis. Lancet 368:1106–1118. https://doi.org/10.1016/S0140-6736(06)69440-3
Guegan F, Bento F, Neves D et al (2020) A long non-coding RNA controls parasite differentiation in African trypanosomes. bioRxiv 2020.05.03.074625. https://doi.org/10.1101/2020.05.03.074625
Guo C-J, Ma X-K, Xing Y-H et al (2020) Distinct processing of lncRNAs contributes to non-conserved functions in stem cells. Cell 181:621-636.e22. https://doi.org/10.1016/j.cell.2020.03.006
Gupta Y, Goicoechea S, Pearce CM et al (2021) The emerging paradigm of calcium homeostasis as a new therapeutic target for protozoan parasites. Med Res Rev 2021:1–27. https://doi.org/10.1002/med.21804
Guttman M, Amit I, Garber M et al (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 458:223–227. https://doi.org/10.1038/nature07672
Guttman M, Russell P, Ingolia NT et al (2013) Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. Cell 154:240–251. https://doi.org/10.1016/j.cell.2013.06.009
Haga CL, Velagapudi SP, Strivelli JR et al (2015) Small molecule inhibition of miR-544 biogenesis disrupts adaptive responses to hypoxia by modulating ATM-mTOR signaling. ACS Chem Biol 10:2267–2276. https://doi.org/10.1021/acschembio.5b00265
Harrow J, Frankish A, Gonzalez JM et al (2012) GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res 22:1760–1774. https://doi.org/10.1101/gr.135350.111
Hartford CCR, Lal A (2020) When long noncoding becomes protein coding. Mol Cell Biol 40(6):e00528-e619. https://doi.org/10.1128/MCB.00528-19
Helwak A, Tollervey D (2016) Identification of miRNA-target RNA interactions using CLASH. In: Dassi E (eds) Post-transcriptional gene regulation. Methods in molecular biology. Humana Press, New York, vol 1358. https://doi.org/10.1007/978-1-4939-3067-8_14
Hezroni H, Koppstein D, Schwartz MG et al (2015) Principles of long noncoding RNA evolution derived from direct comparison of

transcriptomes in 17 species. Cell Rep 11:1110–1122. https://doi.org/10.1016/j.celrep.2015.04.023

Horiuchi A, Satou T, Akao N et al (2005) The effect of free and polyethylene glycol-liposome-entrapped albendazole on larval mobility and number in *Toxocara canis* infected mice. Vet Parasitol 129:83–87. https://doi.org/10.1016/j.vetpar.2004.12.017

Horvath S (2011) Weighted network analysis. Springer New York, New York

Hotez PJ, Brindley PJ, Bethony JM et al (2008) Helminth infections: the great neglected tropical diseases. J Clin Invest 118:1311–1321. https://doi.org/10.1172/JCI34261

Hotez PJ, Molyneux DH, Fenwick A et al (2007) Control of neglected tropical diseases. N Engl J Med 357:1018–1027. https://doi.org/10.1056/NEJMra064142

Howick VM, Russell AJC, Andrews T et al (2019) The Malaria Cell Atlas: single parasite transcriptomes across the complete *Plasmodium* life cycle. Science (80) 365:eaaw2619. https://doi.org/10.1126/science.aaw2619

Ingolia NT (2014) Ribosome profiling: new views of translation, from single codons to genome scale. Nat Rev Genet 15:205–213. https://doi.org/10.1038/nrg3645

Jeelani G, Nozaki T (2014) Metabolomic analysis of *Entamoeba*: applications and implications. Curr Opin Microbiol 20:118–124. https://doi.org/10.1016/j.mib.2014.05.016

Jensen KB, Darnell RB (2008) CLIP: crosslinking and immunoprecipitation of in vivo RNA targets of RNA-binding proteins. In: Lin RJ (eds) RNA-protein interaction protocols. Methods in molecular biology. Humana Press, vol 488. https://doi.org/10.1007/978-1-60327-475-3_6

Jiang W, Qu Y, Yang Q 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. https://doi.org/10.1080/15476286.2019.1649584

Jinek M, Chylinski K, Fonfara I et al (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 80(337):816–821. https://doi.org/10.1126/science.1225829

Kastenmayer JP (2006) Functional genomics of genes with small open reading frames (sORFs) in S. cerevisiae. Genome Res 16:365–373. https://doi.org/10.1101/gr.4355406

Kaufer A, Ellis J, Stark D, Barratt J (2017) The evolution of trypanosomatid taxonomy. Parasit Vectors 10:287. https://doi.org/10.1186/s13071-017-2204-7

Kim HC, Khalil AM, Jolly ER (2020) LncRNAs in molluscan and mammalian stages of parasitic schistosomes are developmentally-regulated and coordinately expressed with protein-coding genes. RNA Biol 17:805–815. https://doi.org/10.1080/15476286.2020.1729594

Kim Y-K (2020) RNA therapy: current status and future potential. Chonnam Med J 56:87. https://doi.org/10.4068/cmj.2020.56.2.87

Kirkcaldy RD, Augostini P, Asbel LE et al (2012) *Trichomonas vaginalis* antimicrobial drug resistance in 6 US cities, STD Surveillance Network, 2009–2010. Emerg Infect Dis 18:939–943. https://doi.org/10.3201/eid1806.111590

Koch H, Raabe M, Urlaub H et al (2016) The polyadenylation complex of *Trypanosoma brucei*: characterization of the functional poly(A) polymerase. RNA Biol 13:221–231. https://doi.org/10.1080/15476286.2015.1130208

Kolev NG, Franklin JB, Carmi S et al (2010) The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. PLoS Pathog 6:1–15. https://doi.org/10.1371/journal.ppat.1001090

Kumar V, Abbas AK, Aster JC (2017) Robbins basic pathology (10th ed.). Elsevier - Health Sciences Division

Kusdian G, Woehle C, Martin WF, Gould SB (2013) The actin-based machinery of *Trichomonas vaginalis* mediates flagellate-amoeboid transition and migration across host tissue. Cell Microbiol 15:1707–1721. https://doi.org/10.1111/cmi.12144

Lee JM (2003) Genomic gene clustering analysis of pathways in eukaryotes. Genome Res 13:875–882. https://doi.org/10.1101/gr.737703

Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75:843–854. https://doi.org/10.1016/0092-8674(93)90529-Y

Leucci E, Vendramin R, Spinazzi M et al (2016) Melanoma addiction to the long non-coding RNA SAMMSON. Nature 531:518–522. https://doi.org/10.1038/nature17161

Li X, Zhou B, Chen L et al (2017) GRID-seq reveals the global RNA–chromatin interactome. Nat Biotechnol 35:940–950. https://doi.org/10.1038/nbt.3968

Li Y, Baptista RP, Kissinger JC (2020) Noncoding RNAs in Apicomplexan parasites: an update. Trends Parasitol 36:835–849. https://doi.org/10.1016/j.pt.2020.07.006

Li Y, Baptista RP, Sateriale A et al (2021) Analysis of long non-coding RNA in *Cryptosporidium parvum* reveals significant stage-specific antisense transcription. Front Cell Infect Microbiol 10:1–16. https://doi.org/10.3389/fcimb.2020.608298

Liao Q, Shen J, Liu J et al (2014) Genome-wide identification and functional annotation of *Plasmodium falciparum* long noncoding RNAs from RNA-seq data. Parasitol Res 113:1269–1281. https://doi.org/10.1007/s00436-014-3765-4

Liao Q, Zhang Y, Zhu Y et al (2018) Identification of long noncoding RNAs in *Schistosoma mansoni* and *Schistosoma japonicum*. Exp Parasitol 191:82–87. https://doi.org/10.1016/j.exppara.2018.07.001

Loda A, Heard E (2019) Xist RNA in action: past, present, and future. PLOS Genet 15:e1008333. https://doi.org/10.1371/journal.pgen.1008333

Loftus B, Anderson I, Davies R et al (2005) The genome of the protist parasite *Entamoeba histolytica*. Nature 433:865–868. https://doi.org/10.1038/nature03291

Lorenzi HA, Puiu D, Miller JR et al (2010) New assembly, reannotation and analysis of the *Entamoeba histolytica* genome reveal new genomic features and protein content information. PLoS Negl Trop Dis 4:e716. https://doi.org/10.1371/journal.pntd.0000716

Lothstein KE, Gause WC (2021) Mining helminths for novel therapeutics. Trends Mol Med 27:345–364. https://doi.org/10.1016/j.molmed.2020.12.010

Loughrey D, Watters KE, Settle AH, Lucks JB (2014) SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. Nucleic Acids Res 42:e165–e165. https://doi.org/10.1093/nar/gku909

Ma L, Bajic VB, Zhang Z (2013) On the classification of long non-coding RNAs. RNA Biol 10:924–933. https://doi.org/10.4161/rna.24604

Maciel L, Morales-Vicente D, Verjovski-Almeida S (2020) Dynamic expression of long non-coding RNAs throughout parasite sexual and neural maturation in *Schistosoma japonicum*. Non-Coding RNA 6:15. https://doi.org/10.3390/ncrna6020015

Maciel LF, Morales-Vicente DA, Silveira GO 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. https://doi.org/10.3389/fgene.2019.00823

Maciel LF, Verjovski-Almeida S (2020) Step-by-step bioinformatics analysis of *Schistosoma mansoni* long non-coding RNA sequences. Timson DJ (ed.) *Schistosoma mansoni*: methods and protocols, Methods in Molecular Biology, vol. 2151,. pp 109–133. https://doi.org/10.1007/978-1-0716-0635-3_10

Marcilla A, Martin-Jaular L, Trelis M et al (2014) Extracellular vesicles in parasitic diseases. J Extracell Vesicles 3:25040. https://doi.org/10.3402/jev.v3.25040
Martinez TF, Chu Q, Donaldson C et al (2020) Accurate annotation of human protein-coding small open reading frames. Nat Chem Biol 16:458–468. https://doi.org/10.1038/s41589-019-0425-0
Matsui M, Corey DR (2017) Non-coding RNAs as drug targets. Nat Rev Drug Discov 16:167–179. https://doi.org/10.1038/nrd.2016.117
McManus DP, Dunne DW, Sacko M et al (2018) Schistosomiasis. Nat Rev Dis Prim 4:13. https://doi.org/10.1038/s41572-018-0013-8
McVeigh P, Maule AG (2019) Can CRISPR help in the fight against parasitic worms? Elife 8:e44382. https://doi.org/10.7554/eLife.44382
Melman SD, Steinauer ML, Cunningham C et al (2009) Reduced susceptibility to praziquantel among naturally occurring Kenyan isolates of *Schistosoma mansoni*. PLoS Negl Trop Dis 3(8):e504. https://doi.org/10.1371/journal.pntd.0000504
Menna-Barreto RFS (2019) Cell death pathways in pathogenic trypanosomatids: lessons of (over)kill. Cell Death Dis 10:93. https://doi.org/10.1038/s41419-019-1370-2
Meri T, Jokiranta TS, Suhonen L, Meri S (2000) Resistance of *Trichomonas vaginalis* to metronidazole: report of the first three cases from Finland and optimization of in vitro susceptibility testing under various oxygen concentrations. J Clin Microbiol 38:763–767. https://doi.org/10.1128/JCM.38.2.763-767.2000
Molina I, Gómez i Prat J, Salvador F et al (2014) Randomized trial of posaconazole and benznidazole for chronic Chagas' disease. N Engl J Med 370:1899–1908. https://doi.org/10.1056/NEJMoa1313122
Mony BM, MacGregor P, Ivens A et al (2014) Genome-wide dissection of the quorum sensing signalling pathway in *Trypanosoma brucei*. Nature 505:681–685. https://doi.org/10.1038/nature12864
Moraes CB, Giardini MA, Kim H et al (2015) Nitroheterocyclic compounds are more efficacious than CYP51 inhibitors against *Trypanosoma cruzi*: implications for Chagas disease drug discovery and development. Sci Rep 4:4703. https://doi.org/10.1038/srep04703
Myers SA, Wright J, Peckner R et al (2018) Discovery of proteins associated with a predefined genomic locus via dCas9–APEX-mediated proximity labeling. Nat Methods 15:437–439. https://doi.org/10.1038/s41592-018-0007-1
Nanes Sarfati D, Li P, Tarashansky AJ, Wang B (2021) Single-cell deconstruction of stem-cell-driven schistosome development. Trends Parasitol 37:790–802. https://doi.org/10.1016/j.pt.2021.03.005
Nawy T (2014) Single-cell sequencing. Nat Methods 11:18–18. https://doi.org/10.1038/nmeth.2771
Nguyen TC, Cao X, Yu P et al (2016) Mapping RNA–RNA interactome and RNA structure in vivo by MARIO. Nat Commun 7:12023. https://doi.org/10.1038/ncomms12023
Novačić A, Vučenović I, Primig M, Stuparević I (2020) Non-coding RNAs as cell wall regulators in *Saccharomyces cerevisiae*. Crit Rev Microbiol 46:15–25. https://doi.org/10.1080/1040841X.2020.1715340
Oliveira KC, Carvalho MLP, Maracaja-Coutinho V et al (2011) Non-coding RNAs in schistosomes: an unexplored world. An Acad Bras Cienc 83:673–694. https://doi.org/10.1590/s0001-37652011000200026
Oliveira VF, Moares LAG, Mota EA et al (2018) Identification of 170 new long noncoding RNAs in *Schistosoma mansoni*. Biomed Res Int 2018:1–9. https://doi.org/10.1155/2018/1264697
Olliaro PL (2010) Drug combinations for visceral leishmaniasis. Curr Opin Infect Dis 23:595–602. https://doi.org/10.1097/QCO.0b013e32833fca9d
Palacino J, Swalley SE, Song C et al (2015) SMN2 splice modulators enhance U1–pre-mRNA association and rescue SMA mice. Nat Chem Biol 11:511–517. https://doi.org/10.1038/nchembio.1837
Paulish-Miller TE, Augostini P, Schuyler JA et al (2014) *Trichomonas vaginalis* metronidazole resistance is associated with single nucleotide polymorphisms in the nitroreductase genes ntr4Tv and ntr6Tv. Antimicrob Agents Chemother 58:2938–2943. https://doi.org/10.1128/AAC.02370-13
Pawar H, Pai K, Patole MS (2017) A novel protein coding potential of long intergenic non-coding RNAs (lincRNAs) in the kinetoplastid protozoan parasite *Leishmania major*. Acta Trop 167:21–25. https://doi.org/10.1016/j.actatropica.2016.12.012
Pearman WS, Freed NE, Silander OK (2020) Testing the advantages and disadvantages of short- and long- read eukaryotic metagenomics using simulated reads. BMC Bioinformatics 21:220. https://doi.org/10.1186/s12859-020-3528-4
Piovesan A, Antonaros F, Vitale L et al (2019) Human protein-coding genes and gene feature statistics in 2019. BMC Res Notes 12:315. https://doi.org/10.1186/s13104-019-4343-8
Podlipaev S (2001) The more insect trypanosomatids under study-the more diverse Trypanosomatidae appears. 31:648–652. https://doi.org/10.1016/S0020-7519(01)00139-4
Ponjavic J, Ponting CP, Lunter G (2007) Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. Genome Res 17:556–565. https://doi.org/10.1101/gr.6036807
Prabhakar B, Zhong X-B, Rasmussen TP (2017) Exploiting long non-coding RNAs as pharmacological targets to modulate epigenetic diseases. Yale J Biol Med 90(1):73–86
Prichard RK (2007) Markers for benzimidazole resistance in human parasitic nematodes? Parasitology 134:1087–1092. https://doi.org/10.1017/S003118200700008X
Prichard RK, Roulet A (2007) ABC transporters and β-tubulin in macrocyclic lactone resistance: prospects for marker development. Parasitology 134:1123–1132. https://doi.org/10.1017/S0031182007000091
Quinn JJ, Chang HY (2016) Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet 17:47–62. https://doi.org/10.1038/nrg.2015.10
Quinn JJ, Zhang QC, Georgiev P et al (2016) Rapid evolutionary turnover underlies conserved lncRNA–genome interactions. Genes Dev 30:191–207. https://doi.org/10.1101/gad.272187.115
Rajic ZA, Jankovic GM, Vidovic A et al (2005) Size of the protein-coding genome and rate of molecular evolution. J Hum Genet 50:217–229. https://doi.org/10.1007/s10038-005-0242-z
Ransohoff JD, Wei Y, Khavari PA (2018) The functions and unique features of long intergenic non-coding RNA. Nat Rev Mol Cell Biol 19:143–157. https://doi.org/10.1038/nrm.2017.104
Rastrojo A, Carrasco-Ramiro F, Martín D et al (2013) The transcriptome of *Leishmania major* in the axenic promastigote stage: transcript annotation and relative expression levels by RNA-seq. BMC Genomics 14:1–13. https://doi.org/10.1186/1471-2164-14-223
Reinhart BJ, Slack FJ, Basson M et al (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. Nature 403:901–906. https://doi.org/10.1038/35002607
Ren Y, Song Y, Zhang L et al (2021) Coding of non-coding RNA: insights into the regulatory functions of Pri-MicroRNA-encoded peptides in plants. Front Plant Sci 12:641351. https://doi.org/10.3389/fpls.2021.641351
Rinaldi F (2014) Cystic echinococcosis of the liver: a primer for hepatologists. World J Hepatol 6:293. https://doi.org/10.4254/wjh.v6.i5.293
Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. Annu Rev Biochem 81:145–166. https://doi.org/10.1146/annurev-biochem-051410-092902

Rivas E, Clements J, Eddy SR (2017) A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. Nat Methods 14:45–48. https://doi.org/10.1038/nmeth.4066

Rosenberg M, Utzinger J, Addiss DG (2016) Preventive chemotherapy versus innovative and intensified disease management in neglected tropical diseases: a distinction whose shelf life has expired. PLoS Negl Trop Dis 10:e0004521. https://doi.org/10.1371/journal.pntd.0004521

Ross AGP, Olveda RM, Li Y (2015) An audacious goal: the elimination of schistosomiasis in our lifetime through mass drug administration. Lancet 385:2220–2221. https://doi.org/10.1016/S0140-6736(14)61417-3

Ruy PDC, Monteiro-Teles NM, Miserani Magalhães RD et al (2019) Comparative transcriptomics in *Leishmania braziliensis:* disclosing differential gene expression of coding and putative noncoding RNAs across developmental stages. RNA Biol 16:639–660. https://doi.org/10.1080/15476286.2019.1574161

Saha A, Bhattacharya S, Bhattacharya A (2016) Serum stress responsive gene EhslncRNA of *Entamoeba histolytica* is a novel long noncoding RNA. Sci Rep 6:1–9. https://doi.org/10.1038/srep27476

Sati S, Ghosh S, Jain V et al (2012) Genome-wide analysis reveals distinct patterns of epigenetic features in long non-coding RNA loci. Nucleic Acids Res 40:10018–10031. https://doi.org/10.1093/nar/gks776

Sberro H, Fremin BJ, Zlitni S et al (2019) Large-scale analyses of human microbiomes reveal thousands of small, novel genes. Cell 178:1245-1259.e14. https://doi.org/10.1016/j.cell.2019.07.016

Schmitt AM, Chang HY (2016) Long noncoding RNAs in cancer pathways. Cancer Cell 29:452–463. https://doi.org/10.1016/j.ccell.2016.03.010

Schwebke JR, Barrientes FJ (2006) Prevalence of *Trichomonas vaginalis* isolates with resistance to metronidazole and tinidazole. Antimicrob Agents Chemother 50:4209–4210. https://doi.org/10.1128/AAC.00814-06

Shao Y, Zhang QC (2020) Targeting RNA structures in diseases with small molecules. Essays Biochem 64:955–966. https://doi.org/10.1042/EBC20200011

Shapiro TA, Englund PT (1995) The structure and replication of kinetoplast DNA. Annu Rev Microbiol 49:117–143. https://doi.org/10.1146/annurev.mi.49.100195.001001

Sharma E, Sterne-Weiler T, O'Hanlon D, Blencowe BJ (2016) Global mapping of human RNA-RNA interactions. Mol Cell 62:618–626. https://doi.org/10.1016/j.molcel.2016.04.030

Shirley D-AT, Farr L, Watanabe K, Moonah S (2018) A review of the global burden, new diagnostics, and current therapeutics for amebiasis. Open Forum Infect Dis 5(7):ofy161. https://doi.org/10.1093/ofid/ofy161

Simon MD (2013) Capture Hybridization Analysis of RNA Targets (CHART). In: Current protocols in molecular biology. Wiley, Hoboken. https://doi.org/10.1002/0471142727.mb2125s101

Simpson AGB, Stevens JR, Lukeš J (2006) The evolution and diversity of kinetoplastid flagellates. Trends Parasitol 22:168–174. https://doi.org/10.1016/j.pt.2006.02.006

Sirekbasan S, Gurkok Tan T (2021) In silico analysis of common long noncoding RNAs in *Schistosoma mansoni* and *Schistosoma haematobium*. J Trop Med 2021:1–8. https://doi.org/10.1155/2021/6617118

Sivaramakrishnan M, McCarthy KD, Campagne S et al (2017) Binding to SMN2 pre-mRNA-protein complex elicits specificity for small molecule splicing modifiers. Nat Commun 8:1476. https://doi.org/10.1038/s41467-017-01559-4

Smallegan MJ, Rinn JL (2019) Linking long noncoding RNA to drug resistance. Proc Natl Acad Sci 116:21963–21965. https://doi.org/10.1073/pnas.1915690116

Smith A, Johnson P (2011) Gene expression in the unicellular eukaryote *Trichomonas vaginalis*. Res Microbiol 162:646–654. https://doi.org/10.1016/j.resmic.2011.04.007

Sridhar B, Rivas-Astroza M, Nguyen TC et al (2017) Systematic mapping of RNA-chromatin interactions in vivo. Curr Biol 27:602–609. https://doi.org/10.1016/j.cub.2017.01.011

Stanley SL (2003) Amoebiasis. Lancet 361:1025–1034. https://doi.org/10.1016/S0140-6736(03)12830-9

Statello L, Guo C-J, Chen L-L, Huarte M (2021) Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol 22:96–118. https://doi.org/10.1038/s41580-020-00315-9

Stephenson ML, Zamecnik PC (1978) Inhibition of Rous sarcoma viral RNA translation by a specific oligodeoxyribonucleotide. Proc Natl Acad Sci 75:285–288. https://doi.org/10.1073/pnas.75.1.285

Steverding D (2010) The development of drugs for treatment of sleeping sickness: a historical review. Parasit Vectors 3:15. https://doi.org/10.1186/1756-3305-3-15

Stolk WA, Kulik MC, le Rutte EA et al (2016) Between-country inequalities in the neglected tropical disease burden in 1990 and 2010, with projections for 2020. PLoS Negl Trop Dis 10:e0004560. https://doi.org/10.1371/journal.pntd.0004560

Stresemann C, Lyko F (2008) Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. Int J Cancer 123:8–13. https://doi.org/10.1002/ijc.23607

Sundar S, Chakravarty J (2015) Investigational drugs for visceral leishmaniasis. Expert Opin Investig Drugs 24:43–59. https://doi.org/10.1517/13543784.2014.954035

Tavares RCA, Pyle AM, Somarowthu S (2019) Phylogenetic analysis with improved parameters reveals conservation in lncRNA structures. J Mol Biol 431:1592–1603. https://doi.org/10.1016/j.jmb.2019.03.012

Torres M, Becquet D, Guillen S et al (2018) RNA pull-down procedure to identify RNA targets of a long non-coding RNA. J vis Exp 134:57379. https://doi.org/10.3791/57379

Townley-Tilson WHD (2006) Genome-wide analysis of mRNAs bound to the histone stem-loop binding protein. RNA 12:1853–1867. https://doi.org/10.1261/rna.76006

Urbina JA (2015) Recent clinical trials for the etiological treatment of chronic Chagas disease: advances, challenges and perspectives. J Eukaryot Microbiol 62:149–156. https://doi.org/10.1111/jeu.12184

Vale N, Gouveia MJ, Rinaldi G et al (2017) Praziquantel for schistosomiasis: single-drug metabolism revisited, mode of action, and resistance. Antimicrob Agents Chemother 61(5):e02582-e2616. https://doi.org/10.1128/AAC.02582-16

Vasconcelos EJR, Mesel VC, DaSilva LF et al (2018) Atlas of *Schistosoma mansoni* long non-coding RNAs and their expression correlation to protein-coding genes. Database 2018:1–5. https://doi.org/10.1093/database/bay068

Vasconcelos EJRR, Dasilva LF, Pires DS et al (2017) The *Schistosoma mansoni* genome encodes thousands of long non-coding RNAs predicted to be functional at different parasite life-cycle stages. Sci Rep 7:10508. https://doi.org/10.1038/s41598-017-10853-6

Vassella E, Reuner B, Yutzy B, Boshart M (1997) Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. J Cell Sci 110(Pt 21):2661–2671

Vendramin R, Verheyden Y, Ishikawa H et al (2018) SAMMSON fosters cancer cell fitness by concertedly enhancing mitochondrial and cytosolic translation. Nat Struct Mol Biol 25:1035–1046. https://doi.org/10.1038/s41594-018-0143-4

Walters HA, Temesvari LA (2021) Target acquired: transcriptional regulators as drug targets for protozoan parasites. Int J Parasitol 51:599–611. https://doi.org/10.1016/j.ijpara.2020.12.007

Wang F, Zuroske T, Watts JK (2020a) RNA therapeutics on the rise. Nat Rev Drug Discov 19:441–442. https://doi.org/10.1038/d41573-020-00078-0

Wang J (2021) Genomics of the parasitic nematode *Ascaris* and its relatives. Genes (basel) 12:493. https://doi.org/10.3390/genes12040493

Wang J, Chen R, Collins JJ (2019) Systematically improved in vitro culture conditions reveal new insights into the reproductive biology of the human parasite *Schistosoma mansoni*. PLOS Biol 17:e3000254. https://doi.org/10.1371/journal.pbio.3000254

Wang KC, Chang HY (2011) Molecular mechanisms of long noncoding RNAs. Mol Cell 43:904–914. https://doi.org/10.1016/j.molcel.2011.08.018

Wang W, Wang L, Liang YS (2012) Susceptibility or resistance of praziquantel in human schistosomiasis: a review. Parasitol Res 111:1871–1877. https://doi.org/10.1007/s00436-012-3151-z

Wang X, Lin J, Li F et al (2017) Screening and functional identification of lncRNAs under β-diketone antibiotic exposure to zebrafish (*Danio rerio*) using high-throughput sequencing. Aquat Toxicol 182:214–225. https://doi.org/10.1016/j.aquatox.2016.12.003

Wang Y, Chen S, Li W et al (2020) Associating divergent lncRNAs with target genes by integrating genome sequence, gene expression and chromatin accessibility data. NAR Genomics Bioinforma 2:lqaa019. https://doi.org/10.1093/nargab/lqaa019

Wang Y, Wang Z, Xu J et al (2018) Systematic identification of non-coding pharmacogenomic landscape in cancer. Nat Commun 9:3192. https://doi.org/10.1038/s41467-018-05495-9

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63. https://doi.org/10.1038/nrg2484

Wendt G, Zhao L, Chen R et al (2020) A single-cell RNA-seq atlas of *Schistosoma mansoni* identifies a key regulator of blood feeding. Science 80(369):1644–1649. https://doi.org/10.1126/science.abb7709

WHO Team: Global Malaria Programme (2020) World malaria report 2020. https://www.who.int/publications/i/item/9789240015791. Accessed 17 May 2021

Woehle C, Kusdian G, Radine C et al (2014) The parasite *Trichomonas vaginalis* expresses thousands of pseudogenes and long non-coding RNAs independently from functional neighbouring genes. BMC Genomics 15:1–12. https://doi.org/10.1186/1471-2164-15-906

World Health Organization (2014) Echinococcosis Fact sheet N°377. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/echinococcosis. Accessed 17 May 2021

World Health Organization (WHO) (2019) Sexually transmitted infections (STIs). World Health Organization. https://www.who.int/news-room/fact-sheets/detail/sexually-transmitted-infections-(stis). Accessed 17 May 2021

Xing J, Liu H, Jiang W, Wang L (2021) LncRNA-encoded peptide: functions and predicting methods. Front Oncol 10:622294. https://doi.org/10.3389/fonc.2020.622294

Yang M, Shang X, Zhou Y et al (2021) Full-length transcriptome analysis of *Plasmodium falciparum* by single-molecule long-read sequencing. Front Cell Infect Microbiol 11:1–11. https://doi.org/10.3389/fcimb.2021.631545

Yang X, Meng T (2019) Long noncoding RNA in preeclampsia: transcriptional noise or innovative indicators? Biomed Res Int 2019:1–7. https://doi.org/10.1155/2019/5437621

Young AP, Jackson DJ, Wyeth RC (2020) A technical review and guide to RNA fluorescence in situ hybridization. PeerJ 8:e8806. https://doi.org/10.7717/peerj.8806

Young DD, Connelly CM, Grohmann C, Deiters A (2010) Small molecule modifiers of microrna miR-122 function for the treatment of hepatitis C virus infection and hepatocellular carcinoma. J Am Chem Soc 132:7976–7981. https://doi.org/10.1021/ja910275u

Yuan L, Xu Z-Y, Ruan S-M et al (2020) Long non-coding RNAs towards precision medicine in gastric cancer: early diagnosis, treatment, and drug resistance. Mol Cancer 19:96. https://doi.org/10.1186/s12943-020-01219-0

Zamecnik PC, Stephenson ML (1978) Inhibition of Rous sarcoma virus replication and cell transformation by a specific oligodeoxynucleotide. Proc Natl Acad Sci 75:280–284. https://doi.org/10.1073/pnas.75.1.280

Zampetaki A, Albrecht A, Steinhofel K (2018) Long non-coding RNA structure and function: is there a link? Front Physiol 9:1201. https://doi.org/10.3389/fphys.2018.01201

Zhang X, Gong W, Cao S et al (2020) Comprehensive analysis of non-coding RNA profiles of exosome-like vesicles from the protoscoleces and hydatid cyst fluid of *Echinococcus granulosus*. Front Cell Infect Microbiol 10:1–14. https://doi.org/10.3389/fcimb.2020.00316

Zuber JA, Takala-Harrison S (2018) Multidrug-resistant malaria and the impact of mass drug administration. Infect Drug Resist 11:299–306. https://doi.org/10.2147/IDR.S123887

Zucca M, Savoia D (2011) Current developments in the therapy of protozoan infections. Open Med Chem J 5:4–10. https://doi.org/10.2174/1874104501105010004

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

2. CAPÍTULOS

2.1. Avaliação de genes de referência em seis diferentes estágios de desenvolvimento de *Schistosoma mansoni* para RT-PCR quantitativo

PREÂMBULO

Contribuições do Doutorando Gilbert de Oliveira Silveira para o manuscrito apresentado nesta sessão:

Design dos experimentos, execução dos experimentos de bancada, execução dos experimentos *in silico* relacionados com os três modelos de análise estatística para avaliação da estabilidade da expressão dos genes nos diversos estágios de vida do parasita, análise dos dados, escrita do primeiro rascunho do manuscrito, revisão do manuscrito.

CCC RightsLink

Help Live Chat

SPRINGER NATURE

Assessment of reference genes at six different developmental stages of Schistosoma mansoni for quantitative RT-PCR

Author: Gilbert O. Silveira et al
Publication: Scientific Reports
Publisher: Springer Nature
Date: Aug 19, 2021

Copyright © 2021, The Author(s)

Creative Commons

This is an open access article distributed under the terms of the Creative Commons CC BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article.
To request permission for a type of use not listed, please contact Springer Nature

© 2022 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Data Security and Privacy | For California Residents

scientific reports

Check for updates

OPEN

Assessment of reference genes at six different developmental stages of *Schistosoma mansoni* for quantitative RT-PCR

Gilbert O. Silveira[1,2,3], Murilo S. Amaral[1,3], Helena S. Coelho[1], Lucas F. Maciel[1], Adriana S. A. Pereira[1,2], Giovanna G. O. Olberg[1], Patricia A. Miyasato[1], Eliana Nakano[1] & Sergio Verjovski-Almeida[1,2]✉

Reverse-transcription quantitative real-time polymerase chain reaction (RT-qPCR) is the most used, fast, and reproducible method to confirm large-scale gene expression data. The use of stable reference genes for the normalization of RT-qPCR assays is recognized worldwide. No systematic study for selecting appropriate reference genes for usage in RT-qPCR experiments comparing gene expression levels at different *Schistosoma mansoni* life-cycle stages has been performed. Most studies rely on genes commonly used in other organisms, such as *actin*, *tubulin*, and *GAPDH*. Therefore, the present study focused on identifying reference genes suitable for RT-qPCR assays across six *S. mansoni* developmental stages. The expression levels of 25 novel candidates that we selected based on the analysis of public RNA-Seq datasets, along with eight commonly used reference genes, were systematically tested by RT-qPCR across six developmental stages of *S. mansoni* (eggs, miracidia, cercariae, schistosomula, adult males and adult females). The stability of genes was evaluated with geNorm, NormFinder and RefFinder algorithms. The least stable candidate reference genes tested were *actin*, *tubulin* and *GAPDH*. The two most stable reference genes suitable for RT-qPCR normalization were Smp_101310 (*Histone H4 transcription factor*) and Smp_196510 (*Ubiquitin recognition factor in ER-associated degradation protein 1*). Performance of these two genes as normalizers was successfully evaluated with females maintained unpaired or paired to males in culture for 8 days, or with worm pairs exposed for 16 days to double-stranded RNAs to silence a protein-coding gene. This study provides reliable reference genes for RT-qPCR analysis using samples from six different *S. mansoni* life-cycle stages.

Along with malaria, schistosomiasis is one of the most important parasitic diseases due to its high impact on morbidity and mortality rates, affecting more than 230 million people in 76 different tropical and subtropical countries[1]. The three main species that cause schistosomiasis are *Schistosoma japonicum*, *S. haematobium* and *S. mansoni*, the latter being prevalent in Africa and Latin America. In America, it is estimated that *S. mansoni* infects 1–3 million people, and over 25 million live in risk areas, being Brazil and Venezuela the most affected[2]. Sexual dimorphism is one of the main characteristics of these blood flukes. Furthermore, blood flukes have a complex developmental life cycle, comprising at least seven different stages (eggs, miracidia, sporocysts, cercariae, schistosomula, adult males and adult females). Life-cycle progression also requires two hosts, the definitive mammalian and the invertebrate Mollusca host from the genus *Biomphalaria*, resulting in complex development and making schistosomiasis very hard to control and prevent[3].

With the advent of the *S. mansoni* transcriptome[4] and genome[5,6] sequences, the field was open for a large number of transcriptomic studies reported in the last two decades, enlightening gene expression patterns in different life-cycle stages[7–13], as well as in tissues[14,15], organs[16] and single-cells[17,18]. A new *S. mansoni* transcriptome and genome version 7 have recently become available[19]. In addition, we built a new transcriptome assembly[20] that merged all known protein-coding genes[19] with long non-coding RNAs expressed by the parasite, which were identified using 633 publicly available *S. mansoni* RNA-Seq libraries[20]. The way is paved, therefore, for

[1]Laboratório de Parasitologia, Instituto Butantan, São Paulo, SP 05503-900, Brazil. [2]Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, SP 05508-900, Brazil. [3]These authors contributed equally: Gilbert O. Silveira and Murilo S. Amaral. ✉email: verjo@iq.usp.br

further functional characterization of protein-coding and long non-coding RNA genes across the parasite life cycle. For all transcriptomic and functional gene studies, the expression levels of each gene of interest found in the large-scale analyses need to be first confirmed by a more specific method such as RT-qPCR before further functional studies are applied.

Reverse transcription-quantitative real-time polymerase chain reaction (RT-qPCR) relies on two different ways to quantify gene expression levels: absolute or relative quantification. Relative quantification is widely used because it is less sensitive to sample preparation, RNA quality, and cDNA synthesis. However, a careful selection of appropriate reference genes (also known as housekeeping genes) for adequate[21] normalizations is required before relative quantification RT-qPCR is applied.

Reference genes chosen for RT-qPCR normalization are frequently associated with housekeeping functions such as glycolysis, respiration, cell transport, and cytoskeleton, with *actin*, *tubulin* and *GAPDH* being some of the most frequently used genes. Nonetheless, pieces of evidence pointing out that those genes may exhibit varying expression levels in different stages/tissues or conditions tested are emerging[22]. Thus, it is now globally accepted that it is necessary to search for stable reference genes in the condition that is being specifically tested[22], and that more than one reference gene should be used in the RT-qPCR normalization steps[23]. In *S. japonicum,* the identification of reference genes that are not differentially expressed in four life stages of the parasite has been accomplished, suggesting PSMD4 (*26S proteasome subunit*), TPC2L (*longin-like protein*), and NDUFV2 (*core subunit of respiratory chain Complex I*) as the most stable genes[24].

As for *S. mansoni*, specific reference genes have been pointed so far only to study pairing-dependent gene expression in males[25]. No report has used high-throughput data to select specific reference genes for correct normalization of RT-qPCR assays comprising samples from different life-cycle stages in *S. mansoni*. Noteworthy, about 8% (1164 genes) out of the 14,548 protein-coding genes in the *S. mansoni* genome v.7 is annotated as "Hypothetical protein" (949 transcripts) or "Uncharacterized protein" (215 transcripts). In this context, as a first step towards the characterization of an uncharacterized gene of interest, it is important to have its correct quantification by RT-qPCR among the different *S. mansoni* life-cycle stages using a set of reliable reference genes.

In the present study, we selected and tested a group of twenty-five candidate reference genes for RT-qPCR assays across six *S. mansoni* life-cycle stages. Our selection was based on their expression stability, which was detected by using publicly available data of *S. mansoni* RNA-Seq libraries[18,26–29] and by using four different methods (DESeq2[30], TMM/CPM[31], UQ[32], and TPM[33]) to normalize and compare the large-scale expression data across the different samples. In addition, eight reference genes used by previous publications for the normalization of data from at least three different life-cycle stages of *S. mansoni* were also selected. Altogether, 33 candidate reference genes were selected and tested for their expression stability in a series of RT-qPCR experiments with RNA samples extracted from six different *S. mansoni* developmental stages (eggs, miracidia, cercariae, 48-h-schistosomula, adult males and adult females). The geNorm, NormFinder and RefFinder algorithms were used for evaluation of the candidate genes' stability. The most stable genes were identified and pointed as reliable reference genes for use in RT-qPCR analyses among six different *S. mansoni* stages. These reference genes were also tested under two other assay conditions, the first involving adult females maintained in culture for 8 days as unpaired worms or as paired female/male couples, the second involving adult worm pairs treated with dsRNAs to silence the protein-coding gene *EED* (*Embryonic Ectoderm Development*), a component of the histone modifying complex Polycomb Repressive Complex 2 (PRC2), which tri-methylates H3K27.

Results

Identification of reference genes for the study by RT-qPCR of gene differential expression across *S. mansoni* life-cycle stages. To identify reference genes suitable for the study by RT-qPCR of gene differential expression across life-cycle stages, we selected RNA-Seq libraries from previous works[18,26–29] available at the Sequence Read Archive (SRA) databank. These libraries had to accomplish a minimum of 50% aligned reads, having at least 4 million aligned reads in each library (Supplementary Table S1). Since most of the *S. mansoni* RNA-Seq data available is from schistosomula, adult male and female stages, we prioritized the six RNA-Seq libraries from each of these stages with the highest numbers of aligned reads and the highest percentage of aligned reads. For miracidium, there was only one available library, while two libraries were available from sporocysts. Thus, those three libraries were analyzed as a single group of Miracidium/Sporocyst (M/S). Exceptionally, library SRR11248243 (from sporocysts) had a 24.65% read alignment rate but was kept in the analysis because it had more than 28 million reads aligned. For cercariae, just three libraries passed the two cut-offs of the number of aligned reads and percentage of aligned reads. Overall, 24 RNA-Seq libraries that comprised miracidium/sporocysts (n = 3), cercariae (n = 3), schistosomula (n = 6), adult males (n = 6) and adult females (n = 6) (Supplementary Table S1) were used in our analyses.

We next established a bioinformatics pipeline to analyze the selected RNA-Seq libraries and search for the best candidate reference genes (see "Methods"). Among the RNA-seq data analysis steps, normalization is crucial to accurately interpret the results of transcriptomic experiments. Thus, we used the following four different normalization methods: DESeq2[30], Trimmed mean of M values (TMM.CPM)[31], Upper Quartile (UQ)[32] and Transcripts per Million (TPM)[33]. The expression values for all *S. mansoni* transcripts in each tested normalization method are provided in Supplementary Tables S2–S5 for DESeq2, TMM.CPM, UQ, and TPM, respectively. A visualization platform for the expression values obtained with each normalization method was created and is available at https://verjolab.shinyapps.io/Reference-genes/.

The first step to identify the most stable reference genes was to calculate for each of the 13,624 *S. mansoni* protein-coding genes (Smps) under analysis, the coefficient of variation across all libraries (CV = Standard Deviation of expression values across all libraries/Average of expression values across all libraries). The calculation was repeated for each of the four normalization methods. The analysis proceeded with the identification, for each

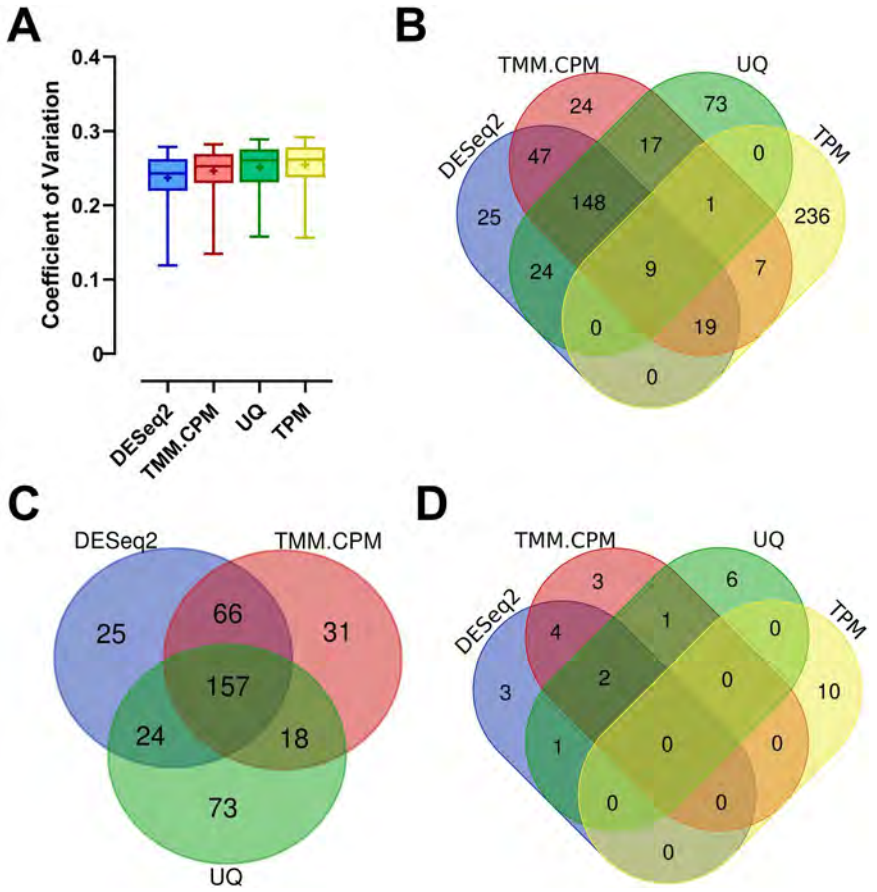


Figure 1. Transcriptome-wide analysis of the stability of candidate reference genes for quantitative RT-PCR assays across six life-cycle stages of *S. mansoni*. Twenty-four RNA-Seq libraries comprising six *S. mansoni* life-cycle stages were analyzed using four different normalization methods, namely DESeq2, TMM.CPM, UQ, and TPM. (**A**) Box plots of the coefficients of variation (CV), which were calculated based on the expression levels of the 272 genes with the lowest coefficients of variation, representing 2% of all genes detected in the RNA-Seq libraries, and analyzed with each of the four normalization methods indicated in the x-axis. The horizontal line represents the CVs' median for each normalization method, while the + signal represents the mean. The boxes and whiskers represent the inter-quantile and min to max ranges, respectively. (**B**) Venn Diagram shows the number of genes and the overlap among the 272 most stable genes found in each of the four RNA-Seq normalization methods. (**C**) Venn Diagram showing the number of genes and the overlap among the 272 most stable genes in the DESeq2, TMM.CPM and UQ normalization methods. (**D**) Venn Diagram showing the number of genes and overlap among the ten most stable genes according to their CVs in each RNA-Seq normalization method tested. *UQ* upper quartile, *TMM* trimmed mean of M-values, *CPM* counts per million, *TPM* transcripts per million.

normalization method, of the 272 Smps comprising the top 2% of all genes that showed the lowest CVs as the most stable ones. The means and medians of the coefficients of variation across these 272 Smps were similar for the four different normalization methods (Fig. 1A), with no significant differences. However, the group of 272 genes obtained in each normalization method was different, especially for the TPM normalization method. From all 272 Smps belonging to the 2% genes with the lowest CVs in the TPM normalization method, only 36 Smps (13% of them) were also present in at least one of the other three normalization methods (Fig. 1B).

A comparison among DESeq2, TMM.CPM and UQ normalization methods showed that 157 (or 58%) of the 272 Smps were common to all three normalization methods (Fig. 1C). Less than 27% of the 272 Smps in each method were not present in at least one of the other two normalization methods (Fig. 1C). With that in mind, we selected the top ten most stable Smps (those genes with the lowest CVs) from each normalization method to be compared against each other. We found out that at least 4 Smps identified in each of the three methods as the best candidate reference genes, out of top 10 genes, were common among DESeq2, TMM.CPM and UQ (Fig. 1D), resulting in a list of 20 unique Smps comprising all three normalization methods. Furthermore, all the ten most stable candidate reference genes identified with the TPM method did not overlap with any of the top 10 genes from the other three methods (Fig. 1D).

Thus, we selected for RT-qPCR assays the 20 candidate reference genes that comprise the set of unique genes in the overlap among the three lists of the top ten most stable genes in the analyses with DESeq2, TMM.CPM and UQ (Fig. 1D), plus 5 candidate reference genes comprising the top, most stable genes from the TPM

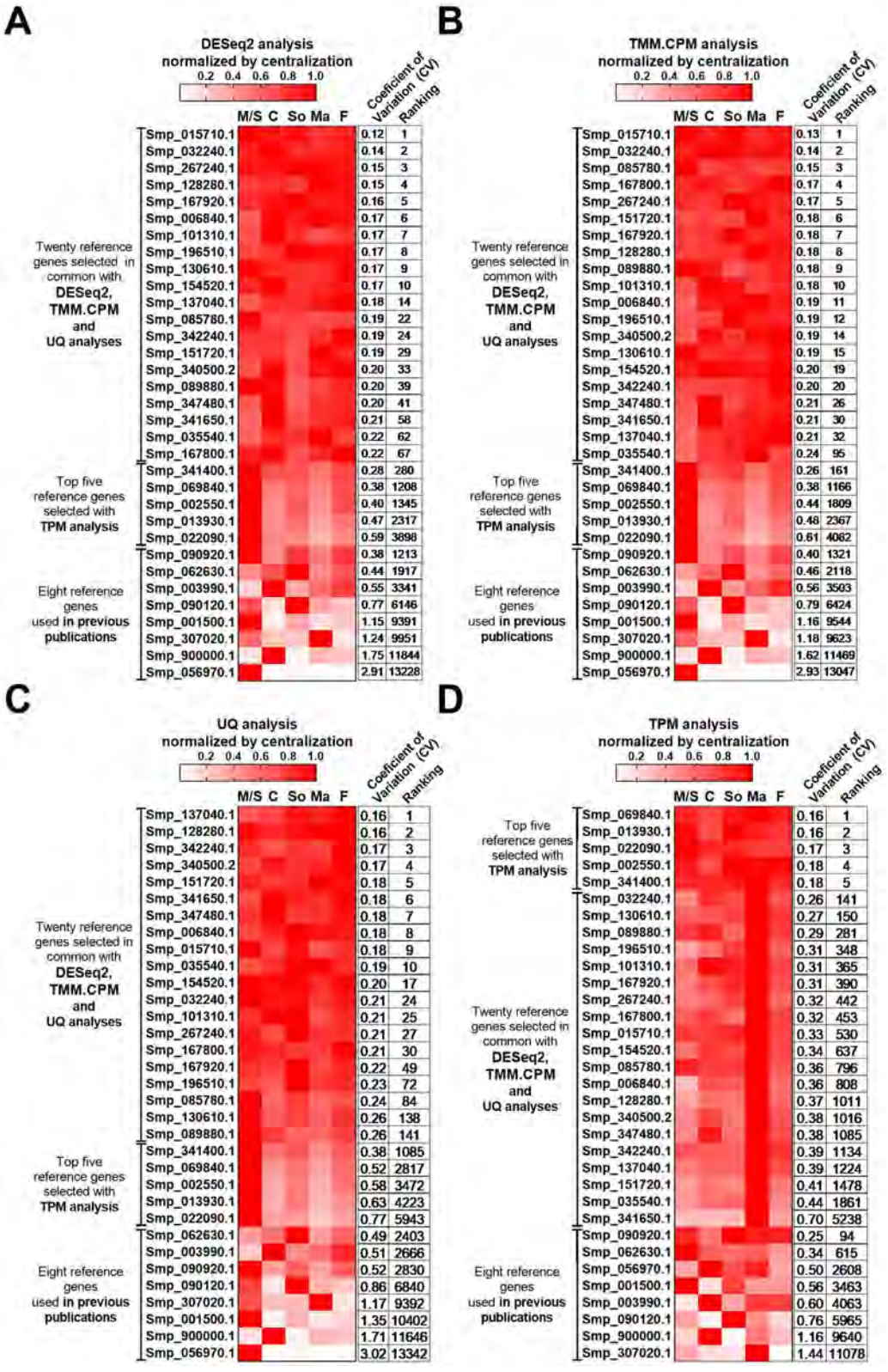


Figure 2. Expression patterns of candidate reference genes at different *S. mansoni* life-cycle stages. Twenty-five candidate reference genes with the lowest coefficients of variation of their expression values were selected to evaluate their stability by RT-qPCR assays. The selection was based on the analysis of 24 RNA-Seq libraries from six different *S. mansoni* life-cycle stages (see "Methods"). In addition, eight candidate reference genes commonly used in previous publications were also selected. The heatmaps were built based on the average expression values of each gene from all libraries used for each life cycle stage analyzed and were calculated for each RNA-Seq normalization method, namely (**A**), DESeq2; (**B**), TMM.CPM; (**C**), UQ; and (**D**) TPM. Smp stands for *S. mansoni* protein, and the codes are related to the gene annotation. *M/S* miracidia/sporocysts, *C* cercariae, *S* 48-h-schistosomula, *Ma* adult males, *F* adult females. The expression coefficient of variation and each gene's ranking across the different RNA-Seq normalization methods are reported on the two columns at right in each panel.

Gene product name	Smp Code[a]	Forward primer sequence (5′–3′)	Reverse primer sequence (5′–3′)	Mean Cq[a]	Refs.[a]
NCK-interacting protein with SH3 domain	Smp_128280.1	CTCTGCGATCTGTTCTCTTACTC	TGAGAAGTAGCAGTAAATGAGGC	21.7	This paper
6-phosphofructo-2-kinase/fructose-2,6-bis-phosphatase 1	Smp_015710.1	TATGGAATTAGAACGACAAACATCAG	GCAGTCGGTGTTAATTTGAATACAG	27.2	
Histone H4 transcription factor	Smp_101310.1	AAGTCAACCGATCCAGTTCTAC	TCTGCTTGAACATGTGGTAAGG	22.5	
Mitotic-spindle organizing protein 2B	Smp_032240.1	GACCGTCCCAAATGATGTTG	TGACATCTGGACTGCAACC	20.2	
Ubinuclein-2	Smp_167920.1	TTTAGTGTCCAAGGCGGTAC	ACCTTTATTTGAGCAGTGGGAG	20.8	
Palmitoyltransferase ZDHHC3	Smp_267240.1	AAGCTTATCCAGATGGTAGCG	GGCTTGAAAAGATATGAACATGGTAG	19.8	
Protein UBASH3A homolog	Smp_151720.1	GGCGATGGGTCGTTCAGATT	GCAGTACGATCACAATAGCCT	20.9	
UPF0060 membrane protein ESA_01751	Smp_006840.1	GTCCTAAATGTCGTAATGTGGC	AAAATACACCTACCAAAAGCTGC	19.5	
39S ribosomal protein L10, mitochondrial	Smp_085780.1	CTGTTTTCTGAGCCTACTCCTG	CCTTGAGATGACAATGCAGC	26.6	
FAD-dependent oxidoreductase domain-containing protein 1	Smp_089880.1	ACTGGTTTTCATCCATTTCATACAAA	ACGAAAAACGTACTAGATCTATTGTTT	31.0	
Zinc finger CCHC domain-containing protein 7	Smp_167800.1	GATAACCGAACTCATTCTTTGTCTTAC	CGCTGTTTTATGGCCTTCTG	25.7	
Protein FAM60A	Smp_154520.1	ATGACCCATAATCCCGAAGTG	AGCAGCAGTATGAATGGTACG	20.5	
tRNA (guanine-N(7)-)-methyltransferase	Smp_130610.1	AAAGCATTGCGTGTGAAGC	CACGTTTAAAATGAGGGTCTGG	26.1	
Ubiquitin recognition factor in ER-associated degradation protein 1	Smp_196510.1	GCGGTACAGGTTATCGGTTAG	ACTTCCAGGTTGATAATTGTAGTTTG	18.8	
Zinc finger SWIM domain-containing protein 3	Smp_035540.1	TGGTGTGTGTACTTGTTCGC	ATGAGATCTAGACCACCGCG	22.2	
Vacuolar protein sorting-associated protein 51 homolog	Smp_340500.2	CGCGAATCATTACTTTTGGGTG	ACGAAATGCCACAGATGAATTTG	21.5	
Uncharacterized aarF domain-containing protein kinase 5	Smp_341650.1	TGTCGTTCTCATGGTGATCC	ACTTGACCTATAGCATATAACAATTG	20.6	
Ubiquitin-protein ligase E3B	Smp_137040.1	GCGTAATTCGATGGTTATGGG	CAAGATTAGCGAACCCCAAAAG	21.4	
Zinc finger and BTB domain-containing protein 17	Smp_347480.1	TGATGGTGACGTTGCTGTAG	TCGACTCTTCCTAATTCTGACGA	22.4	
Transcription factor tau subunit sfc6	Smp_342240.1	GATGCAGAGCTTCAACACTTTC	AAATCCAATTCAGGTTTCTTAGGTAC	24.5	
Ankyrin repeat domain-containing protein 49	Smp_069840.1	CAGCGGGTGCAGATTTAAATG	AATGTCTCAGGTCCTTGGTTG	23.8	
SAP30-binding protein	Smp_013930.1	TTACTAGCACAGACAAACCCAG	ATTCTGAGGATGTGTCATGGG	21.8	
Ribose-phosphate pyrophosphokinase 1	Smp_022090.1	ACATGCCAGTCAAATTCAAGGA	ACATCTCCAACAAGAACCATACG	22.1	
RNA-binding protein 8A	Smp_002550.1	TCCGTAGAAGGTTGGATCTTGT	TAACCAGTTCGCCGATCAAG	20.6	
Nuclear factor related to kappa-B-binding protein	Smp_341400.1	GAAATTGTTCGAGACGCTGTG	TCGATCCAACGCAGAACTAAC	20.2	
Mitochondrial 28S ribosomal protein S14	Smp_090920.1	CACCAGCTCATCATAATAATCCA	TAGCATCCTGAAAGCCACGA	20.2	20,26,34
Putative dynactin subunit	Smp_062630.1	GGAATGATGTGGCCGATAGT	CGCAGAGATTGGCTAAATTG	19.2	20,26
Eukaryotic translation initiation factor 4e	Smp_001500.1	TGTTCCAACCACGGTCTCG	TCGCCTTCCAATGCTTAGG	20.9	35,36
Glyceraldehyde-3-Phosphate dehydrogenase	Smp_056970.1	TGAGGAAATCAAGGCTGCAGT	CCCTTCAATGGTCCAGATGC	17.7	37–39
Triosephosphate isomerase	Smp_003990.1	CATACTTGGACATTCTGAGCGTAGA	ACCTTCAGCAAGTGCATGTTGA	20.5	40,41
α-Tubulin	Smp_090120.1	CCATTTATGATATTTGTCGACGGA	TTTGTGTAGGTTGGACGCTCTATATCTA	19.3	41–46
Cytochrome c oxidase subunit I	Smp_900000.1	TACGGTTGGTGGTGTCACAG	ACGGCCATCACCATACTAGC	13.9	47–49
Actin	Smp_307020.1	CGTTGGACGACCTCGACAT	TGTCTTTCTGACCCATACCAACC	16.2	38,50

Table 1. Information on the RT-qPCR amplification characteristics of the 33 candidate reference genes. [a]Smp code stands for *Schistosoma mansoni* protein-coding gene ID. Mean Cq stands for Mean Corrected Cq, as described in the Methods. Refs. stands for the previous publications in which the corresponding primer pair of a reference gene was used.

normalization method. We also selected eight genes used in previous works as reference genes across different *S. mansoni* life-cycle stages, totalizing 33 candidate reference genes to be tested. Information regarding the eight previously reported reference genes, including the life-cycle stages in which they were measured, and if the gene expression change identified was validated by another method (such as western blotting or northern blotting), is reported in Supplementary Table S6.

The ranking for all 33 selected candidate reference genes is presented in Fig. 2, along with a heatmap of their expression values determined by DESeq2 (Fig. 2A), TMM.CPM (Fig. 2B), UQ (Fig. 2C), and TPM (Fig. 2D) normalized by centralization (which represents the expression of that gene in each of the life-cycle stages). The Log_{10} transformed expression values for each of the 33 candidate reference genes across all analyzed libraries are shown in Supplementary Fig. S1 for each of the normalization methods used.

Pairs of primers were designed to amplify each of the 25 candidate reference genes selected based on our RNA-seq data analyses. Their amplification efficiencies were measured using a dilution series of cDNA from male and female *S. mansoni* adult worms. We also measured the primers' amplification efficiency for the other

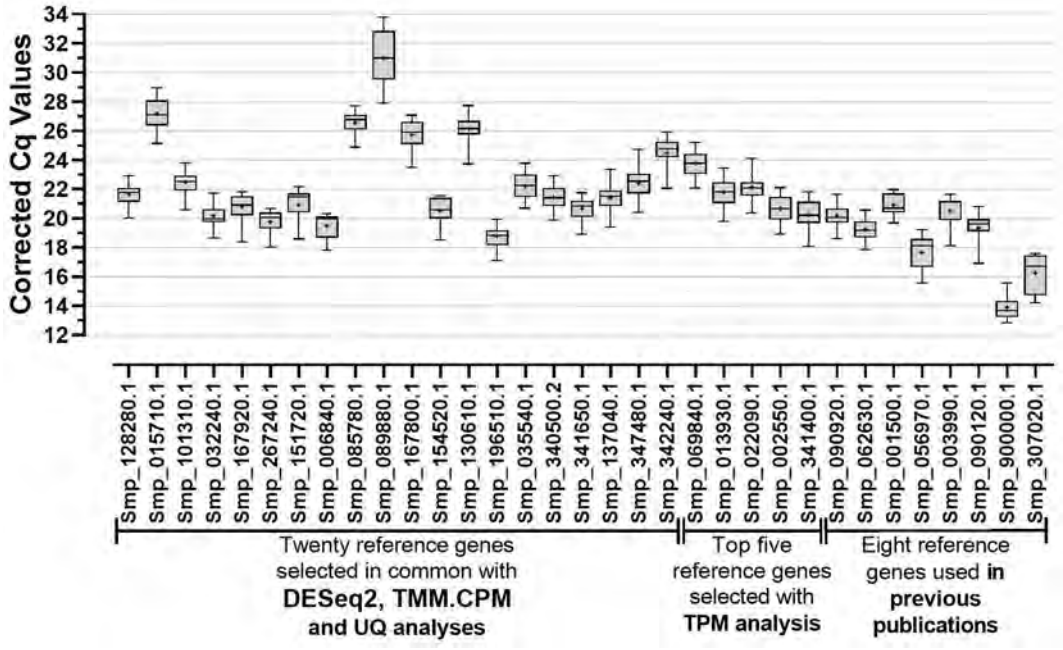


Figure 3. Gene expression analysis performed by RT-qPCR for all 33 candidate reference genes across six different *S. mansoni* life-cycle stages. The data are presented as corrected RT-qPCR quantification cycle (Cq) values as described in the Methods. The Cq values derive from the quantitative RT-qPCR analysis of the genes' expression in the following *S. mansoni* life-cycle stages: eggs, miracidia, cercariae, 48-h-schistosomula, adult males, and adult females. For each gene, the horizontal line represents the median while the + signal represents the average. The boxes and whiskers represent the inter-quantile and min to max ranges, respectively.

eight reference genes used in previous works. Amplification efficiencies for all the primer pairs were higher than 92.6%. Their RT-qPCR amplification characteristics are presented in Supplementary Table S7. Gene annotation of all 33 genes, primer sequences and their mean Cq values are given in Table 1.

Expression levels and stability of candidate reference genes measured in RT-qPCR assays. RNAs extracted from six different life-cycle stages of *S. mansoni* (eggs, miracidia, cercariae, 48-h-schistosomula, adult males and adult females), were used in RT-qPCR assays. The RNA integrity of samples in each of the four biological replicates of each *S. mansoni* life-cycle stage was confirmed as being very good, as visualized by their Electropherogram Summary in Supplementary Fig. S2. The amplicon specificities of the 33 candidate reference genes were confirmed by the presence of a single peak in the melting curve analysis obtained at the end of the real-time RT-qPCR assay (Supplementary Fig. S3).

The quantification cycle (Cq) values for each candidate reference gene were retrieved from each analyzed sample's real-time PCR amplification curves. Our real-time RT-qPCR gene expression analysis in eggs, miracidia, cercariae, 48-h-schistosomula, adult males, and adult females showed a Cq in the range 13–34 across all candidate reference genes (Fig. 3). The previously reported reference genes *Cytochrome C oxidase subunit I* (Smp_900000.1), *Actin* (Smp_307020.1) and *GAPDH* (Smp_056970.1) exhibited the highest expression levels (lowest Cq values) (Fig. 3). Conversely, the genes with the lowest expression levels (highest Cq values) were *FAD-dependent oxidoreductase domain-containing protein 1* (Smp_089880.1), *6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 1* (Smp_015710.1), and *39S ribosomal protein L10, mitochondrial* (Smp_085780.1) (Fig. 3 and Supplementary Table S8). Of note, Smp_089880.1 showed a mean corrected Cq value across the six developmental stages higher than 30 (Cq = 31.0) and, therefore, is not recommended to be used as a reference gene.

The Stability Score (M value) of each of the 33 candidate reference genes was calculated using the geNorm software (version 3.5)[21]. This software recommends using a stability score below the threshold of 1.5 to correctly identify reference genes with stable expression. Notably, Smp_341650.1 (M = 0.28), Smp_101310.1 (M = 0.28), and Smp_128280.1 (M = 0.28) were the most stably expressed genes with M values way lower than those from the previously used reference genes (Fig. 4A). The genes with the highest M values (i.e., least stable genes) were Smp_089880.1 (*FAD-dependent oxidoreductase domain-containing protein 1*), Smp_307020.1 (*Actin*), and Smp_900000.1 (*Cytochrome C oxidase subunit I*) (Fig. 4A). Notably, apart from Smp_062630.1, a *Putative dynactin subunit*, all the reference genes used in previous publications were amongst the least stable candidates, as shown by black bars in Fig. 4A. One of the candidate reference genes selected from the TPM normalization method was determined as the least stable candidate (Smp_089880.1). The best scored from this group ranked sixth place (Smp_069840.1, an *Ankyrin repeat domain-containing protein 49*).

Pairwise variation analysis demonstrated that the combination of Smp_101310.1 and Smp_341650.1 was sufficient for accurate gene expression normalization (V2/3 = 0.084) across the six developmental stages (Fig. 4B).

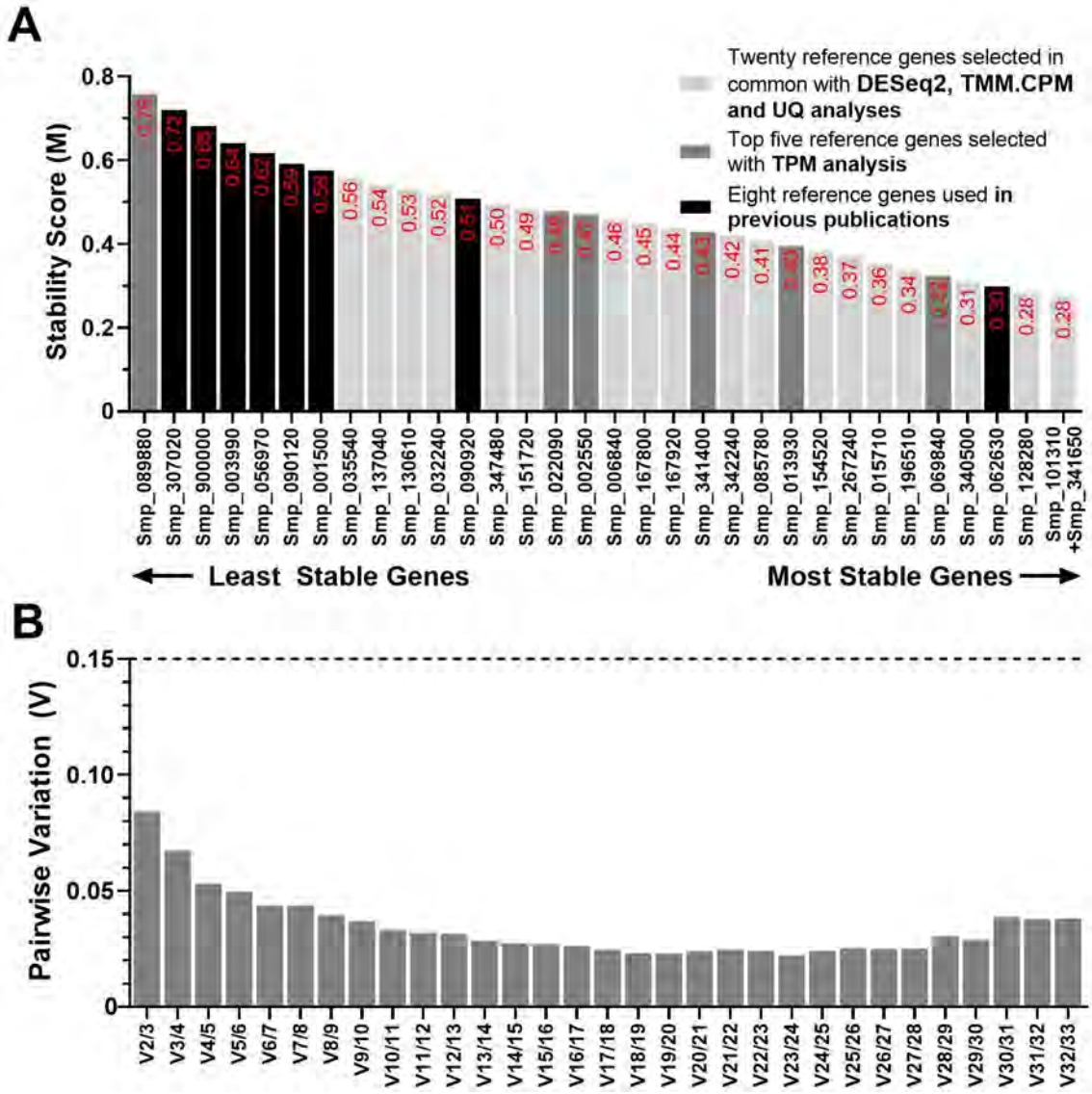


Figure 4. Gene expression stability and ranking of the 33 candidate reference genes across six *S. mansoni* developmental stages measured by RT-qPCR and analyzed with geNorm. (**A**) Expression stability scores (M) of all candidate reference genes are represented at the column bar's top. Genes were colored based on their selection method: genes selected with DESeq2, TMM.CPM and UQ RNA-Seq normalization methods are shown in light grey bars; genes selected with TPM RNA-Seq normalization method are shown in dark grey bars, while genes used in previous works are shown in black bars. The higher the stability score, the least stable the gene. (**B**) Pairwise variation (Vn/n + 1) analysis between the normalization factors NFn and NFn + 1 to determine the optimal number of reference genes required for accurate normalization, where n is the number of genes involved in the normalization factor. The cut-off value determined by this analysis is 0.15.

Thus, based on geNorm analysis, we conclude that *Histone H4 transcription factor* (Smp_101310.1) and *Uncharacterized aarF domain-containing protein kinase 5* (Smp_341650.1) were the most stable reference genes.

We used a second software, namely NormFinder[51], to evaluate the most stable candidate reference gene. Notably, the six best reference genes determined by this analysis were also predicted as the most stable genes by our analysis with DESeq2, TMM.CPM and UQ normalization methods (Fig. 5A). The three most stable reference genes were *Uncharacterized aarF domain-containing protein kinase 5* (Smp_341650.1, stability value = 0.11), *Ubiquitin recognition factor in ER-associated degradation protein 1* (Smp_196510.1, stability value = 0.18) and *39S ribosomal protein L10, mitochondrial* (Smp_085780.1, stability value = 0.19). Notwithstanding, as for the geNorm analysis, the least stable reference genes defined with NormFinder were *FAD-dependent oxidoreductase domain-containing protein 1* followed by the *Cytochrome C Oxidase subunit 1*, *Actin*, *GAPDH,* and *Tubulin* (Smp_089880.1, Smp_900000.1, Smp_307020.1, Smp_056970.1, and Smp_090120.1, respectively) (Fig. 5A). The most stable candidate reference gene determined with the TPM analysis was Smp_069840.1, an *Ankyrin repeat domain-containing protein* 49, that scored seventh with the NormFinder software.

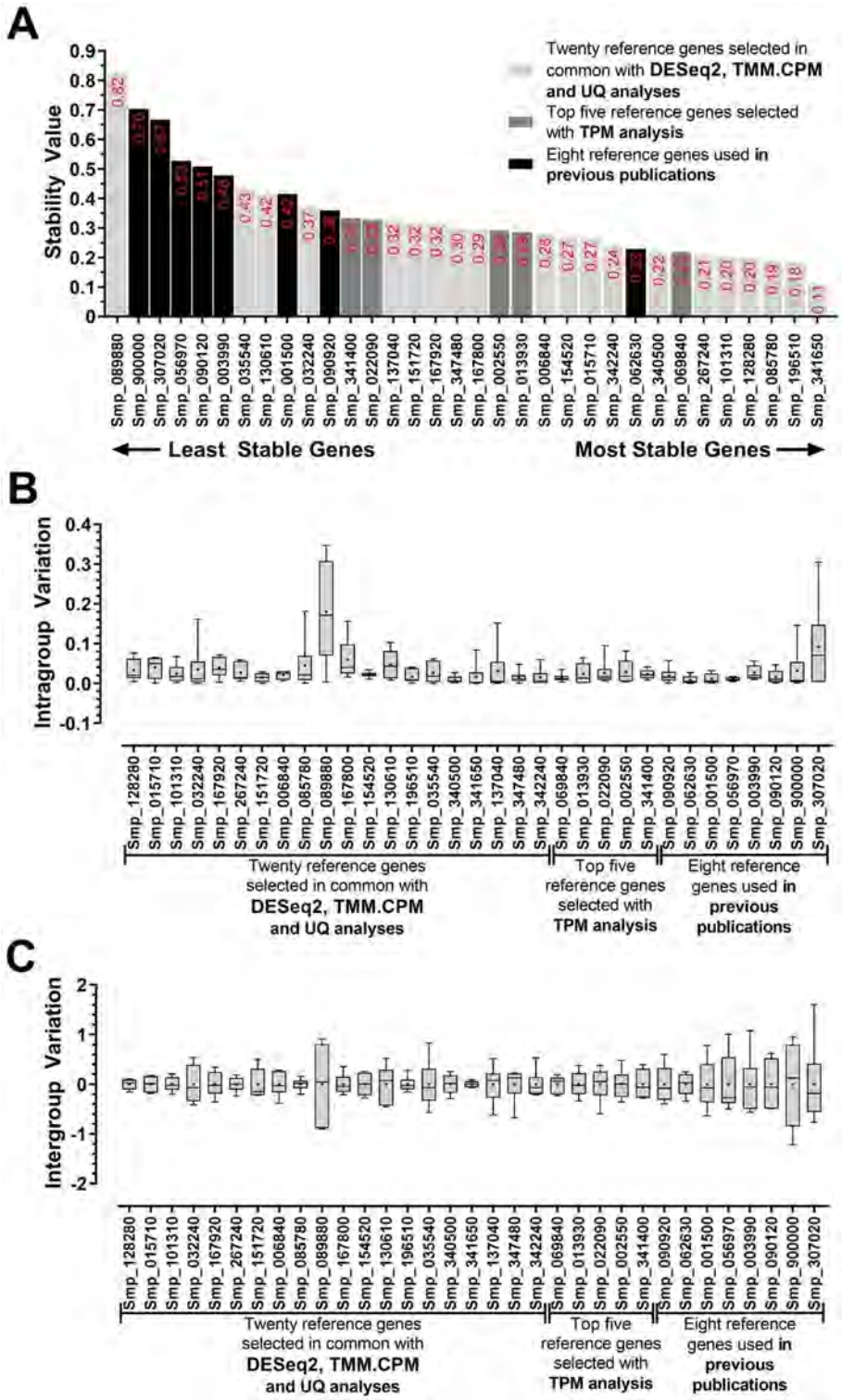


Figure 5. Gene expression stability and ranking of the 33 candidate reference genes across six *S. mansoni* developmental stages measured by RT-qPCR and analyzed with NormFinder. (**A**) The stability value of all candidate reference genes is represented at the column bar's top. Genes were colored based on their selection method: genes selected with DESeq2, TMM.CPM and UQ RNA-Seq normalization methods are shown in light grey bars; genes selected with TPM RNA-Seq normalization method are shown in dark grey bars, while genes used in previous works are shown in black bars. The higher the stability score, the least stable the gene. (**B**) Intragroup variation represents how much the biological replicates varied from each other in each of the six developmental stages. (**C**) Intergroup variation represents how much the average of the biological replicates of each stage varied among the six developmental stages. For each gene, the horizontal line represents the median while the + signal represents the average. The boxes and whiskers represent the inter-quantile and min to max ranges, respectively.

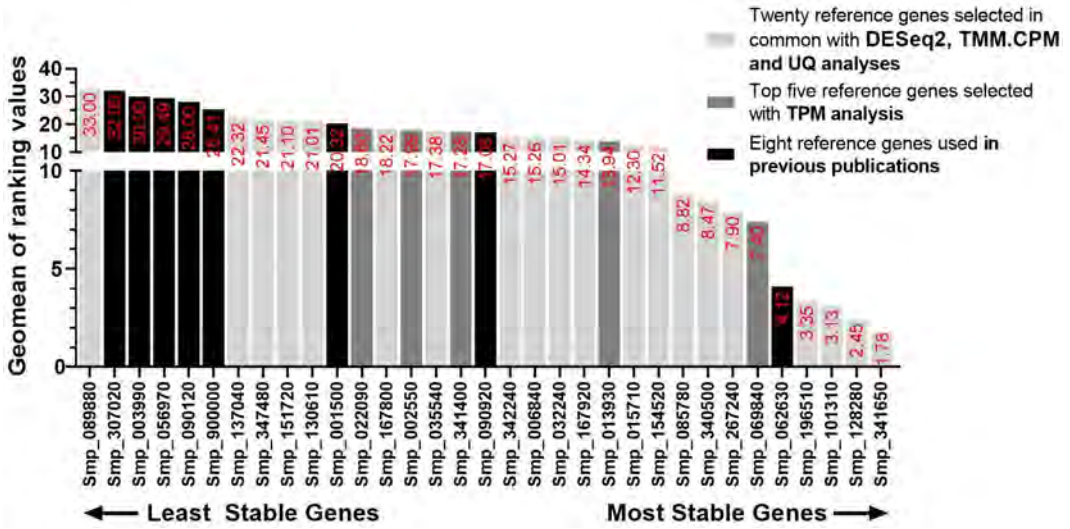


Figure 6. RefFinder comprehensive analysis with corrected Cp values retrieved from RT-qPCR of *S. mansoni* life-cycle stages samples. The geometric mean of all candidate reference genes' stability ranking values calculated by RefFinder is represented at the top of the column bar. Genes were colored based on their selection method: genes selected with DESeq2, TMM.CPM and UQ RNA-Seq normalization methods are shown in light gray bars, genes selected with TPM RNA-Seq normalization method are shown in dark gray bars, while genes used in previous works are shown in black bars. The higher the stability ranking value, the least stable the gene.

NormFinder also allows determining the Intra- and Intergroup variation, defined respectively as the variation among the sample biological replicates within each analyzed stage, and the variation among the stages of the biological replicates average of each different stage analyzed. Conspicuously, intragroup variation was less dispersed than intergroup variation for all candidate reference genes analyzed, but Smp_307020.1 and Smp_089880.1 stood out when it comes to intragroup variation, having a more dispersed expression within the biological replicates of the cercariae and schistosomula groups (Fig. 5B). On the other hand, intergroup variation with all normalization methods was small for almost all candidate reference genes selected by our analysis, except Smp_089880.1 (Fig. 5C). Simultaneously, the genes from previous works, *eIF4e*, *GAPDH, TPI, Tubulin, Cytochrome C oxidase subunit 1* and *Actin* (Smp_001500.1, Smp_056970.1, Smp_003990.1, Smp_090120.1, Smp_900000.1, and Smp_307020.1, respectively) were more dispersed than the other genes. The intergroup variation agreed with our ranking pointed in Fig. 2.

We used a third software, namely RefFinder[52], to identify the most stable reference gene. RefFinder uses Cq values from all samples and genes analyzed as input data and performs a group measurement from all available methods to determine the most to the least stable reference gene. As RefFinder does not take into consideration the primer efficiency for each candidate reference gene tested, we performed this analysis using the corrected Cq value (using the formula = $Log_2 E^{Cp\ value}$). RefFinder provides a ranking order for all the candidate reference genes for each of the four methods tested: DeltaCT, BestKeeper, NormFinder, and geNorm (Supplementary Table S9). In the end, the analysis also provides a comprehensive ranking value for each gene that takes into consideration the ranking for each of the four methods tested (Fig. 6). As expected, the ranking obtained by RefFinder was similar to our previous individual analyses with geNorm and NormFinder, pointing as the four most stable genes the ones that were selected based on the DESeq2, TMM.CPM and UQ RNA-Seq normalization methods (Smp_341650.1, Smp_128280.1, Smp_101310.1 and Smp_106510.1) (Fig. 6). In fifth place was Smp_062630.1, a *Putative dynactin subunit* previously used by us[20,26]; the *Ankyrin repeat domain-containing protein 49* (Smp_069840.1) scored sixth place (Fig. 6). The least stable candidate reference was *FAD-dependent oxidoreductase domain-containing protein 1* followed by *Actin*, *TPI*, *GAPDH*, *Tubulin,* and *Cytochrome C Oxidase subunit 1* (Fig. 6), which was all in agreement with the previously tested methods. This agreement could be confirmed when a Pearson correlation was calculated with the results from each of the three tested software (geNorm, NormFinder and RefFinder), one by one, with Pearson correlation coefficient values (r) above 0.93 (a positive linear correlation) and p-values < 0.001 for all cases (Supplementary Fig. S4).

Works in the literature recommend the selection of reference genes that possess different expression levels and biological functions[21,51,53]. This would reduce incorrect normalization eventually due to the reference genes being affected by any treatment to the samples. Furthermore, we have considered that genes without a predicted function or with putative annotation should be avoided to favor best-characterized genes. Unquestionably, Smp_341650.1 has consistently ranked as the best candidate in three of the tested methods. However, it is annotated as an *Uncharacterized aarF domain-containing protein kinase 5*. This protein's function is not yet clear, and no information of its protein kinase activity and what type of substrate it would phosphorylate has been provided until now. The second, consistently most stable reference gene was the *Histone H4 transcription factor* (Smp_101310.1), which was singled out here as the first recommended reference gene, followed by *Ubiquitin recognition factor in ER-associated degradation protein 1* (Smp_196510.1), due to their biological functions and

expression levels across all samples. Therefore, our work pointed to Smp_101310.1 and Smp_196510.1 as the reference genes to be used in RT-qPCR experiments with the six tested *S. mansoni* developmental stages.

Differential gene expression across six developmental stages using the reference genes identified in the present work. We evaluated the differential gene expression levels across the six different developmental stages of the eight reference genes previously used in the literature, using for normalization the geometric mean of the two reference genes recommended here: *Histone H4 transcription factor* (Smp_101310.1) and *Ubiquitin recognition factor in ER-associated degradation protein 1* (Smp_196510.1). All genes previously used in the literature as reference showed a statistically significant differential expression in at least one *S. mansoni* developmental stage tested (Fig. 7). Smp_062630.1, *Dynactin subunit 2* presented the lowest expression level fold-change (1.6-fold, expression level in miracidium compared with schistosomula, Fig. 7B). In contrast, the highest expression level fold-change (11-fold) was observed for Smp_307020.1, *Actin* (expression levels in males compared with eggs, Fig. 7H).

Evaluation of the performance of the reference genes identified in the present work under different conditions: female pairing status or gene silencing. We evaluated under two other different conditions the performance of the two most stable reference genes found in this work. The first condition involved adult females cultured in vitro for 8 days as unpaired worms or paired with males. The stability of the two normalizer genes identified here was compared with the stability of *Actin*, *GAPDH*, *Tubulin*, as well as of *LETM1* (Smp_065110) and *PSMB7* (Smp_073410), two genes identified as normalizers to study gene expression in *S. mansoni* males from single-sex and dual-sex infections[25]. RT-qPCR data were obtained for these genes from a total of 21 samples covering three biological replicates of females cultured for 0, 2, 4 and 8 days either paired with males or as separated female worms. Gene stability was calculated from the RT-qPCR data, which was analyzed with the three algorithms, geNorm, NormFinder and RefFinder (Supplementary Table S10); the analysis showed that *Ubiquitin recognition factor in ER-associated degradation protein 1* (Smp_196510.1) was identified by the three algorithms as the most stable normalizer gene, whereas *Histone H4 transcription factor* (Smp_101310.1) was identified as the second most stable normalizer by geNorm and RefFinder (Supplementary Table S10). We measured in females the pairing-dependent expression of the *eggshell protein p14* (Smp_316140.1)[54,55] using the two reference genes identified here (Fig. 8). Expression of *p14* was significantly reduced ~eightfold along the eight days in culture in unpaired females compared with paired ones (Fig. 8A). A similar expression pattern was obtained when using *Actin*, *GAPDH* and *Tubulin* as normalizers (Fig. S5A).

The second condition involved adult worm pairs treated in culture for 16 days with double-stranded RNAs that targeted and silenced the protein-coding gene *EED* (dsEED) (Smp_165220, *Embryonic Ectoderm Development*), a component of the Polycomb Repressive Complex 2 (PRC2), which is the histone modifying complex that tri-methylates H3K27. A non-related double-stranded RNA targeting the mCherry gene (dsmCherry) was used as a negative control. Again, the stability of the two reference genes identified here was compared with the stability of three reference genes from the literature, namely *Actin*, *GAPDH* and *Tubulin*, using RT-qPCR data obtained from 48 samples corresponding to three biological replicates of males and females exposed in vitro to dsEED or dsmCherry along 2, 4, 8 and 16 days in culture. The stability analysis showed that *Tubulin* (Smp_090120.1) was the most stable gene (Supplementary Table S11), whereas *Ubiquitin recognition factor in ER-associated degradation protein 1* (Smp_196510.1) and *Histone H4 transcription factor* (Smp_101310.1) were identified as the second and third most stable normalizers (Supplementary Table S11). Using the latter two normalizer genes, a significant knockdown of *EED* was detected by RT-qPCR upon the exposure of adult males and females in culture for 2, 4 and 8 days to dsRNAs targeting the *EED* gene compared with a control dsRNA (Fig. 8B). A similar expression pattern was obtained when using *Actin*, *GAPDH* and *Tubulin* as normalizers (Fig. S5B).

Discussion

A decade ago, the use of high throughput transcriptomic technologies, such as RNA-Seq, was restricted due to its high price and to the limited number of powerful informatics tools for proper analysis. Nowadays, high throughput expression analyses are usually just the starting point of major studies, providing us with a plethora of processed data. While large amounts of data have been explored, further steps to provide the functional characterization of candidate genes and gene families have to be performed gene by gene. The first step in this direction is to confirm gene expression patterns individually, and real time quantitative RT-PCR has been one of the most used techniques.

The raw read counts data from RNA-Seq experiments need to be converted into informative measurements of gene expression. Normalization is a crucial step towards resolving the variabilities that affect the number of reads mapped to a gene, such as the gene length[56], GC-content[57] and the sequencing depth[58]. These RNA-Seq library features have to be considered when performing comparisons between different samples, and the major source of variation between RNA-Seq data is the sequencing depth that is reflected by the library size (i.e., how many reads are generated from the experiment). Library size normalization relies on scaling raw read counts in each sample by a single sample-specific factor reflecting its library size. The most commonly used methods include Relative Log Expression that is the basis of DESeq2 analysis, Upper Quartile (UQ) and Trimmed Mean of M-values (TMM), and all these methods have been thoroughly reviewed[59–64]. Another frequently used RNA-Seq normalization method is TPM, which relies on both the gene length and the sequencing read length corrections. However, TPM has been pointed as a flawed normalization method for RNA-Seq normalization between highly different samples[56,65], and could be less appropriate for the available *S. mansoni* data at different life-cycle stages. Some recent evidence points to the misuse of TPM for normalization of RNA-Seq data[61,66]. Our RT-qPCR

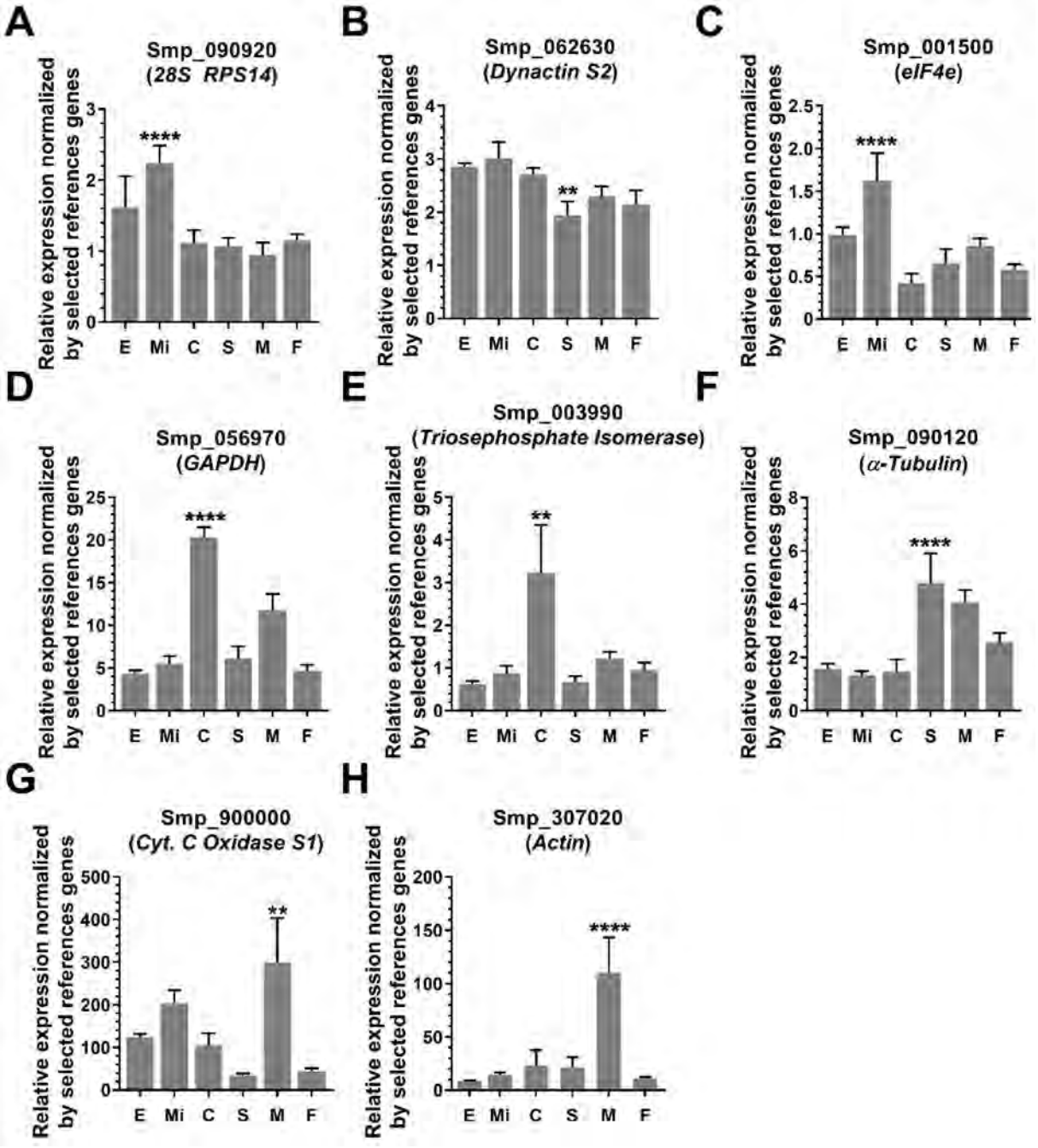


Figure 7. Relative expression at six different *S. mansoni* developmental stages of reference genes previously used in the literature, normalized by the two most stable reference genes found in the present work. (**A**) *26s Ribosomal Protein Subunit 14*; (**B**) *Dynactin subunit 2*; (**C**) *Eukaryotic translation initiation factor 4e*; (**D**) *Glyceraldehyde-3-Phosphate dehydrogenase*; (**E**) *Triosephosphate isomerase*; (**F**) *α-Tubulin*; (**G**) *Cytochrome C oxidase subunit* 1 and (**H**) *Actin*. Quantitative RT-qPCR was performed for the eight genes at six different parasite life-cycle stages, namely egg (E), miracidium (Mi), cercaria (C), 48-h-schistosomulum (S), adult male (M) and adult female (F) (x-axis). The expression values are represented as the relative expression using as normalizer the geometric mean of the two reference genes selected in this work, namely Smp_101310.1 and Smp_196510.1. Bars represent the standard deviation of the mean from four biological replicates for each stage. Three technical replicates were assayed for each of the four biological replicates. One-way ANOVA test was used to calculate the statistical significance of the expression differences among the parasite life-cycle stages (*p-value ≤ 0.05; **p-value ≤ 0.01; ***p-value ≤ 0.001; ****p-value ≤ 0.0001). For clarity purposes, we show only the highest p-value for the stage in which the Smp was detected with the highest expression difference among the life-cycle stages.

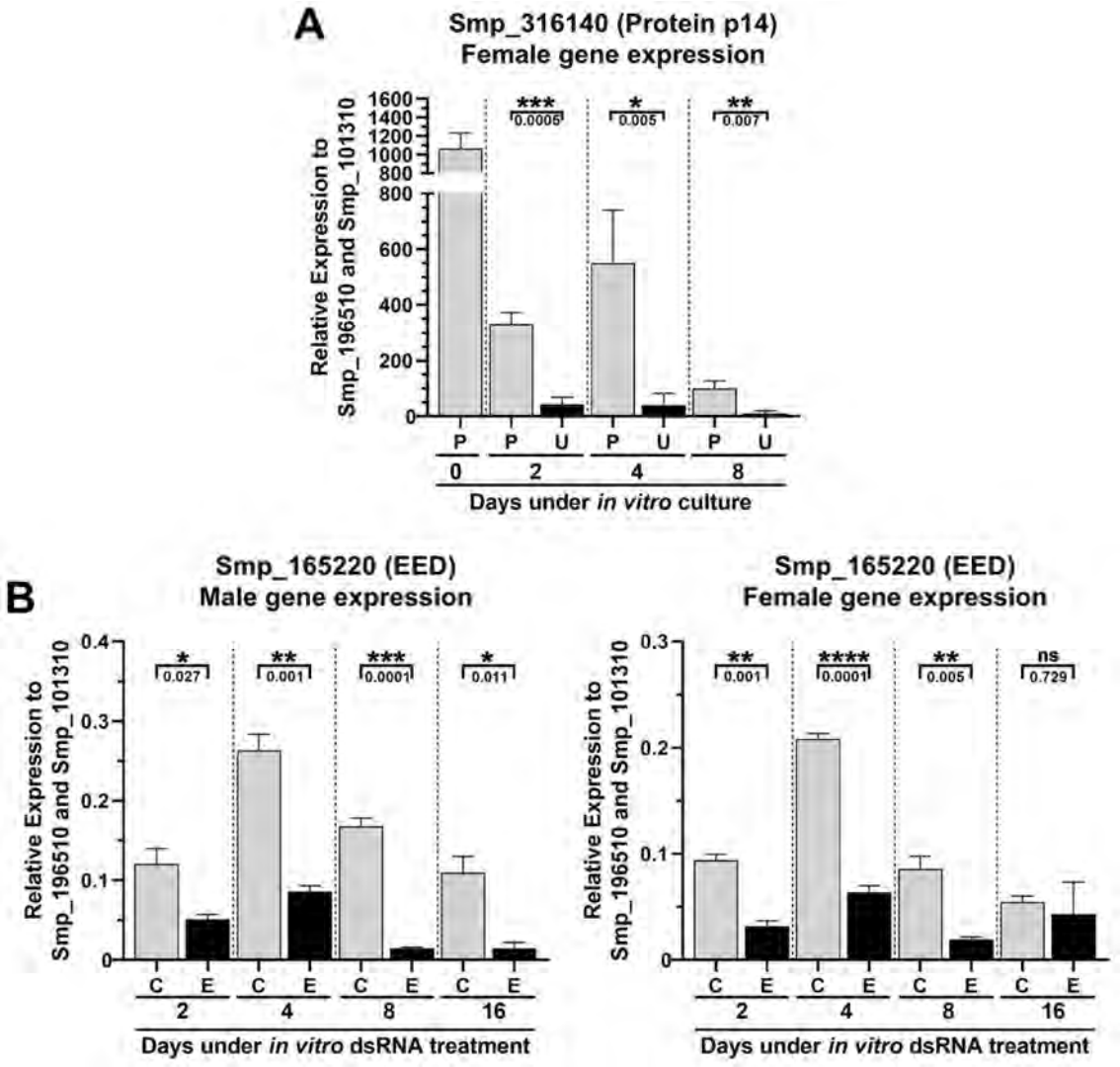


Figure 8. Relative expression of two protein coding genes in *S. mansoni* under different culturing conditions, normalized by the two most stable reference genes found in the present work. (**A**) Female adult worm gene expression pattern of Smp_316140 (Protein p14) across different in vitro culturing conditions. Quantitative RT-qPCR was performed with RNA samples from females that were paired (P) or unpaired (U) to males and cultivated in vitro for 2, 4 or 8 days. Day 0 stands for paired females retrieved right after perfusion. (**B**) Male and female adult worm Smp_165220 (EED) gene expression in samples obtained from couples treated with dsRNA targeting either the control unrelated dsmCherry gene (C) or the Smp_165220 (EED) gene (E) in vitro for 2, 4, 8 or 16 days. The expression values are represented as the relative expression using as normalizer the geometric mean of the two reference genes selected in this work, namely Smp_101310.1 and Smp_196510.1. Bars represent the standard deviation of the mean from three biological replicates for each experiment. Three technical replicates were assayed for each of the three biological replicates. Unpaired student t-test was used to calculate the statistical significance of the expression differences in the comparisons (*p-value ≤ 0.05; **p-value ≤ 0.01; ***p-value ≤ 0.001; ****p-value ≤ 0.0001; ns = p-value > 0.05). The p-value obtained from the Student's t-test is represented under the brackets.

results show that candidate reference genes from the TPM normalization method have been detected as the least stable reference genes for the correct normalization in samples from the six S. *mansoni* life-cycle stages tested.

We used three different and well-established RT-qPCR analysis tools that follow distinct strategies to determine the most stable reference genes. It has already been shown that the golden standard for selecting reference genes relies on using these three different approaches[67–70]. GeNorm[21] is based on paired dependent gene comparison. The expression level of each candidate reference gene is compared with the normalization factor, excluding the least stable genes one by one until the two genes with the highest stability (least variation) remain. NormFinder[51], on the other hand, determines intragroup and intergroup variation between the different sample groups to identify the genes with the highest expression stability. RefFinder[52] is a tool that gathers four computational programs (geNorm, NormFinder, BestKeeper and the comparative Delta-Ct method) to compare and rank

the tested candidate reference genes. Although it does not use primer efficiency for the ranking, this limitation can be overcome by using corrected Cq values.

We have defined Smp_101310.1 and Smp_196510.1 as the most stable reference genes in our expression stability analysis of the six different *S. mansoni* life-cycle stages derived from the RT-qPCR data. Smp_101310.1 (*Histone H4 transcription factor*) has been computationally annotated as a transcription factor that activates Histone H4 gene expression. In humans, the gene *Histone H4 Transcription Factor/HINFP* is categorized as having "Low tissue and cell specificities", is detected in all tissues and is expressed at similar levels in almost all tissues. Human HINFP is a critical component of a signaling pathway that controls expression of histone H4 genes[71]. Human HINFP has been shown to possess a HINFP-specific conserved region that is present in HINFP homologues of all metazoan species that have been examined[72]. Therefore, it is likely that in *S. mansoni*, Smp_101310 gene expression pattern is similar to that of human HINFP, with low differential expression along the tissues and stages of the parasite, which would be necessary to maintain the expression of histone H4 target genes.

Smp_196510.1 was computationally annotated as the *Ubiquitin recognition factor in ER-associated degradation protein 1*, an essential component of the ubiquitin-dependent proteolytic pathway that degrades ubiquitin fusion proteins. In humans, the gene *Ubiquitin Recognition Factor In ER Associated Degradation 1/UFD1* is categorized as having "Low tissue and cell specificities", is detected in all tissues and is expressed at similar levels in almost all tissues. In humans, Ufd1 facilitates the exportation of misfolded proteins from the endoplasmic reticulum to the cytosol for ubiquitin–proteasome pathway (UPP) degradation[73]. This regulated protein degradation by the ubiquitin–proteasome system plays a central role in diverse cellular processes. In *S. mansoni*, it has been already shown that functional UPP components are required for parasite development in the vertebrate host[74]. In fact, siRNA knockdown of the SmRPN11/POH1 proteasome subunit gene decreased schistosomula viability by 78%, suggesting an important role for UPP in *S. mansoni* development and survival[75]. Therefore, some of the genes that participate in the UPP, such as Smp_196510, are likely to be highly controlled and expressed in stable levels throughout the life-cycle stages of the parasite. On the other hand, an interesting observation arising from our work is that *GAPDH*[76], a key enzyme of glycolysis, *actin*[77], a cytoskeleton component, and *TPI* (*triosephosphate isomerase*)[78], another glycolysis component enzyme, were the least stable genes in our analysis. Human *GAPDH* gene is enriched in skeletal muscle and shows higher tissue and cell specificities. In *S. mansoni*, while *SmGAPDH* is found within all schistosome tissues, the protein has also been particularly identified in the parasite tegument[79–81], which is an abundant tissue in adult worms when compared with other life-cycle stages.

Smp_101310.1 and Smp_196510.1 have average Cq values of 22 and 19, respectively, across the six developmental life-cycle stages tested. It is recommended that reference genes should have expression levels in the same range as the genes of interest being tested. Therefore, if genes of interest that are being quantified across life-cycle stages are lowly expressed, such as long non-coding RNA genes[82,83], we recommend the inclusion of a third reference gene Smp_085780.1 (*39S ribosomal mitochondrial protein L10*), a mitochondrial large ribosomal subunit component gene that was detected with an average Cq of 26 in the RT-qPCR assays.

By using the geometric mean of Smp_101310.1 and Smp_196510.1, the two reference genes recommended here, for the expression normalization of the genes commonly used as reference in previous works, we found that Smp_090920.1 (*28s ribosomal protein subunit 14*), Smp_062630.1 (*Dynactin subunit* 2), Smp_001500.1 (*eIF4e*), Smp_056970.1 (*GAPDH*), Smp_003990.1 (*Triosephosphate isomerase*), Smp_090120.1 (*α-tubulin*), Smp_900000.1 (*Cytochrome C oxidase Subunit 1*) and Smp_307020.1 (*Actin*) were differentially expressed in at least one *S. mansoni* developmental stage, and therefore should not be used for RT-qPCR normalization purposes when analyzing different parasite stages. It is also worth mentioning that whereas our analysis comprised six different developmental stages, this might not be the case for all the previous works in the literature mentioned here, and therefore the reference genes used in some of them may have been adequate for those specific works.

Expression stability analysis indicated that Smp_101310.1 and Smp_196510.1 were also suitable for RT-qPCR data normalization in experiments involving the pairing status of females in culture, or the exposure of parasites to dsRNAs in gene-silencing assays. Therefore, while it is recommended to define stable reference genes in the condition that is being specifically tested, the genes pointed here can be used as starting points when selecting reference genes for correct normalization under such similar experimental conditions.

In conclusion, this work is the first systematic study that screened and validated the optimal reference genes for RT-qPCR relative gene quantification in six *S. mansoni* life-cycle stages. Twenty-five novel candidate reference genes and eight previously used candidate reference genes were selected, and their expression stabilities evaluated. We recommend as the reference genes of choice two genes, Smp_101310.1 and Smp_196510.1, which were the most stably expressed at the six different *S. mansoni* developmental stages that were analyzed.

Methods

RNA-Seq analyses. The selected RNA-Seq libraries were downloaded from the SRA-NCBI database using the SRA Toolkit[84] v.2.10.8. Adapters and bad quality reads were filtered out using fastp v. 0.20.0 with default parameters[85]. Filtered reads were mapped against the *S. mansoni* genome v.7 using STAR[86] v. 2.7.0f, and transcripts were quantified with RSEM[33] v1.3.0 (estimate-rspd parameter on). The GTF file (available at http://schistosoma.usp.br/) containing the protein-coding transcriptome v 7.1 (WBPS14)[19] merged with the lncRNAs transcriptome[20] identified by Maciel et al. was used as a reference during the quantification process. As suggested by DESeq2[30], after reads counting with RSEM[33] we performed a minimal pre-filtering to keep only genes that had at least 10 reads total when adding all stages. This resulted in 13,624 protein-coding genes (out of 14,548 genes in the v 7.1 transcriptome[19]) and in 9391 lncRNAs (out of 16,583 lncRNAs[20]) that were considered for further analyses.

Counts and TPM values obtained with RSEM were imported to R[87] environment v. 3.6.3 with the tximport package[88] v. 1.14.2. Normalized counts were obtained with DESeq2[30] v. 1.24.0, edgeR[31] v. 3.28.1 (calcNormFactors

with TMM method and exported in CPM) and EBSeq[32] v. 1.2.6 (QuantileNorm with quantile = 0.75) packages. The Coefficient of Variance (CV) for each transcript in each normalization method was calculated as the ratio of the standard deviation (SD) σ to the arithmetical mean μ.

A user-friendly web application, built with shiny R[89], showing the expression along the stages for each of the normalization methods and their respective CV values is available at https://verjolab.shinyapps.io/Reference-genes/.

Ethics statement. The experimental protocols were in accordance with the Ethical Principles in Animal Research adopted by the Conselho Nacional de Controle da Experimentação Animal (CONCEA) and the protocol/experiments have been approved by the Comissão de Ética no Uso de Animais do Instituto Butantan (CEUAIB number 8859090919). This study was carried out in compliance with the ARRIVE guidelines (http://www.nc3rs.org.uk/page.asp?id=1357).

Parasite material. The BH strain (Belo Horizonte, Brazil) of *S. mansoni* was maintained in the intermediate snail host *Biomphalaria glabrata* and as definitive host the golden hamster (*Mesocricetus auratus*). Female hamsters aged 4 weeks, freshly weaned, weighing 50–60 g, were housed in cages (30 × 20 × 13 cm) containing a sterile bed of wood shavings. A standard diet (Nuvilab CR-1 Irradiada, Quimtia S/A, Paraná, Brazil) and water were made available ad libitum. The room temperature was kept at 22 ± 2 °C, and a 12:12 h light–dark cycle was maintained.

Hamsters were infected with an *S. mansoni* cercariae suspension containing approximately 200–250 cercariae via subcutaneous injection[90]. After 49 days of infection, *S. mansoni* adult worms were recovered by perfusion of the hepatic portal system[91]. *S. mansoni* eggs were extracted from *S. mansoni* infected female hamsters livers, and miracidia were hatched from *S. mansoni* eggs, both as previously described[92]. Cercariae were harvested from infected *B. glabrata* snails exposed to light and mechanically transformed to obtain schistosomula in vitro[93]. The newly transformed schistosomula were maintained for 48 h in M169 medium (Vitrocell, cat number 00464) supplemented with penicillin/streptomycin, amphotericin, gentamicin (Vitrocell, cat number 00148), 2% fetal bovine serum, 1 μM serotonin, 0.5 μM hypoxanthine, 1 μM hydrocortisone, and 0.2 μM triiodothyronine at 37 °C and 5% CO2 before collection and RNA extraction.

Adult worm pairing and EED silencing assays. Adult worm pairing and silencing experiments were performed in ABC media as previously reported[94]. At day 42 after infection, *S. mansoni* adult worms were recovered by perfusion of the hamster hepatic portal system. For the pairing experiments, ten adult worm paired couples, or unpaired females naturally recovered from the hamster perfusion were cultivated in 6-well plates containing 5 mL of ABC media for 2, 4, or 8 days and 70% of the media was exchanged every other day. For the silencing experiments, adult worm paired couples were treated with 30 μg/mL/day of dsRNA targeting either Smp_165220 (EED) or a control unrelated gene (mCherry) in 6-well plates containing 5 mL of ABC media for 2, 4, 8 or 16 days.

At the end of the experiment, the adult worm couples were collected and stored in *RNAlater* (Thermo Fisher) for further manual separation of males and females before RNA extraction.

dsRNA synthesis. Double-stranded RNA (dsRNA) was synthesized from DNA templates amplified from cDNA of male and female adult worms, using the specific primer sequences indicated below, all of them containing in addition a 17-nt T7 RNA Polymerase Promoter sequence at the 5′-end. For Smp_165220.1 (EED) silencing, three different regions of the transcript were targeted, with three different dsRNAs that were obtained from the in vitro transcription of the three amplicons generated: Region 1—Primer F 5′–ACAGATAGTTCTGTG CAGACTCAA–3′ and Primer R 5′–AGCGGATCAGTTGGTTGACTT–3′; Region 2—Primer F 5′–CCTGTG CTTGTTCAACACTTCC–3′ and Primer R 5′–GGACCAACTCCACTAACTGTAGG–3′ Region 3—Primer F 5′–TGTAGTCTGAAGAATGATCTGGAAGA–3′ and Primer R 5′–CGATCCGTGACCAACAAGACTA–3′. The in vitro dsRNA transcription reaction was adapted from a tRNA transcription protocol[95]. Briefly, reactions were performed at 37 °C for 12 h in a buffer containing 40 mM Tris–HCl (pH 8.0), 22 mM $MgCl_2$, 5 mM DTT, 2 mM spermidine, 0.05% BSA, 15 mM guanosine monophosphate, 7.5 mM of each nucleoside triphosphate, amplified template DNA (0.1 μg/μL) and 5 μM of T7 RNA polymerase. The transcribed dsRNA was treated with DNase at 37 °C for 30 min and precipitated using 1:10 (v/v) 3 M sodium acetate pH 5.2 and 1:1 (v/v) of isopropanol. The pellet was washed twice with 70% ethanol and then eluted in water to reach a final concentration of 3 μg/μL. Double-stranded RNA (30 μg/mL/day) was provided to the parasites via soaking. The mCherry gene was used as a non-related dsRNA control and was amplified from a pPLOT-mCherry plasmid containing the mCherry gene with the following primers, containing in addition a T7 RNA Polymerase Promoter sequence at the 5′-end: Primer F 5′–TGGAAGGTTCTGTAAATGGACA–3′ and Primer R 5′–CTCCCTCAGCCCTTTCGTAT–3′.

RNA extraction and cDNA synthesis. Total RNA from the eggs, miracidia, cercariae, schistosomula, adult males and adult females was extracted using the Qiagen RNeasy Plus Micro Kit (Cat number 74034). Briefly, 55,000–70,000 eggs, 30,000–45,000 miracidia, 40,000–55,000 cercariae, 40,000–65,000 schistosomula, 40 adult paired males and adult paired females for each of four biological replicates were grounded with glass beads in buffer RLT supplemented with 2-mercaptoethanol, according to Qiagen recommendation, for 2 min and then frozen in liquid nitrogen. After freezing, the samples from eggs, cercariae, and schistosomula were thawed in a heat block at 65 °C and submitted to a grinding, freezing, and de-freezing process twice more. Miracidia, male and female adult worms were ground only once and frozen and thawed three times. The protocol was followed according to the manufacturer's instructions, including gDNA exclusion by the provided gDNA elimination col-

umn. All RNA samples were quantified using the Qubit RNA HS Assay Kit (Q32852, Thermo Fisher Scientific), and the integrity of RNAs was verified using the Agilent RNA 6000 Pico Kit (5067–1513, Agilent Technologies) in a 2100 Bioanalyzer Instrument (Agilent Technologies). For quantitative RT-qPCR analysis across the different life cycle stages of *S. mansoni*, complementary DNAs were obtained by reverse transcription (RT) of 400 ng total RNA using SuperScript IV Reverse Transcriptase (18091050, Invitrogen) and random hexamer primers in a 20 µL volume, according to the manufacturer's recommendations. For the pairing status experiments, cDNA was synthetized from 1000 ng of total RNA and for the silencing experiments, cDNA was obtained from 200 ng of total RNA.

Quantitative RT-qPCR assays and analyses. Thirty-three candidate reference genes were selected for RT-qPCR assays and analysis in this study. Twenty-five candidate reference genes were selected based on a re-analysis of published data, and eight candidate reference genes were selected because of their frequent use in previous publications (Table 1 and Supplementary Table S6). All primer pairs were designed using Primer-Quest Tool provided by IDT Integrated DNA Technologies (https://www.idtdna.com/PrimerQuest/) with PCR amplicon length ranging from 50 to 300 bp and melting temperature (Tm) of approximately 60 °C. All primer sequences are reported in Table 1 and the predicted amplicon sizes and primers efficiencies are shown in Supplementary Table S7. After reverse transcription, the obtained complementary DNAs (cDNAs) were diluted eight times in water. Quantitative PCR was performed using 2.5 µL of each diluted cDNA in a total volume of 10 µL containing 1× LightCycler 480 SYBR Green I Master Mix (04707516001, Roche Diagnostics), 800 nM of each primer in a LightCycler 480 System (Roche Diagnostics), and each real-time qPCR was run in three technical replicates. The PCR conditions included initial activation at 95 °C for 10 min; 45 cycles with denaturation at 95 °C for 10 s, annealing at 60 °C for 10 s and extension at 72 °C for 20 s. A dissociation step (95 °C for 15 s, 60 °C for 1 min, 95 °C for 15 s, 60 °C for 15 s) was added at the end of the run to confirm the amplicon specificity for each gene. The quantitative RT-qPCR assays were performed following the MIQE guidelines[21,51,53].

The amplification efficiency for each primer was calculated using a diluted series of cDNA synthesized using five µg of RNA from *S. mansoni* male and female adult worms. The cDNAs were mixed and diluted in a 3× factor starting from 1:8 dilution. The average Cq values from all three technical replicates in at least four different dilutions were used to retrieve a linear regression. The x-axis corresponds to the $\log_{10}$ dilution factor. The slope from the linear regression curve was used to get the primer efficiency value (**E**), where $\mathbf{E} = 10^{\frac{1}{Slope}}$. The **E** value is then used in the equation where Primer Efficiency (%) = ((**E** − 1)*100). An **E** value of 2 equals to a Primer Efficiency of 100%; therefore, the slope needs to be around 3.32.

Three different tools were used to evaluate gene expression stability using RT-qPCR data. The first one is geNorm[21], which requires the transformation of quantification cycle (Cq) values to relative quantities, ranging from 0 to 1. The mean Cq values from the triplicate runs were used as input data and converted into relative quantities using the lowest Cq values mean of each primer as reference for normalization. It is important to note that the primer efficiency value **(E)** is used to retrieve the relative quantities by using the equation Relative quantity $= E^{\text{Lowest Cp value}-\text{Cp value}}$. This program estimates an expression stability value (M) for each gene, and the lowest M values correspond to the most stable expressed genes. The determination of the optimal number of reference genes for reliable normalization by geNorm relies on the calculation of the pairwise variation values (V), which measures pairwise variation (Vn/n + 1) between the sequentially ranked normalization factors NFn and NFn + 1, where *n* is the number of genes involved in the normalization factor. The recommended cut-off value[21] below which there is no need for inclusion of another gene is 0.15.

The second tool used was NormFinder[51] that requires the Cq values to be transformed to a linear scale expression quantity (without any negative values) by using the primer efficiency value **(E)**, where Linear quantity=$E^{-\text{Cq value}}$. This software determines a stability value ranging from 0 to 1, where the lower values correspond to the best reference genes. Furthermore, this software provides additional information about the Intragroup and Intergroup variation. While intragroup variation describes how much the biological replicates in each life-cycle stage differ from each other, intergroup variation describes how much the averages of the replicates from a given life-cycle stage differ from the averages of other stages.

The third software used was RefFinder[52] (http://blooge.cn/RefFinder/) that integrates four different software, namely BestKeeper and NormFinder, geNorm, and the Comparative delta-Ct method. RefFinder only requires inputting the Cq values from each sample and primers, and then it retrieves each software result and a comprehensive ranking gathering all results. RefFinder does not use the primer efficiency value in its data input; thus, each primer's corrected Cq values were submitted instead. The corrected Cq value is obtained with the following equation: Corrected Cq value = $\text{Log}_2 E^{\text{Cq value}}$.

The ΔCt method[96] was used to determine if the reference genes most commonly used in the literature were differentially expressed among the six *S. mansoni* life-cycle stages, to calculate p14 differential expression in the pairing experiments, and to calculate EDD knockdown in the silencing experiments. For this, the geometric mean of the Cq values from the two best reference genes found in the present work (Smp_101310.1 and Smp_196510.1), measured across all samples, was used to normalize the expression of the genes of interest. Primers were retrieved from previous publications for Smp_316140 (p14)[97] and Smp_165220 (EED)[28].

Statistical analyses. One-way ANOVA with Tukey correction for multiple comparisons was used to compare the differential expression across the *S. mansoni* stages of the eight reference genes frequently used in the literature. Unpaired Student t-test was used to compare the differential expression of Smp_316140 (p14) or of Smp_165220 (EED) across the pairing and silencing samples. GraphPad Prism software (version 8.0) was used to perform the analyses. p-value thresholds were *< 0.05, **< 0.01, ***< 0.001, and ****< 0.0001.

Received: 6 April 2021; Accepted: 31 July 2021
Published online: 19 August 2021

References

1. Colley, D. G., Bustinduy, A. L., Secor, W. E. & King, C. H. Human schistosomiasis. *Lancet* **383**, 2253–2264 (2014).
2. Zoni, A. C., Catalá, L. & Ault, S. K. Schistosomiasis prevalence and intensity of infection in Latin America and the Caribbean Countries, 1942–2014: A systematic review in the context of a regional elimination goal. *PLoS Negl. Trop. Dis.* **10**, e0004493 (2016).
3. Lewis, F. A. & Tucker, M. S. Schistosomiasis. In *Digenetic Trematodes. Advances in Experimental Medicine and Biology*, vol 766 (eds Toledo, R. & Fried, B.) (Springer, New York, 2014). https://doi.org/10.1007/978-1-4939-0915-5_3.
4. Verjovski-Almeida, S. *et al.* Transcriptome analysis of the acoelomate human parasite *Schistosoma mansoni*. *Nat. Genet.* **35**, 148–157 (2003).
5. Berriman, M. *et al.* The genome of the blood fluke *Schistosoma mansoni*. *Nature* **460**, 352–358 (2009).
6. Protasio, A. V. *et al.* A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. *PLoS Negl. Trop. Dis.* **6**, e1455 (2012).
7. Almeida, G. T. *et al.* Exploring the *Schistosoma mansoni* adult male transcriptome using RNA-seq. *Exp. Parasitol.* **132**, 22–31 (2012).
8. Anderson, L. *et al. Schistosoma mansoni* egg, adult male and female comparative gene expression analysis and identification of novel genes by RNA-Seq. *PLoS Negl. Trop. Dis.* **9**, e0004334 (2015).
9. Leutner, S. *et al.* Combinatory microarray and SuperSAGE analyses identify pairing-dependently transcribed genes in *Schistosoma mansoni* males, including follistatin. *PLoS Negl. Trop. Dis.* **7**, e2532 (2013).
10. Ojopi, E. P. B. *et al.* A quantitative view of the transcriptome of *Schistosoma mansoni* adult-worms using SAGE. *BMC Genom.* **8**, 186–186 (2007).
11. Parker-Manuel, S. J., Ivens, A. C., Dillon, G. P. & Wilson, R. A. Gene expression patterns in larval *Schistosoma mansoni* associated with infection of the mammalian host. *PLoS Negl. Trop. Dis.* **5**, e1274 (2011).
12. Picard, M. A. L. *et al.* Sex-biased transcriptome of *Schistosoma mansoni*: Host–parasite interaction, genetic determinants and epigenetic regulators are associated with sexual differentiation. *PLoS Negl. Trop. Dis.* **10**, e0004930 (2016).
13. Taft, A. S. *et al.* Transcriptome analysis of *Schistosoma mansoni* larval development using serial analysis of gene expression (SAGE). *Parasitology* **136**, 469–485 (2009).
14. Nawaratna, S. S. K. *et al.* Transcriptional profiling of the oesophageal gland region of male worms of *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* **196**, 82–89 (2014).
15. Nawaratna, S. S. K., McManus, D. P., Moertel, L., Gobert, G. N. & Jones, M. K. Gene Atlasing of digestive and reproductive tissues in *Schistosoma mansoni*. *PLoS Negl. Trop. Dis.* **5**, e1043 (2011).
16. Lu, Z. *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).
17. Tarashansky, A. J., Xue, Y., Li, P., Quake, S. R. & Wang, B. Self-assembling manifolds in single-cell RNA sequencing data. *Elife* **8**, e48994 (2019).
18. Wang, B. *et al.* Stem cell heterogeneity drives the parasitic life cycle of Schistosoma mansoni. *Elife* **7**, e35449 (2018).
19. Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P. & Berriman, M. WormBase ParaSite—A comprehensive resource for helminth genomics. *Mol. Biochem. Parasitol.* **215**, 2–10 (2017).
20. Maciel, L. F. *et al.* Weighted gene co-expression analyses point to long non-coding RNA hub genes at different *Schistosoma mansoni* life-cycle stages. *Front. Genet.* **10**, 823 (2019).
21. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, RESEARCH0034 (2002).
22. Dos Santos, K. C. G., Desgagné-Penix, I. & Germain, H. Custom selected reference genes outperform pre-defined reference genes in transcriptomic analysis. *BMC Genom.* **21**, 35 (2020).
23. Chapman, J. R. & Waldenström, J. With reference to reference genes: A systematic review of endogenous controls in gene expression studies. *PLoS ONE* **10**, e0141853 (2015).
24. Liu, S. *et al.* Genome-wide identification and characterization of a panel of house-keeping genes in *Schistosoma japonicum*. *Mol. Biochem. Parasitol.* **182**, 75–82 (2012).
25. Haeberlein, S. *et al.* Identification of a new panel of reference genes to study pairing-dependent gene expression in *Schistosoma mansoni*. *Int. J. Parasitol.* **49**, 615–624 (2019).
26. Carneiro, V. C. *et al.* Pharmacological inhibition of lysine-specific demethylase 1 (LSD1) induces global transcriptional deregulation and ultrastructural alterations that impair viability in *Schistosoma mansoni*. *PLoS Negl. Trop. Dis.* **14**, e0008332 (2020).
27. Kim, H. C., Khalil, A. M. & Jolly, E. R. LncRNAs in molluscan and mammalian stages of parasitic schistosomes are developmentally-regulated and coordinately expressed with protein-coding genes. *RNA Biol.* **17**, 805–815 (2020).
28. Pereira, A. S. A. *et al.* Inhibition of histone methyltransferase EZH2 in *Schistosoma mansoni* in vitro by GSK343 reduces egg laying and decreases the expression of genes implicated in DNA replication and noncoding RNA metabolism. *PLoS Negl. Trop. Dis.* **12**, e0006873 (2018).
29. Wang, B., Collins, J. J. 3rd. & Newmark, P. A. Functional genomic characterization of neoblast-like stem cells in larval *Schistosoma mansoni*. *Elife* **2**, e00768 (2013).
30. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
31. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
32. Leng, N. & Kendziorski, C. EBSeq: An R package for gene and isoform differential expression analysis of RNA-seq data. *R Package Version 1.30.0* https://doi.org/10.18129/B9.bioc.EBSeq (2020).
33. Li, B. & Dewey, C. N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **12**, 323 (2011).
34. Vasconcelos, E. J. R. *et al.* The *Schistosoma mansoni* genome encodes thousands of long non-coding RNAs predicted to be functional at different parasite life-cycle stages. *Sci. Rep.* **7**, 10508 (2017).
35. Pereira, R. V. *et al.* Ubiquitin-specific proteases are differentially expressed throughout the *Schistosoma mansoni* life cycle. *Parasit. Vectors* **8**, 349 (2015).
36. Pereira, R. V., de Gomes, M. S., Jannotti-Passos, L. K., Borges, W. C. & Guerra-Sá, R. Characterisation of the COP9 signalosome in Schistosoma mansoni parasites. *Parasitol. Res.* **112**, 2245–2253 (2013).
37. El-Shehabi, F., Vermeire, J. J., Yoshino, T. P. & Ribeiro, P. Developmental expression analysis and immunolocalization of a biogenic amine receptor in *Schistosoma mansoni*. *Exp. Parasitol.* **122**, 17–27 (2009).
38. Farias, L. P. *et al.* Screening the *Schistosoma mansoni* transcriptome for genes differentially expressed in the schistosomulum stage in search for vaccine candidates. *Parasitol. Res.* **108**, 123–135 (2010).
39. Oke, T. T., Moskovitz, J. & Williams, D. L. Characterization of the methionine sulfoxide reductases of *Schistosoma mansoni*. *J. Parasitol.* **95**, 1421–1428 (2009).

40. Figueiredo, B. C., Da'dara, A. A., Oliveira, S. C. & Skelly, P. J. Schistosomes enhance plasminogen activation: The role of tegumental enolase. *PLoS Pathog* **11**, e1005335 (2015).
41. Krautz-Peterson, G. *et al.* Amino acid transport in schistosomes. *J. Biol. Chem.* **282**, 21767–21775 (2007).
42. Araujo-Montoya, B. O. *et al. Schistosoma mansoni*: Molecular characterization of Alkaline Phosphatase and expression patterns across life cycle stages. *Exp. Parasitol.* **129**, 284–291 (2011).
43. Chalmers, I. W. *et al.* Developmentally regulated expression, alternative splicing and distinct sub-groupings in members of the *Schistosoma mansoni* venom allergen-like (SmVAL) gene family. *BMC Genom.* **9**, 89 (2008).
44. Skinner, D. E. *et al.* Vasa-Like DEAD-Box RNA helicases of *Schistosoma mansoni*. *PLoS Negl. Trop. Dis.* **6**, e1686 (2012).
45. Soares, C. S. *et al.* Molecular and functional characterization of a putative PA28γ proteasome activator orthologue in *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* **189**, 14–25 (2013).
46. Wu, W., Niles, E. G., Hirai, H. & LoVerde, P. T. Identification and characterization of a nuclear receptor subfamily I member in the Platyhelminth *Schistosoma mansoni* (SmNR1). *FEBS J.* **274**, 390–405 (2006).
47. Dvorák, J. *et al.* SmCL3, a gastrodermal cysteine protease of the human blood fluke *Schistosoma mansoni*. *PLoS Negl. Trop. Dis.* **3**, e449 (2009).
48. Fajtová, P. *et al.* Prolyl oligopeptidase from the blood fluke schistosoma mansoni: from functional analysis to anti-schistosomal inhibitors. *PLoS Negl. Trop. Dis.* **9**, e0003827 (2015).
49. Horn, M. *et al.* Trypsin- and chymotrypsin-like serine proteases in schistosoma mansoni—'The undiscovered country'". *PLoS Negl. Trop. Dis.* **8**, e2766 (2014).
50. Rofatto, H. K. *et al.* Characterization of phosphodiesterase-5 as a surface protein in the tegument of *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* **166**, 32–41 (2009).
51. Andersen, C. L., Jensen, J. L. & Ørntoft, T. F. Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Can. Res.* **64**, 5245–5250 (2004).
52. Xie, F., Xiao, P., Chen, D., Xu, L. & Zhang, B. miRDeepFinder: A miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Mol. Biol.* **80**, 75–84 (2012).
53. Bustin, S. A. *et al.* The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **55**, 611–622 (2009).
54. Galanti, S. E., Huang, S. C. & Pearce, E. J. Cell death and reproductive regression in female Schistosoma mansoni. *PLoS Negl. Trop. Dis.* **6**, e1509 (2012).
55. Grevelding, C. G., Sommer, G. & Kunz, W. Female-specific gene expression in *Schistosoma mansoni* is regulated by pairing. *Parasitology* **115**(Pt 6), 635–640 (1997).
56. Oshlack, A. & Wakefield, M. J. Transcript length bias in RNA-seq data confounds systems biology. *Biol. Direct* **4**, 14 (2009).
57. Risso, D., Schwartz, K., Sherlock, G. & Dudoit, S. GC-content normalization for RNA-Seq data. *BMC Bioinform.* **12**, 480 (2011).
58. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
59. Abbas-Aghababazadeh, F., Li, Q. & Fridley, B. L. Comparison of normalization approaches for gene expression studies completed with high-throughput sequencing. *PLoS ONE* **13**, e0206312 (2018).
60. Abrams, Z. B., Johnson, T. S., Huang, K., Payne, P. R. O. & Coombes, K. A protocol to evaluate RNA sequencing normalization methods. *BMC Bioinform.* **20**, 679 (2019).
61. Evans, C., Hardin, J. & Stoebel, D. M. Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Brief Bioinform.* **19**, 776–792 (2018).
62. Lafzi, A., Moutinho, C., Picelli, S. & Heyn, H. Tutorial: Guidelines for the experimental design of single-cell RNA sequencing studies. *Nat. Protoc.* **13**, 2742–2757 (2018).
63. Li, P., Piao, Y., Shon, H. S. & Ryu, K. H. Comparing the normalization methods for the differential analysis of Illumina high-throughput RNA-Seq data. *BMC Bioinform.* **16**, 347 (2015).
64. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: A tutorial. *Mol. Syst. Biol.* **15**, e8746 (2019).
65. Bullard, J. H., Purdom, E., Hansen, K. D. & Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinform.* **11**, 94 (2010).
66. Zhao, S., Ye, Z. & Stanton, R. Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. *RNA* **26**, 903–909 (2020).
67. Dheda, K. *et al.* Validation of housekeeping genes for normalizing RNA expression in real-time PCR. *Biotechniques* **37**, 112–119 (2004).
68. Jacob, F. *et al.* Careful selection of reference genes is required for reliable performance of RT-qPCR in human normal and cancer cell lines. *PLoS ONE* **8**, e59180 (2013).
69. Shakeel, M., Rodriguez, A., Tahir, U. B. & Jin, F. Gene expression studies of reference genes for quantitative real-time PCR: An overview in insects. *Biotech. Lett.* **40**, 227–236 (2017).
70. Svingen, T., Spiller, C. M., Kashimada, K., Harley, V. R. & Koopman, P. Identification of suitable normalizing genes for quantitative real-time RT-PCR analysis of gene expression in fetal mouse gonads. *Sexual Development* **3**, 194–204 (2009).
71. Mitra, P. *et al.* HiNF-P is a bifunctional regulator of cell cycle controlled histone H4 gene transcription. *J. Cell. Biochem.* **101**, 181–191 (2007).
72. Medina, R. *et al.* The histone gene cell cycle regulator HiNF-P is a unique zinc finger transcription factor with a novel conserved auxiliary DNA-binding motif. *Biochemistry* **47**, 11415–11423 (2008).
73. Chen, M., Gutierrez, G. J. & Ronai, Z. A. Ubiquitin-recognition protein Ufd1 couples the endoplasmic reticulum (ER) stress response to cell cycle control. *Proc. Natl. Acad. Sci. USA* **108**, 9119–9124 (2011).
74. Guerra-Sa, R., Castro-Borges, W., Evangelista, E. A., Kettelhut, I. C. & Rodrigues, V. *Schistosoma mansoni*: Functional proteasomes are required for development in the vertebrate host. *Exp. Parasitol.* **109**, 228–236 (2005).
75. Nabhan, J. F., El-Shehabi, F., Patocka, N. & Ribeiro, P. The 26S proteasome in *Schistosoma mansoni*: Bioinformatics analysis, developmental expression, and RNA interference (RNAi) studies. *Exp. Parasitol.* **117**, 337–347 (2007).
76. Ercolani, L., Florence, B., Denaro, M. & Alexander, M. Isolation and complete sequence of a functional human glyceraldehyde-3-phosphate dehydrogenase gene. *J. Biol. Chem.* **263**, 15335–15341 (1988).
77. Zhou, Y. & Podesta, R. B. Surface spines of human blood flukes (*Schistosoma mansoni*) contain bundles of actin filaments having identical polarity. *Eur. J. Cell Biol.* **48**, 150–153 (1989).
78. Zinsser, V. L., Farnell, E., Dunne, D. W. & Timson, D. J. Triose phosphate isomerase from the blood flukeSchistosoma mansoni: Biochemical characterisation of a potential drug and vaccine target. *FEBS Lett.* **587**, 3422–3427 (2013).
79. van Balkom, B. W. *et al.* Mass spectrometric analysis of the *Schistosoma mansoni* tegumental sub-proteome. *J. Proteome Res.* **4**, 958–966 (2005).
80. Pirovich, D. B., Da'dara, A. A. & Skelly, P. J. *Schistosoma mansoni* glyceraldehyde-3-phosphate dehydrogenase enhances formation of the blood-clot lysis protein plasmin. *Biol. Open* **9**, bio050385 (2020).
81. Sotillo, J., Pearson, M., Becker, L., Mulvenna, J. & Loukas, A. A quantitative proteomic analysis of the tegumental proteins from *Schistosoma mansoni* schistosomula reveals novel potential therapeutic targets. *Int. J. Parasitol.* **45**, 505–516 (2015).

82. Amaral, M. S. *et al.* Long non-coding RNA levels can be modulated by 5-azacytidine in *Schistosoma mansoni*. *Sci. Rep.* **10**, 21565 (2020).
83. Oliveira, V. F. *et al.* Identification of 170 new long noncoding RNAs in *Schistosoma mansoni*. *Biomed. Res. Int.* **2018**, 1264697 (2018).
84. Thornton, J. Downloading SRA data using the SRA toolkit v1. In *protocols.io*. (ZappyLab, Inc, 2016). https://doi.org/10.17504/protocols.io.fthbnj6.
85. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
86. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
87. R Development Core Team. R: A language and environment for statistical computing. *R Found. Stat. Comput.* https://doi.org/10.1007/978-3-540-74686-7 (2011).
88. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1521 (2015).
89. Chang, W. *et al.* Shiny: Web Application Framework for R. https://CRAN.R-project.org/package=shiny (2021). https://shiny.rstudio.com.
90. De Souza, C. P., Dias, E. P., De Azevedo, M. D. & Paulini, E. Observations upon some factors which influence the laboratory maintenance of *Schistosoma mansoni* (author's transl). *Revista brasileira de pesquisas medicas e biologicas* **12**, 411–419 (1979).
91. Smithers, S. R. & Terry, R. J. The infection of laboratory hosts with cercariae of *Schistosoma mansoni* and the recovery of the adult worms. *Parasitology* **55**, 695–700 (1965).
92. Dalton, J. P., Day, S. R., Drew, A. C. & Brindley, P. J. A method for the isolation of schistosome eggs and miracidia free of contaminating host tissues. *Parasitology* **115**, 29–32 (1997).
93. Basch, P. F. Cultivation of *Schistosoma mansoni* in vitro. I. Establishment of cultures from cercariae and development until pairing. *J. Parasitol.* **67**, 179 (1981).
94. Wang, J., Chen, R. & Collins, J. J. 3rd. Systematically improved in vitro culture conditions reveal new insights into the reproductive biology of the human parasite *Schistosoma mansoni*. *PLoS Biol.* **17**, e3000254 (2019).
95. Sampson, J. R. & Uhlenbeck, O. C. Biochemical and physical characterization of an unmodified yeast phenylalanine transfer RNA transcribed in vitro. *Proc. Natl. Acad. Sci. USA* **85**, 1033–1037 (1988).
96. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402–408 (2001).
97. Buro, C. *et al.* Imatinib treatment causes substantial transcriptional changes in adult *Schistosoma mansoni* in vitro exhibiting pleiotropic effects. *PLoS Negl. Trop. Dis.* **8**, e2923 (2014).

Author contributions

M.S.A. conceived the project. G.O.S. and M.S.A. designed the experiments. G.O.S., M.S.A., H.S.C., G.G.O.O., and A.S.A.P. carried out the wet-lab experiments. G.O.S., M.S.A and L.F.M. carried out the in-silico experiments. G.O.S. analyzed the data and wrote the first draft of the manuscript. M.S.A. and S.V.-A. revised the manuscript. P.A.M. and E.N. provided biological material. S.V.-A. coordinated the study and obtained funding. All authors reviewed and approved the manuscript.

Funding

This work was supported by a Grant from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; thematic Grant number 2018/23693-5 to SV-A). G.O.S., L.F.M., and A.S.A.P. received fellowships from FAPESP (18/24015-0, 18/19591-2 and 2016/10046-6, respectively). H.S.C. and G.G.O.O. were supported by fellowships from Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (CNPq) (164533/2019-2 and 116733/2019-5, respectively). S.V.-A. received an established investigator fellowship award from CNPq, Brasil. S.V.-A. laboratory was also supported by institutional funds from Fundação Butantan.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-96055-7.

Correspondence and requests for materials should be addressed to S.V.-A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021

Supplementary Figures for

Assessment of reference genes at six different developmental stages of *Schistosoma mansoni* for quantitative RT-PCR

Gilbert O. Silveira[1,2, #], Murilo S. Amaral[1, #], Helena S. Coelho[1], Lucas F. Maciel[1], Adriana S. A. Pereira[1,2], Giovanna G. O. Olberg[1], Patricia A. Miyasato[1], Eliana Nakano[1] & Sergio Verjovski-Almeida[1,2*]

[1] Laboratório de Parasitologia, Instituto Butantan, 05503-900, Sao Paulo, SP, Brazil.

[2] Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, 05508-900, Sao Paulo, SP, Brazil.

#Authors contributed equally

[*]Correspondence should be addressed to S.V.A. (email: verjo@iq.usp.br)

Supplementary Figures

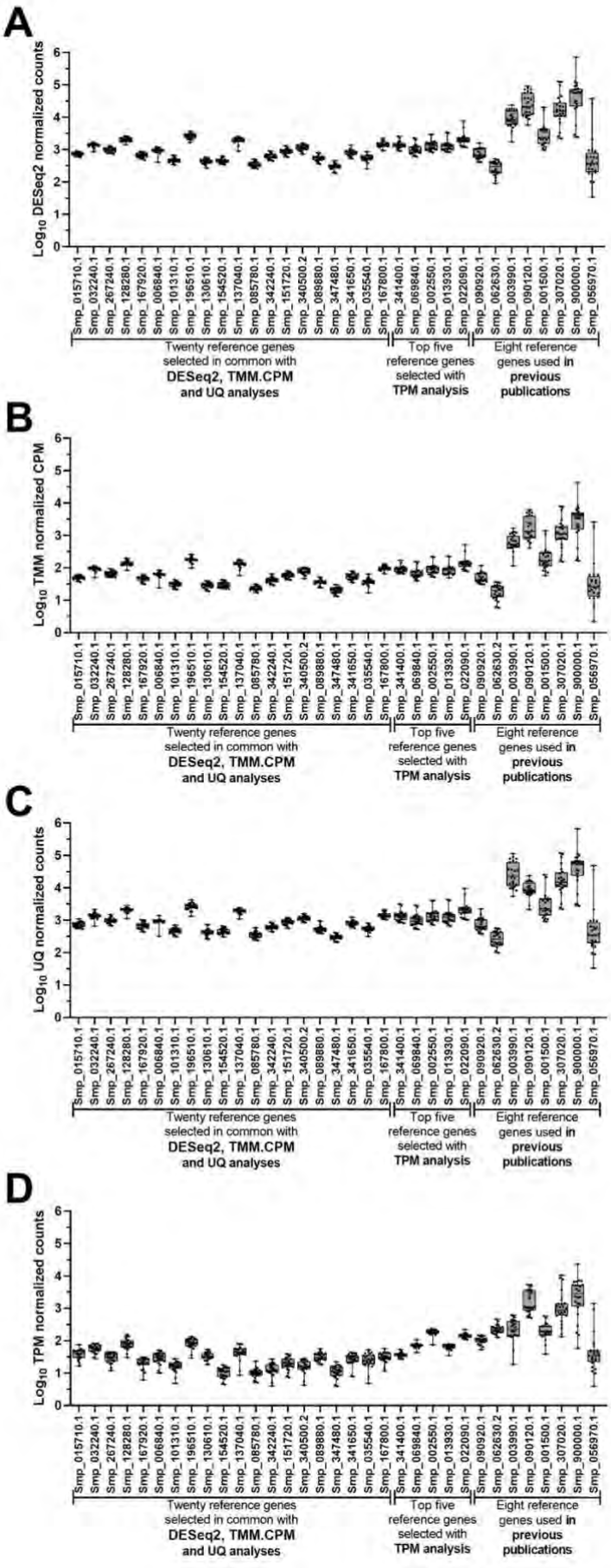

A
Log10 DESeq2 normalized counts
B
Log10 TMM normalized CPM
C
Log10 UQ normalized counts
D
Log10 TPM normalized counts
Smp_015710.1
Smp_032240.1
Smp_267240.1
Smp_128280.1
Smp_167920.1
Smp_006840.1
Smp_101310.1
Smp_196510.1
Smp_130610.1
Smp_154520.1
Smp_137040.1
Smp_085780.1
Smp_342240.1
Smp_151720.1
Smp_340500.2
Smp_089880.1
Smp_347480.1
Smp_341650.1
Smp_035540.1
Smp_167800.1
Smp_341400.1
Smp_069840.1
Smp_002550.1
Smp_013930.1
Smp_022090.1
Smp_090920.1
Smp_062630.1
Smp_062630.2
Smp_003990.1
Smp_090120.1
Smp_001500.1
Smp_307020.1
Smp_900000.1
Smp_056970.1
Twenty reference genes selected in common with DESeq2, TMM.CPM and UQ analyses
Top five reference genes selected with TPM analysis
Eight reference genes used in previous publications

Figure S1. Expression patterns of candidate reference genes at different *Schistosoma mansoni* life-cycle stages. Twenty-five candidate reference genes that have the lowest coefficients of variation of their expression values across 24 RNA-Seq libraries were selected in this work for evaluation of their stability by RT-qPCR assays using six different *S. mansoni* life-cycle stages (see Methods). The libraries used in the analysis are described in Supplementary Table S1. Eight candidate reference genes commonly used in previous publications were also selected. Log_{10} transformed expression profiles from all selected reference genes are represented by their minimum to maximum values for each of the four RNA-Seq normalization methods tested: **(A)** DESeq2; **(B)** TMM.CPM; **(C)** UQ; and **(D)** TPM. The horizontal line represents the median, and the boxes and whiskers represent the inter-quantile and min to max ranges, respectively.

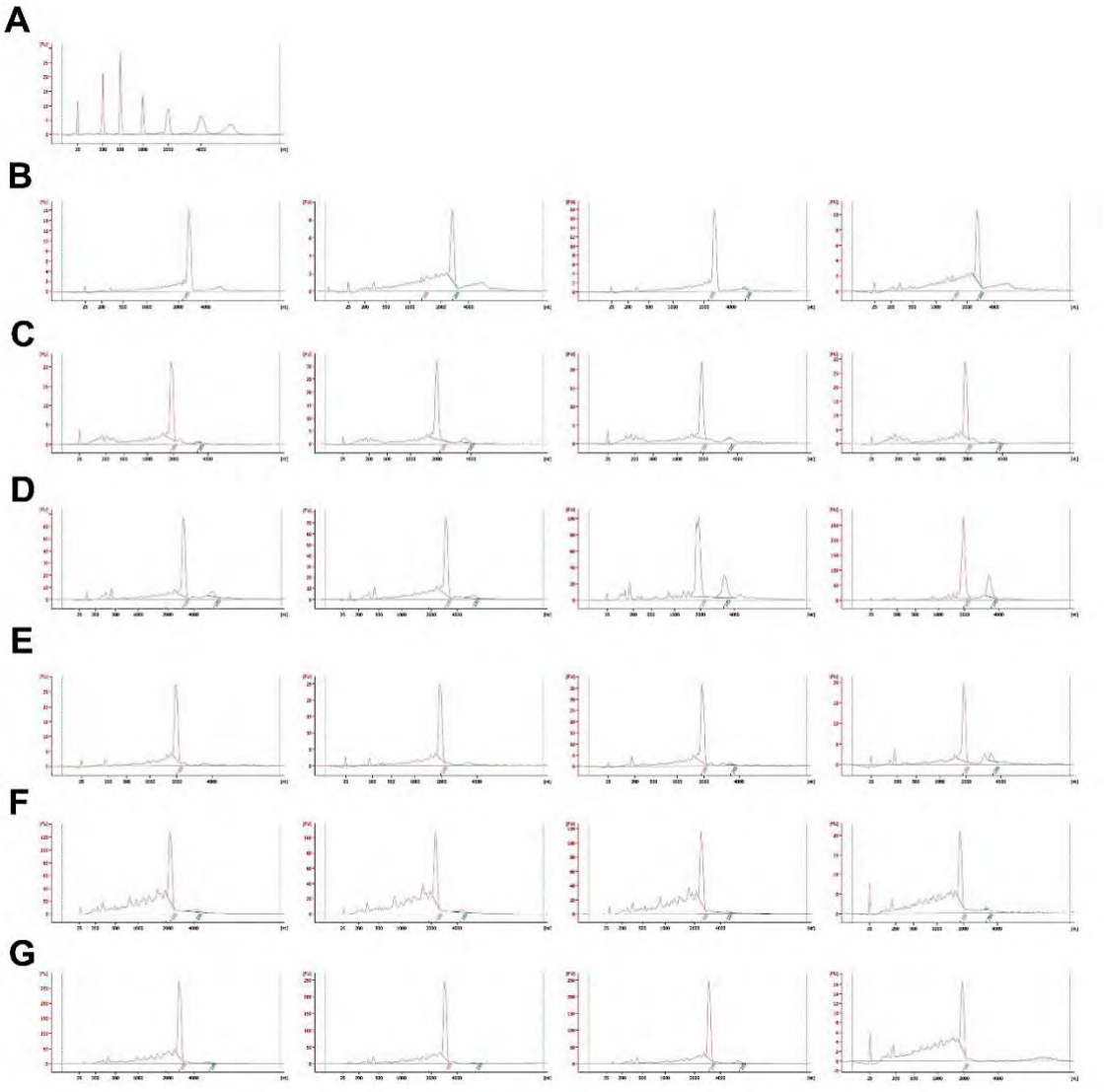


Figure S2. RNA integrity measurement with the Agilent RNA 6000 Pico Kit in a 2100 Bioanalyzer Instrument for all the life cycle stage samples used for RT-qPCR assays. **(A)** Ladder run, **(B)** Eggs samples, **(C)** Miracidia samples, **(D)** Cercariae samples, **(E)** 48-h-schistosomula samples, **(F)** Adult males samples and **(G)** Adult females samples. The y-axis represents the fluorescence intensity for each of the graphs, while the x-axis represents the run time. All four biological replicates of each life cycle stage sample are represented side by side. A typical single peak (denoting 18S ribosomal subunit RNA) is expected.

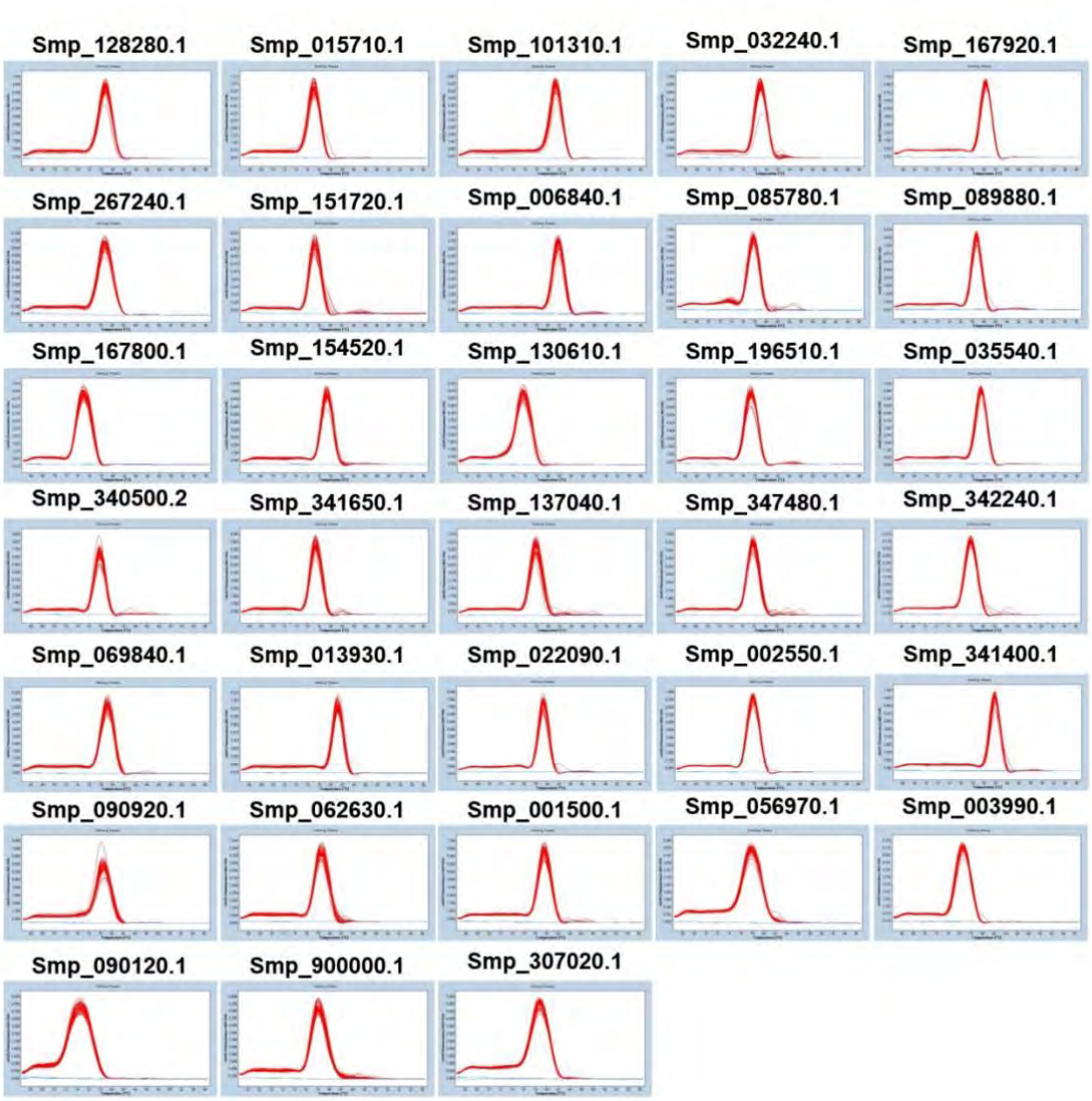


Figure S3. Melting curves for all amplicons obtained in the RT-qPCR assays. For each of the graphs, the y-axis represents the fluorescence intensity while the x-axis represents temperature. Smp_nnnnnn is the code for each of the genes selected for RT-qPCR normalization in different life-cycle stages of *S. mansoni.* The sequences of the primer pairs used for each gene are shown in Supplementary Table S7.

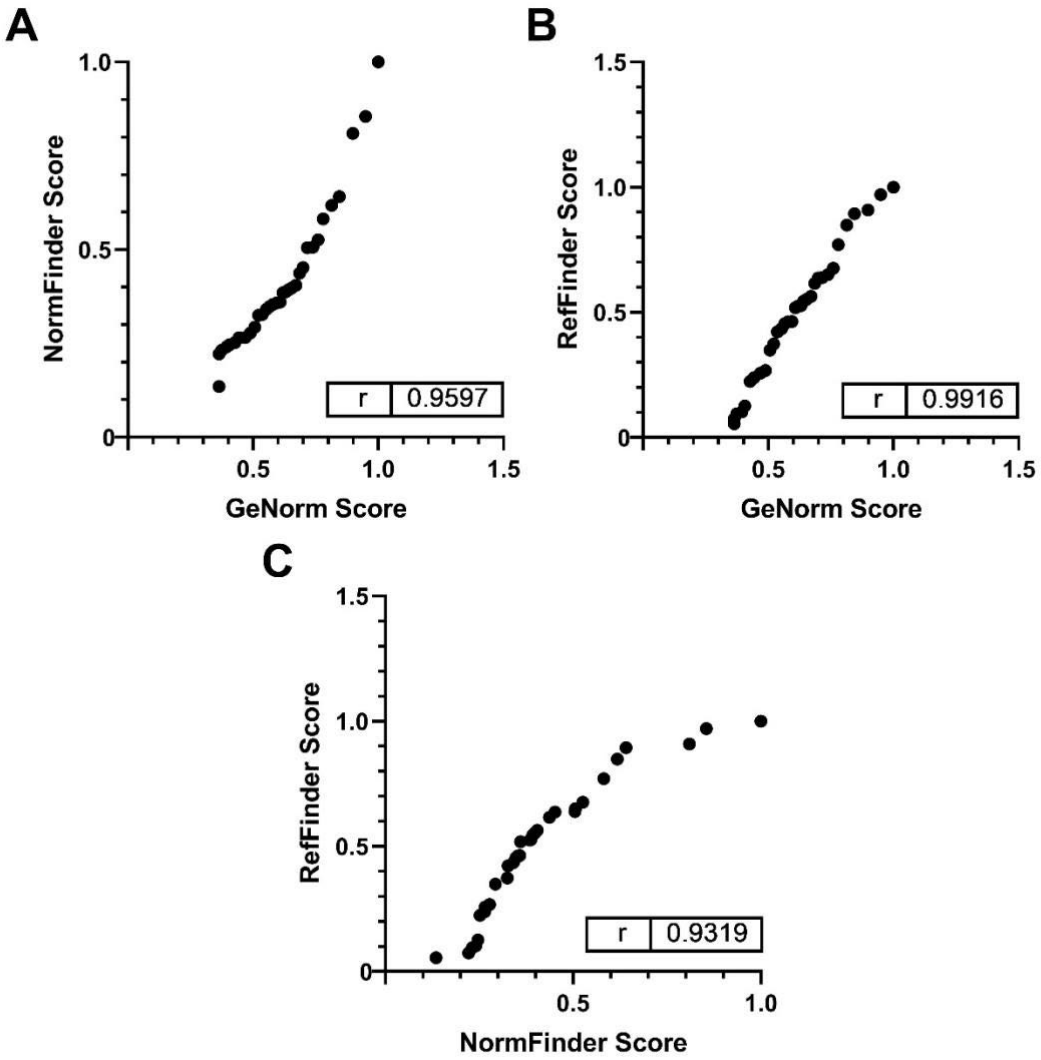


Figure S4. Pearson correlation calculated between the results of GeNorm, NormFinder, and RefFinder. The stability scores for each of the candidate reference genes obtained from the three different software (GeNorm, NormFinder, and RefFinder) were used to calculate the Pearson correlation coefficient r. In **(A)** correlation between the results from GeNorm and NormFinder, in **(B)** correlation between GeNorm and RefFinder, and in **(C)** correlation between RefFinder and NormFinder results. Each dot represents one of the 33 candidate reference genes tested. For **(A)**, **(B)** and **(C),** p-value <0.001.

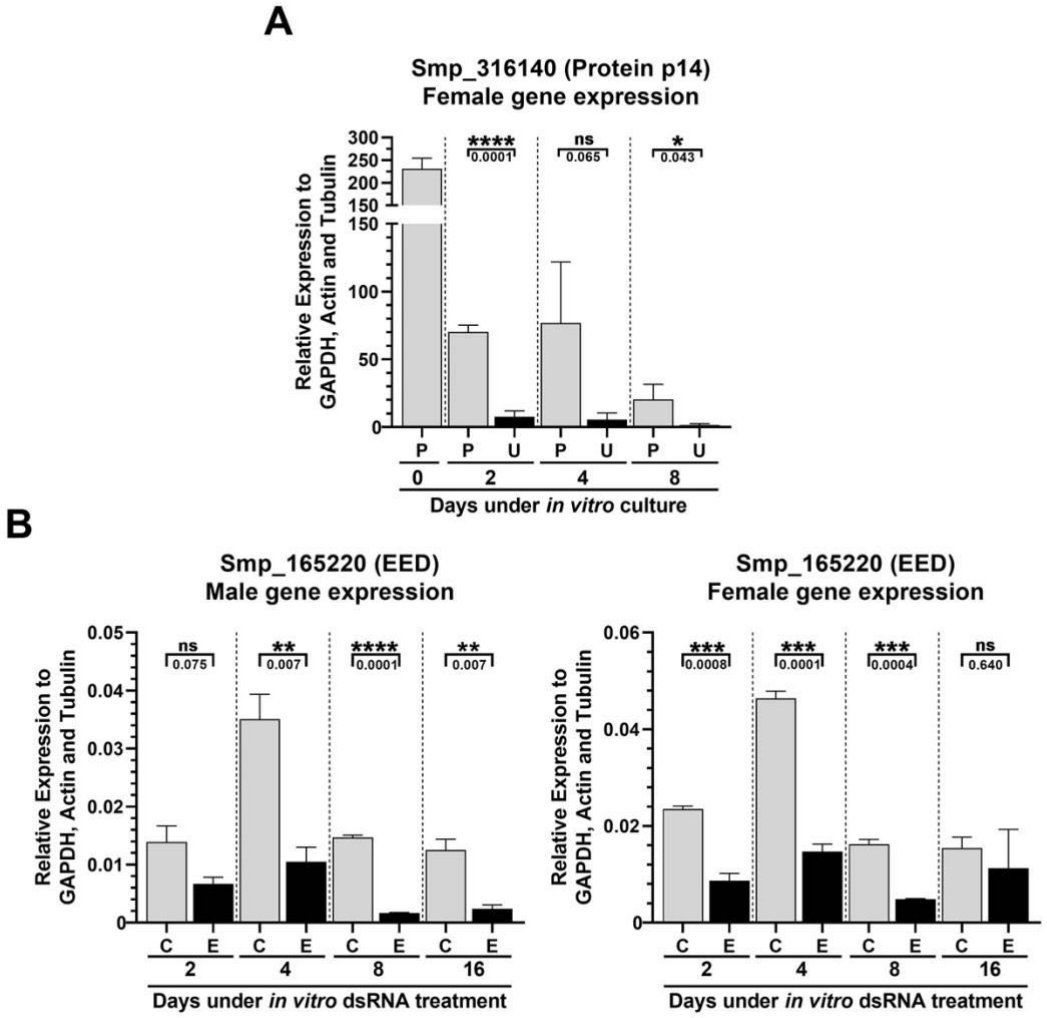


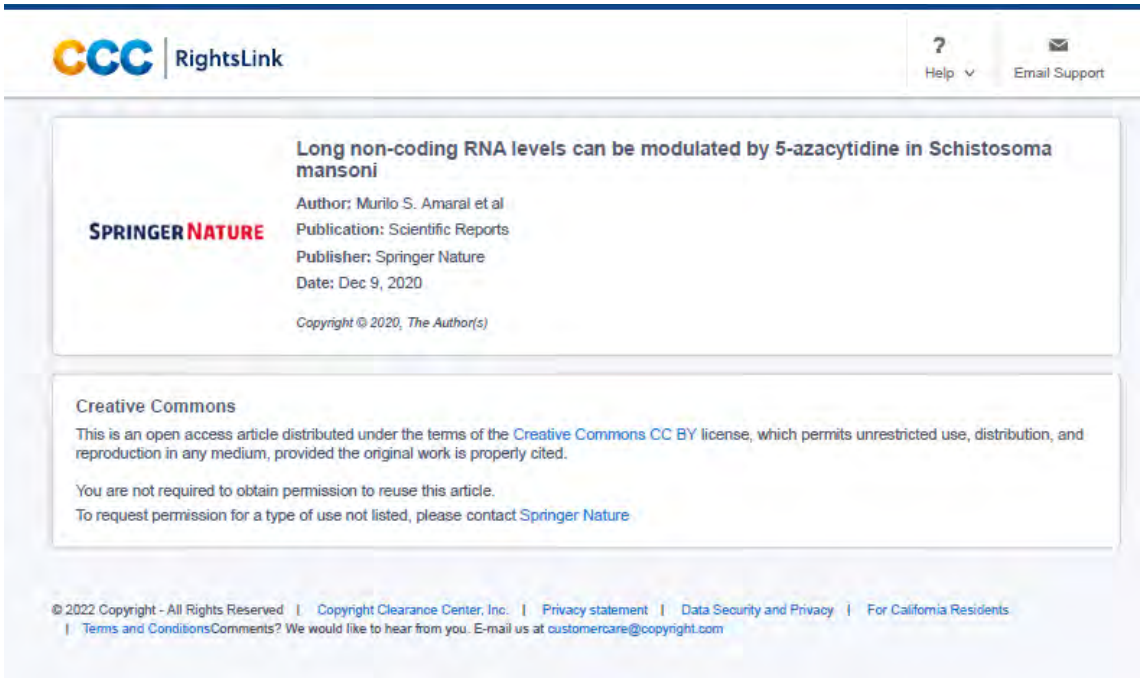
Figure S5. Relative expression of two different protein coding genes in *S. mansoni* under several culturing conditions, normalized by three reference genes found in the literature. **(A)** Female adult worm gene expression pattern of Smp_316140 gene (Protein p14) across different *in vitro* culturing conditions. Quantitative RT-qPCR was performed with RNA samples from females that were paired (P) or unpaired (U) to males and cultivated *in vitro* for 2, 4 or 8 days. Day 0 stands for paired females retrieved right after perfusion. **(B)** Male and female adult worm Smp_165220 (*EED*) gene expression in samples obtained from couples treated with dsRNA targeting either the control unrelated dsmCherry gene (C) or the Smp_165220 (*EED*) gene (E) *in vitro* for 2, 4, 8 or 16 days. The expression values are represented as the relative expression using as normalizer the geometric mean of three commonly used reference genes, Smp_056970 (*GAPDH*), Smp_307020 (*Actin*) and Smp_090120 (*α-tubulin*). Bars represent the standard deviation of the mean from three biological replicates for each experiment. Three technical replicates were assayed for each of the three biological replicates. Unpaired student t-test was used to calculate the statistical significance of the expression differences in the comparisons (* p-value ≤ 0.05; ** p-value ≤ 0.01; *** p-value ≤ 0.001; **** p-value ≤ 0.0001; ns = p-value > 0.05). The p-value obtained from the Student t-test is represented under the brackets.

2.3. Níveis de RNA longos não-codificadores podem ser modulados por 5-azacitidina em *Schistosoma mansoni*

PREÂMBULO

Contribuições do Doutorando Gilbert de Oliveira Silveira para o manuscrito apresentado nesta sessão:

Elaboração e desenvolvimento de todos os experimentos de RT-qPCR.



CCC | RightsLink

Help | Email Support

SPRINGER NATURE

Long non-coding RNA levels can be modulated by 5-azacytidine in Schistosoma mansoni

Author: Murilo S. Amaral et al
Publication: Scientific Reports
Publisher: Springer Nature
Date: Dec 9, 2020

Copyright © 2020, The Author(s)

Creative Commons

This is an open access article distributed under the terms of the Creative Commons CC BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article.
To request permission for a type of use not listed, please contact Springer Nature

© 2022 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Data Security and Privacy | For California Residents | Terms and ConditionsComments? We would like to hear from you. E-mail us at customercare@copyright.com

scientific reports

Check for updates

OPEN

Long non-coding RNA levels can be modulated by 5-azacytidine in *Schistosoma mansoni*

Murilo S. Amaral[1], Lucas F. Maciel[1], Gilbert O. Silveira[1,2], Giovanna G. O. Olberg[1], João V. P. Leite[1], Lucas K. Imamura[1], Adriana S. A. Pereira[1,2], Patricia A. Miyasato[1], Eliana Nakano[1] & Sergio Verjovski-Almeida[1,2]✉

Schistosoma mansoni **is a flatworm that causes schistosomiasis, a neglected tropical disease that affects more than 200 million people worldwide. There is only one drug indicated for treatment, praziquantel, which may lead to parasite resistance emergence. The ribonucleoside analogue 5-azacytidine (5-AzaC) is an epigenetic drug that inhibits *S. mansoni* oviposition and ovarian development through interference with parasite transcription, translation and stem cell activities. Therefore, studying the downstream pathways affected by 5-AzaC in *S. mansoni* may contribute to the discovery of new drug targets. Long non-coding RNAs (lncRNAs) are transcripts longer than 200 nucleotides with low or no protein coding potential that have been involved in reproduction, stem cell maintenance and drug resistance. We have recently published a catalog of lncRNAs expressed in *S. mansoni* life-cycle stages, tissues and single cells. However, it remains largely unknown if lncRNAs are responsive to epigenetic drugs in parasites. Here, we show by RNA-Seq re-analyses that hundreds of lncRNAs are differentially expressed after in vitro 5-AzaC treatment of *S. mansoni* females, including intergenic, antisense and sense lncRNAs. Many of these lncRNAs belong to co-expression network modules related to male metabolism and are also differentially expressed in unpaired compared with paired females and ovaries. Half of these lncRNAs possess histone marks at their genomic *loci*, indicating regulation by histone modification. Among a selected set of 8 lncRNAs, half of them were validated by RT-qPCR as differentially expressed in females, and some of them also in males. Interestingly, these lncRNAs are also expressed in other life-cycle stages. This study demonstrates that many lncRNAs potentially involved with *S. mansoni* reproductive biology are modulated by 5-AzaC and sheds light on the relevance of exploring lncRNAs in response to drug treatments in parasites.**

Schistosomiasis is a very debilitating disease, spread across three continents with a global burden estimated by the World Health Organization at 2,543,364 DALYs (Disease Adjusted Life Years)[1]. It is estimated that schistosomiasis affects more than 200 million people in 74 countries[2,3]. The disease is caused by parasitic trematodes of the genus *Schistosoma*, being the three main species *Schistosoma mansoni*, *S. japonicum* and *S. haematobium*[4]. *S. mansoni* is the prevalent species in Latin America, with 1 to 3 million people infected and over 25 million living in risk areas mainly in Brazil and Venezuela[5].

Administration of praziquantel (PZQ) to infected individuals is the basis of current schistosomiasis therapy. PZQ is a safe, cheap and tolerable drug[6], however, cure rates of less than 50% have been recorded[7] and drug tolerance has already been reported[8,9]. This scenario reinforces the need of new and more efficient approaches in reducing morbidity or disease eradication, such as the development of a vaccine[10] or alternative drugs[11].

5-azacytidine (5-AzaC) is a ribonucleoside currently used to treat human myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML)[12]. 5-AzaC is considered an epigenetic drug as it can prevent DNA methylation by inhibition of DNA methyltransferases. It can also impede RNA methylation[13] and decrease protein synthesis[14]. In *S. mansoni*, 5-AzaC has been shown to inhibit biological processes related to female metabolism, including egg production, egg maturation and normal ovarian development[15]. In addition, 5-AzaC also significantly alters *S. mansoni* adult female transcription, translation and stem cell activities[16]. Therefore, the study of the downstream pathways affected by 5-AzaC in *S. mansoni* may contribute to the understanding of the epigenetic

[1]Laboratório de Parasitologia, Instituto Butantan, São Paulo, Brazil. [2]Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil. ✉email: verjo@iq.usp.br

control of gene expression and its physiological consequences in schistosomes and, in the future, to the possible development of new chemotherapeutic strategies against schistosomiasis.

Long non-coding RNAs (lncRNAs) are transcripts longer than 200 nucleotides with low or no protein coding potential[17,18] that in humans are involved in a wide range of biological processes, including cell cycle regulation, reproduction, stem cell maintenance and drug resistance[19]. While the functions of lncRNAs have been explored[20] and growing evidence suggests that they should be considered as drug targets in human diseases[21], the mechanisms of regulation of lncRNA expression are much less understood[17]. In helminths other than *S. mansoni*, just a few works have reported the identification of lncRNAs using transcriptomic approaches[22–24], however no further investigation of the mechanisms of lncRNA regulation or response to drug treatments were performed.

In *S. mansoni*, the expression of lncRNAs at different life-cycle stages was first detected by our group in 2011 using microarrays[25] and then subsequently reported by many groups using RNA-Seq approaches[26–31]. However, it is largely unknown if *S. mansoni* lncRNA levels may be regulated by drugs. In the present work, we have evaluated the effect of 5-AzaC on lncRNA expression in *S. mansoni* adult worms by performing a re-analysis of the public RNA-Seq data from Geyer et al.[16]. We show, for the first time, that an epigenetic drug affects lncRNA levels in *S. mansoni* and that many of these lncRNAs are also differentially expressed in unpaired females and ovaries, indicating involvement in parasite reproductive biology. Understanding the mechanisms of control of lncRNAs expression will help the identification of potential new therapeutic targets and may contribute to the development of novel therapeutic strategies in the future.

Results

A set of lncRNAs is differentially expressed in *S. mansoni* females upon 5-AzaC in vitro treatment. We reanalyzed the RNA-Seq public data generated by Geyer et al.[16] to search for long non-coding RNAs (lncRNAs) possibly affected by 5-AzaC treatment in *S. mansoni* females (Supplementary Table S1 shows the samples used and alignment statistics). In that study, Geyer et al.[16] had treated *S. mansoni* adult worm pairs in vitro (Puerto Rican strain/NMRI, obtained from mice) with 5-AzaC at 491 μM for 48 h, extracted RNA from the females and then performed RNA-Seq. Thirty adult worm couples were used in each of three biological replicates[16]. Geyer et al.[16] analyzed in that RNA-Seq dataset only the protein-coding genes differentially expressed in *S. mansoni* females after treatment with 5-AzaC. As lncRNA levels have been shown to be modulated by nucleoside analogs in other eukaryotes[32,33], we hypothesized that lncRNA levels would also be modulated by 5-AzaC in *S. mansoni*.

Indeed, the re-analysis of Geyer et al.[16] RNA-Seq data with a reference transcriptome that is comprised of protein-coding genes as well as lncRNAs[31] (see "Materials and methods"), found 912 lncRNAs differentially expressed in *S. mansoni* females upon 5-AzaC in vitro treatment. Among them, 522 were long intergenic non-coding RNAs (lincRNAs, being 353 upregulated and 169 downregulated), 358 were antisense non-coding RNAs (SmLNCAs, being 183 upregulated and 175 downregulated), and 32 were sense non-coding RNAs (SmLNCSs, being 16 upregulated and 16 downregulated) (Fig. 1). All differentially regulated protein-coding genes and lncRNAs are shown in Supplementary Table S2 with their transcript per million (TPM) values, and in Supplementary Table S3 with their raw counts.

We also found in our analysis 3219 protein-coding genes (corresponding to 3693 transcripts, 1655 upregulated and 2038 downregulated) differentially expressed after 5-AzaC treatment, of which 1810 have also been found as differentially expressed by Geyer et al.[16]. Geyer et al.[16] previously identified 4036 protein-coding genes differentially expressed after 5-AzaC treatment, and 3221 of these genes are contained in the *S. mansoni* genome v7 annotation; thus, we were able to retrieve 1810 out of 3221 (or 56%) protein-coding genes found as differentially expressed in that work, which is a reasonable proportion considering the difference in the genomes used for reads alignment (genome version 7 was used here versus genome version 5.2 in the previous work) and the different read-mapping and counting tools used for the analysis.

As expected, principle component analysis (PCA) resulted in transcriptomes of the 5-AzaC treated and control groups segregating broadly into two distinct regions with replicates from the same condition clustering together, both for control or 5-AzaC-treated groups (Supplementary Fig. S1).

Most of the lncRNAs differentially expressed upon 5-AzaC treatment in *S. mansoni* females belong to co-expression modules related to male metabolism. Previously, besides building a new *S. mansoni* transcriptome comprised of lncRNAs in addition to protein-coding genes[31], we also showed by weighted gene co-expression network analyses (WGCNA) that 6016 out of 16,583 lncRNAs identified in different *S. mansoni* life-cycle stages and tissues belong to one of 15 different lncRNAs/mRNAs co-expression modules[31] (Fig. 2A). Each of these 15 modules represents one cluster of highly interconnected lncRNA/mRNA genes that are more expressed in one given *S. mansoni* stage/tissue, including miracidia, sporocysts, cercariae, schistosomula, juveniles, adult males, adult females and gonads (testes and ovaries)[31]; also, some of the stage/tissues are represented by more than one module (Fig. 2A) (please refer to "Materials and methods" section for details).

Evaluating to which of the lncRNAs/mRNAs co-expression modules the lncRNAs differentially expressed after 5-AzaC treatment belong to, can help the understanding of the gene expression programs altered by this epigenetic drug on *S. mansoni* females. When we looked at the modules to which the 552 lncRNAs upregulated after 5-AzaC exposure in females belong to, we observed enrichment in a male-related module (Fig. 2B). Out of the 552 lncRNAs upregulated after 5-AzaC treatment, 450 were assigned to any module, being the top three most represented enriched modules: adult males (yellow module, 114 lncRNAs or 20% of the lncRNAs, p value < 0.0001, hypergeometric test), gonads (brown module, 71 lncRNAs or 13% of the lncRNAs, p value < 0.001,

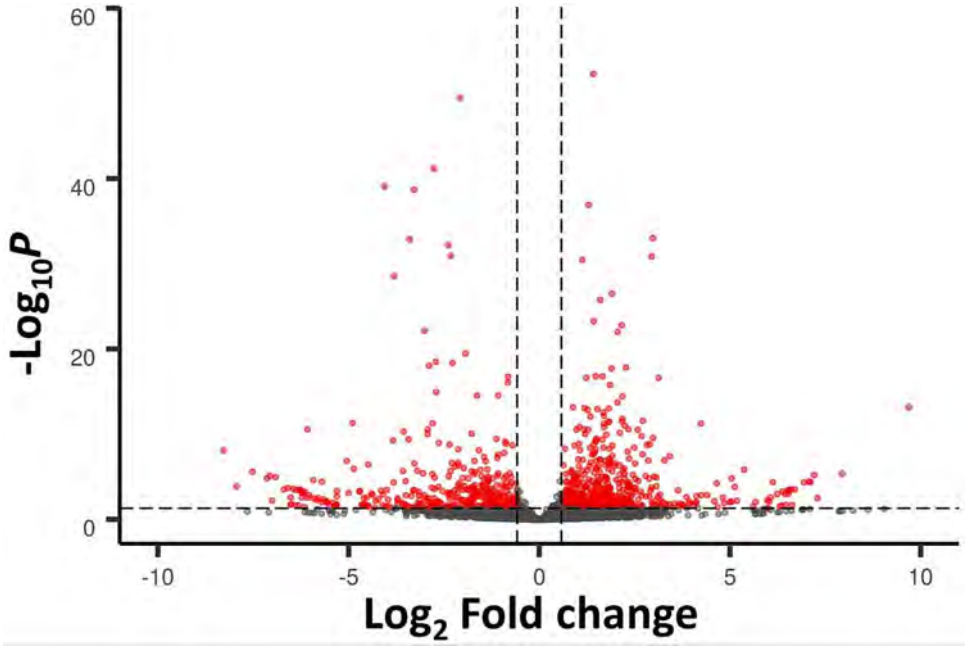


Figure 1. Differentially expressed long non-coding RNAs (lncRNAs) detected by RNA-Seq in adult *S. mansoni* females treated with 5-AzaC. These results were obtained by re-analysis of the RNA-Seq data from Geyer et al.[16] using the *S. mansoni* lncRNA transcriptome published in Maciel et al.[31] as reference. In Geyer et al.[16], parasites were cultured either in the presence or absence of 491 µM 5-AzaC for 48 h. LncRNA gene expression levels identified in this RNA-Seq dataset are shown with a volcano plot, which displays the differentially expressed lncRNAs between 5-AzaC-treated and control *S. mansoni* females (red dots, showing FDR < 0.05 and log2FC > 0.59 or < -0.59, dotted lines). Grey dots represent non-differentially expressed lncRNAs. 912 lncRNAs were considered significantly differentially expressed, being 522 long intergenic ncRNAs, 358 antisense lncRNAs and 32 sense lncRNAs.

hypergeometric test) and schistosomula (magenta module, 49 lncRNAs or 9% of the lncRNAs, p value < 0.01, hypergeometric test) (Fig. 2B).

Out of the 360 lncRNAs downregulated after 5-AzaC treatment, 296 were assigned to any module, being the three top most represented modules: adult males (turquoise module, 98 lncRNAs or 27% of the lncRNAs, p value < 0.0001, hypergeometric test), gonads (brown module, 29 lncRNAs or 8% of the lncRNAs, p value = 0.057, hypergeometric test) and adult females (pink module, 27 lncRNAs or 7% of the lncRNAs, p value < 0.01, hypergeometric test) (Fig. 2C). It is possible that 5-AzaC treatment in females switches the lncRNA transcriptional program to a pattern more similar to that shown by males, as it was shown for protein-coding genes in females treated with GSK343, an histone methyltransferase EZH2 inhibitor[34], or in unpaired females, in which the gonads are not developed[35]. This effect can impact on stem cell activity and egg production by females, as previously shown by Geyer et al.[16].

Similar patterns of module distribution of protein-coding genes in the Maciel et al.[31] dataset (Fig. 2D) and of protein-coding genes differentially expressed after 5-AzaC treatment (Fig. 2E,F) were obtained in our re-analysis of the Geyer et al.[16] dataset.

The list of lncRNAs and protein-coding genes differentially expressed after 5-AzaC treatment, as well as the modules to which they belong are given in Supplementary Table S4.

Involvement of lncRNAs with the parasite reproductive biology. In order to evaluate if the lncRNAs differentially expressed after 5-AzaC treatment could be involved in *S. mansoni* reproductive biology, we checked if these lncRNAs are also differentially expressed in pairing-dependent conditions or in reproductive organs compared with whole worms. To do that, we cross compared the lncRNAs differentially expressed after 5-AzaC exposure in females with lncRNAs that we found to be differentially expressed in a re-analysis of the Lu et al. data[35] (please refer to "Materials and methods" section for details) for lncRNAs differentially expressed between bisex (paired) females (bF) and single-sex females (sF), between bisex ovaries (bO) and single-sex ovaries (sO) and between bisex ovaries(bO) and bisex females (bF). We found that 60% of the lncRNAs downregulated after 5-AzaC exposure (216 out of 360 lncRNAs) are also present in at least one of these comparisons (Supplementary Fig. S2A,B, see overlap between the yellow oval and the other ovals). When the statistical significance of the overlaps between the lncRNAs downregulated after 5-AzaC exposure and each of the above comparisons was calculated, all of the overlaps were statistically significant, with all the p values obtained from the pairwise comparisons lower than 1.254e−10 (hypergeometric test).

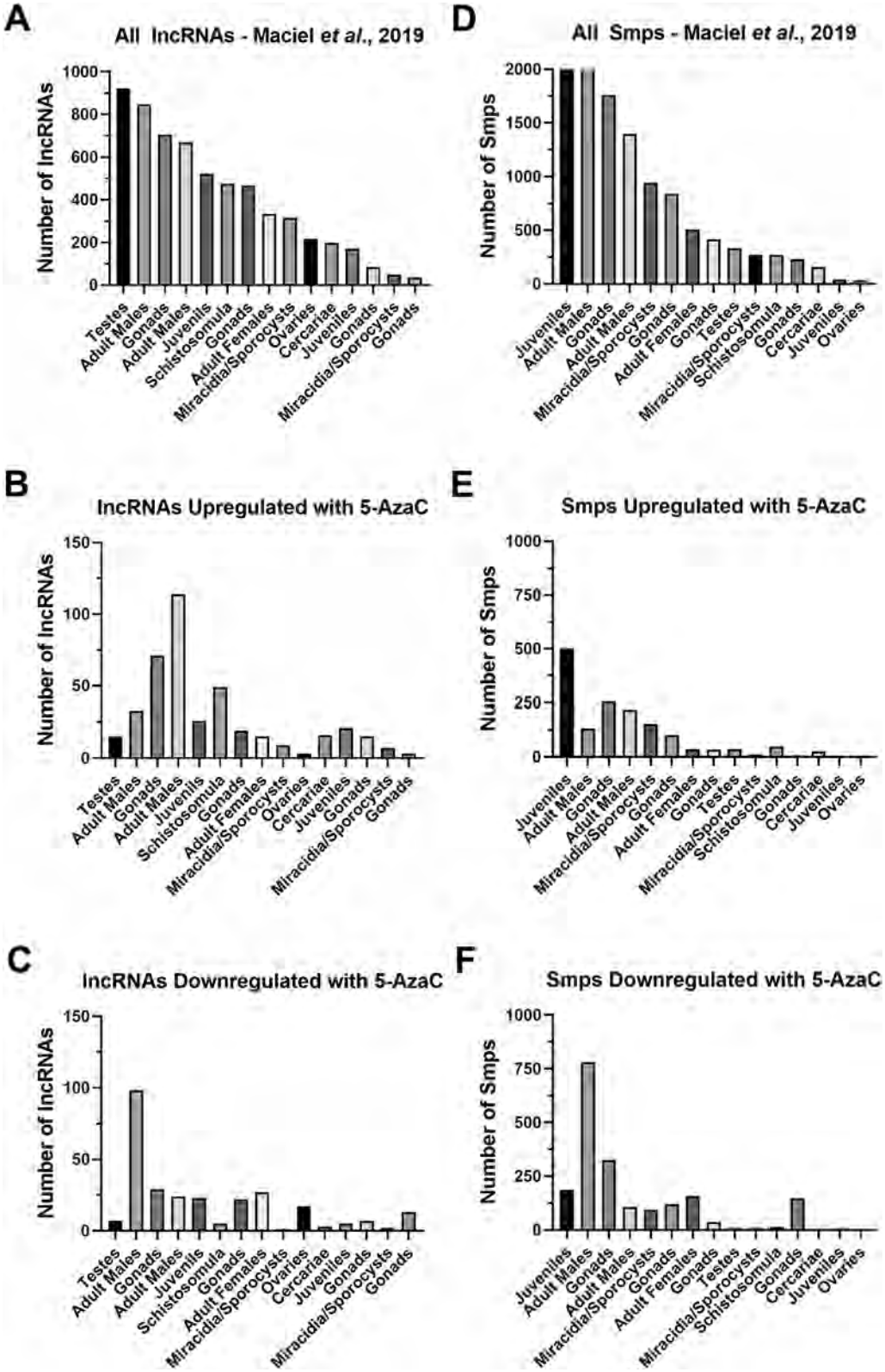


Figure 2. Distribution of 5-AzaC-affected long non-coding RNAs and protein-coding genes among the weighted gene co-expression network (WGCNA) modules (represented by life-cycle stages and tissues). (**A**) Number of lncRNAs detected in each of 15 different WGCNA modules, according to Maciel et al.[31]; note that different modules are associated to the same *S. mansoni* life-cycle stage/tissue. (**B**) Number of lncRNAs upregulated in 5-AzaC treated females in each of the 15 WGCNA modules. (**C**) Number of lncRNAs downregulated in 5-AzaC treated females in each of the 15 WGCNA modules. (**D**) Number of protein-coding genes detected in each of 15 different WGCNA modules, according to Maciel et al.[31]; note that different modules are associated to the same *S. mansoni* life-cycle stage/tissue. (**E**) Number of protein-coding genes upregulated in 5-AzaC treated females in each of the 15 WGCNA modules. (**F**) Number of protein-coding genes downregulated in 5-AzaC treated females in each of the 15 WGCNA modules.

In addition, 23% of the lncRNAs upregulated after 5-AzaC exposure (127 out of 552 lncRNAs) are also present in at least one of these comparisons (Supplementary Fig. S2C,D, see overlap between the yellow oval and the other ovals). When the statistical significance of the overlaps between the lncRNAs upregulated after 5-AzaC exposure and each of the comparisons was calculated, all of the overlaps were statistically significant (with all the p values obtained for the pairwise comparisons lower than 0.037, hypergeometric test), except for two pairwise comparisons: "lncRNAs upregulated after 5-AzaC versus bF > sF" (p value = 0.119) and "lncRNAs upregulated after 5-AzaC versus sF > bF" (p value = 0.181).

It has been shown that juvenile worms and schistosomula co-express transcripts that cluster into modules midnightblue and magenta, respectively[31]. These modules are among those with higher lncRNA/total transcripts ratio when compared with all modules: midnightblue is the second and magenta is the fourth module, out of 15 modules, with the highest lncRNAs/total transcript ratios. In midnightblue and magenta modules, lncRNAs correspond to 80% and 64% of all the transcripts, respectively[31]. As a high proportion of lncRNAs upregulated by 5-AzaC in *S. mansoni* females belongs to midnightblue and magenta modules (16.7% of all upregulated lncRNAs assigned to any module), we tested if 5-AzaC treatment would have any impact on schistosomula viability. We treated schistosomula with different concentrations of 5-AzaC and measured the viability at each day, along 5 days of treatment. No statistically significant reduction in schistosomula viability as measured by ATP levels was observed after 5-AzaC treatment at any of the concentrations and days tested (Supplementary Fig. S3A), with discrete phenotypic alterations observed only at day 5 post-treatment, at 245 µM, the highest concentration tested (Supplementary Fig. S3B). This is in agreement with observations that *S. mansoni* schistosomula possess lower detectable levels of 5-methylcytosine and of mRNAs encoding SmDnmt2 and SmMBD proteins involved with DNA methylation[15], compared with other *S. mansoni* life-cycle stages, thus probably making schistosomula less susceptible to 5-AzaC treatment.

LncRNAs differentially expressed upon 5-AzaC treatment have histone marks at their genomic *loci*. The presence of histone marks at the TSSs of lncRNAs adds another layer of functionality evidence for lncRNAs, indicating regulation by epigenetic mechanisms related to chromatin structure[26,31]. Therefore, to check if the lncRNAs differentially expressed after 5-AzaC treatment would have histone marks at their TSSs, we cross compared the lncRNAs affected by 5-AzaC treatment with lncRNAs expressed in *S. mansoni* and reported by Maciel et al.[31] as having at least one histone mark obtained by ChIP-Seq (H3K4me3, that is generally associated with active transcription or H3K27me3, associated to transcription repression) in non-treated *S. mansoni* cercariae, schistosomula or adults. As reported in that work[31], 8599 out of 16,583 lncRNAs identified in different *S. mansoni* life-cycle stages and tissues have at least one histone modification mark within 1 kb from their TSS[31]. In addition, gene expression control by DNA/RNA methylation, affected by 5-AzaC, has been linked to histone modifications in eukaryotes[36,37].

A total of 461 out of 912 lncRNA transcripts differentially expressed after 5-AzaC treatment have at least one histone modification mark within 1 kb from their TSS, being 274 upregulated lncRNAs (Supplementary Table S5) and 187 downregulated lncRNAs (Supplementary Table S6). This represents 50% of all the 912 lncRNAs differentially expressed after 5-AzaC treatment, which is statistically significant (p value < 0.05, hypergeometric test).

The most abundant mark found individually at the *loci* of the lncRNAs differentially expressed after 5-AzaC treatment was H3K27me3 in adults, for both upregulated lncRNAs (with 35 marks, Fig. 3A) and downregulated lncRNAs (with 31 marks, Fig. 3B). The second and third most present marks were, among the upregulated lncRNAs, H3K27me3 in cercariae and H3K4me3 in schistosomula, and among the downregulated lncRNAs H3K27me3 in schistosomula and H3K27me3 in cercariae.

In addition, when computed together with other marks, the most abundant mark found in the upregulated lncRNAs upon 5-AzaC treatment (Fig. 3A) was the transcriptional repressive mark, H3K27me3, with 18 lncRNAs presenting this mark in adults and schistosomula and other 14 lncRNAs presenting this mark in adults and cercariae simultaneously. For the downregulated lncRNAs upon 5-AzaC treatment (Fig. 3B), H3K27me3 was also the most abundant mark found when the marks were computed together, with 17 lncRNAs presenting this mark in adults and schistosomula and other 10 lncRNAs presenting this mark in adults, schistosomula and cercariae simultaneously.

In Fig. 3C, we show the *locus* on chromosome 3 of SmLINC133371-IBu (orange track), a lincRNA that has H3K27me3 histone mark ChIP-Seq peaks ("ChIP-Seq Control H3K27me3" track at the bottom of the image) in adults (blue/green tracks). This lincRNA is upregulated 1.5 × in females after 5-AzaC treatment (yellow track, "5-AzaC treated RNA-Seq"), belongs to the greenyellow module and also has RNA Polymerase II peaks ("ChIP-Seq Control RNAPol II") at its *locus*.

Validation of lncRNAs differential expression by RT-qPCR. We designed PCR primer pairs for a selected set of ten genes, including eight lincRNAs and two protein-coding genes, to validate their differential expression after 5-AzaC treatment. First, we treated adult worm couples with 5-AzaC at 491 µM for 48 h, extracted RNA from females and males separately and then performed RT-qPCR. As observed by Geyer et al.[15], 5-AzaC was not lethal to adult worms even when they were treated with 5-AzaC at 491 µM, the limit of aqueous solubility. Here, we measured for the first time the amount of ATP in adult worms upon 5-AzaC exposure, as readout for worm viability. 5-AzaC exposure for 48 h did not alter significantly the ATP content of adult worms when compared with the controls (Supplementary Fig. S4A, p = 0.12). In addition, we observed a statistically significant 49% reduction in egg laying by adult worms treated with 5-AzaC at 491 µM (p = 0.02, Supplementary Fig. S4B). Eggs laid by adult worms treated with 5-AzaC show many phenotypic abnormalities, including lack of lateral spine on some eggs and eggs with smaller sizes (Supplementary Fig. S4C).

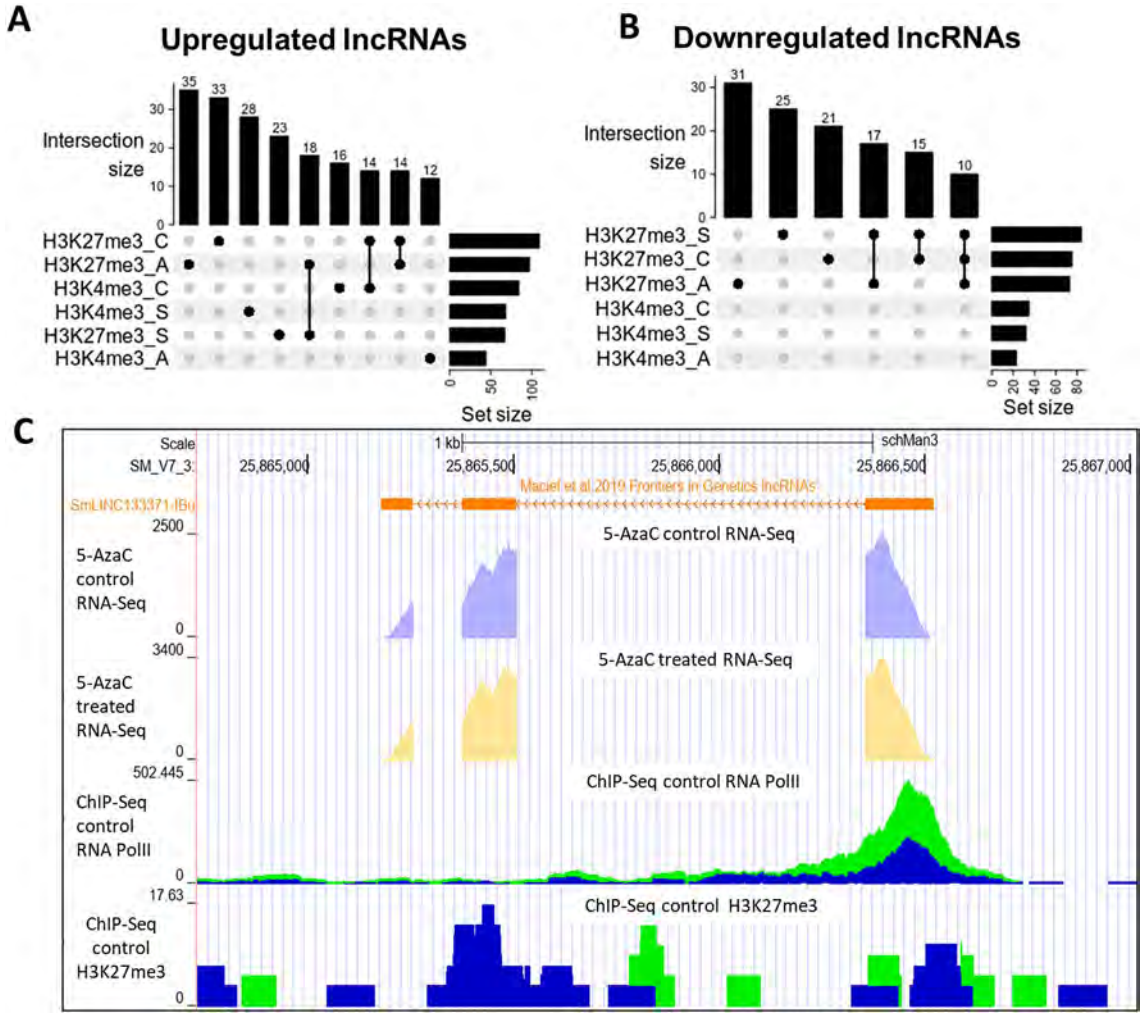


Figure 3. Hundreds of lncRNAs differentially expressed after 5-AzaC exposure in *S. mansoni* females have histone transcriptional activating or repressive marks at their TSSs. The UpSet intersection diagram shows the number of *S. mansoni* lncRNAs differentially expressed after 5-AzaC exposure (y-axis) that have been detected in each of the intersection sets, indicated by the connected points in the lower part of the plot, as having the H3K4me3 transcriptional activating marks and/or the H3K27me3 repressive marks within 1 kb (upstream or downstream) from their TSSs. Six histone mark datasets indicated at the bottom left were analyzed: H3K4me3_A in adults, H3K4me3_C in cercariae, H3K4me3_S in schistosomula, H3K27me3_A in adults, H3K27me3_S in schistosomula, and H3K27me3_C in cercariae, and each set size horizontal black bar represents the number of lncRNAs that contain the indicated histone mark at the indicated stage. The top enriched intersection sets are shown for the 5-AzaC upregulated (**A**) and downregulated (**B**) lncRNAs; all intersection sets and the lists of lncRNAs in each intersection set are shown in Supplementary Table S5, S6. (**C**) Snapshot of a *S. mansoni* genome browser image (www.schistosoma.usp.br), showing a region spanning 3 kb on chromosome 3, where the SmLINC133371-IBu is located. The orange track (top) represents lncRNAs from *S. mansoni* published by Maciel et al.[31]. Below the orange lncRNAs track, two other tracks show RNA-Seq data from control (light purple) or 5-AzaC treated *S. mansoni* females (light yellow). Below, two ChIP-Seq tracks are shown: RNA Polymerase II ChIP-Seq (ChIP-Seq Control RNA Pol II) and H3K27me3 histone mark ChIP-Seq (ChIP-Seq Control H3K27me3). The green and blue colours at the two ChIP-Seq tracks at the bottom represent each of two experimental biological replicates.

Quantitative real-time PCR (RT-qPCR) was then employed to validate results obtained by the RNA-Seq analysis. Two protein-coding genes were used as controls: Smp_151640 (*Insulin-like growth factor I*), which was 14.6 × upregulated in the RNA-Seq after 5-AzaC treatment (Supplementary Fig. S5A, left) and Smp_121390 (*Genome polyprotein*), which was downregulated 3.6 × in the RNA-Seq data (Supplementary Fig. S5B, left). In the RT-qPCR, both protein-coding genes were validated in females: Smp_151640 was upregulated 12.1 × (Supplementary Fig. S5A, right) and Smp_121390 was downregulated 2.5 × (Supplementary Fig. S5B, right) after 5-AzaC treatment. In addition, we also tested the expression of both Smp_151640 and Smp_121390 after 5-AzaC treatment in *S. mansoni* males. While Smp_151640 was found to be 51.4 × upregulated (Supplementary Fig. S6A), Smp_121390 was not differentially expressed after 5-AzaC in vitro treatment in males (Supplementary Fig. S6B).

We then tested by RT-qPCR in *S. mansoni* females and males a selected set of eight lincRNAs found to be differentially expressed in the female RNA-Seq dataset: SmLINC133371-IBu, SmLINC151825-IBu, SmLINC158444-IBu, SmLINC110084-IBu, SmLINC158969-IBu, SmLINC156349-IBu, SmLINC103888-IBu and SmLINC100882-IBu. These lincRNAs were selected because they show a wide range of expression levels in the RNA-Seq (TPM from 4 to 1635 in at least one of the conditions, control or 5-AzaC treated), because they have fold-changes higher than 1.5 × in the RNA-Seq dataset and because they all show only one isoform at their loci, except for SmLINC151825-IBu.

Four of these lincRNAs (SmLINC133371-IBu, SmLINC151825-IBu, SmLINC158969-IBu and SmLINC156349-IBu) were validated by RT-qPCR in females, confirming the RNA-Seq data; in our assays with *S. mansoni* females, they were upregulated 2.8 ×, 5.4 ×, 1.6 × and 2.2 ×, respectively (Fig. 4A,B,E,F). Additionally, four other lincRNAs tested were detected as expressed in the RT-qPCR assays; however, they were not differentially expressed after 5-AzaC treatment as predicted by the RNA-seq data (Fig. 4C,D,G,H). This indicates that there is variability of lncRNA expression and response to 5-AzaC exposure, probably related to the different parasite strains used in our RT-qPCR assays and in the RNA-Seq experiments from the literature[16].

Considering the six genes in which the effect of 5-AzaC was validated in females by RT-qPCR (four lincRNAs and the two protein-coding genes), the extent of the effect measured by RT-qPCR mirrored the one obtained with RNA-Seq, as fold changes in expression were well correlated (Pearson correlation coefficient = 0.9334, p value = 0.0065, Supplementary Fig. S7).

Interestingly, two out of the eight lincRNAs that were tested (SmLINC133371-IBu and SmLINC151825-IBu) were also upregulated in males treated with 5-AzaC (6.0 × and 24.6 ×, respectively, Fig. 5A,B), indicating that these lincRNAs share similar regulatory mechanisms in both sexes. Expression of the other six tested lincRNAs was not significantly affected by 5-AzaC treatment of males (Fig. 5C–H).

LncRNAs modulated by 5-AzaC are differentially expressed along *S. mansoni* life-cycle stages. To evaluate if the lncRNAs differentially expressed after 5-AzaC treatment tested by RT-qPCR here are also expressed in other *S. mansoni* life-cycle stages or tissues, we re-analyzed data from public RNA-Seq libraries from different *S. mansoni* life-cycle stages and tissues (Supplementary Table S7) to look for the expression patterns of the eight selected lincRNAs. First, we evaluated the stage-specificity of the different RNA-Seq datasets that we used for this re-analysis by confirming that five protein-coding genes previously described as stage markers[38,39] were indeed more highly expressed at the predicted stages in our analysis (Supplementary Fig. S8). In addition, PCA analysis (Supplementary Fig. S9) shows that biological replicates of the same sample grouped according to the life-cycle stages and tissues, confirming the clustering of samples in expected segregating groups.

We then looked at the expression levels along *S. mansoni* life-cycle stages and tissues of the eight lincRNAs tested by RT-qPCR and observed a heterogeneous expression pattern distribution (Fig. 6). Expression of SmLINC133371-IBu (Fig. 6A), SmLINC151825-IBu (Fig. 6B) and SmLINC103888-IBu (Fig. 6G) is higher in miracidia and sporocysts stages, with SmLINC133371-IBu and SmLINC151825-IBu being also highly expressed in adult females and schistosomula (Fig. 6A,B). Whereas SmLIN158444-IBu shows higher expression in the posterior adult somatic tissues and tails (Fig. 6C), SmLINC110084-IBu has higher expression in schistosomula and cercariae (Fig. 6D). While SmLINC156349-IBu (Fig. 6E) and SmLINC158969-IBu (Fig. 6F) show broad expression in all the stages, SmLINC100882-IBu (Fig. 6H) is highly expressed in female adult worms and tails. These results show that most of the tested lncRNAs (except SmLINC103888-IBu and SmLINC100882-IBu) are not stage-specific and may play roles in other *S. mansoni* life-cycle stages.

Discussion

Here, we have shown that long non-coding RNAs levels can be modulated in *S. mansoni* by 5-AzaC, a DNA methyltransferase inhibitor that is currently used to treat myelodysplastic syndrome and acute myeloid leukemia in humans[12,40]. Hundreds of the lncRNAs differentially expressed after 5-AzaC exposure in *S. mansoni* females belong to co-expression modules related to male metabolism, have histone marks at their genomic *loci* and are also differentially expressed in unpaired compared with paired *S. mansoni* females and ovaries. While short RNAs (especially miRNAs) have been more explored in various helminths[41–44], lncRNAs have received little attention, being identified by transcriptomic approaches only in a few helminths other than *S. mansoni*[22–24] or studied in a limited number of free-living nematodes[45,46]. In addition, unlike miRNAs[47–52], the mechanisms of regulation of lncRNAs are largely unknown in parasites and, to our knowledge, this is the first report of modulation of lncRNAs levels by an epigenetic drug in any helminth.

In the past few years, human lncRNAs have been proposed as drug targets in many diseases, especially in cancer and neurological syndromes[21,53–55]. In parasitic diseases, there is a clear need to develop new and inexpensive drugs, especially with the emerging reports of drug resistance[6,56–58]. We believe that it is time to consider lncRNAs as possible drug targets also in parasitic diseases, especially because they show lower conservation in their primary sequences between species than protein-coding genes[17,19,59,60], which in principle would reduce side effects in therapeutic strategies.

The choice of lncRNAs to be further validated as drug targets will rest on the appropriate selection of lncRNA candidates. This selection should be guided by functional characterization of the lncRNA as well as by the demonstration of the lncRNA relevance to the parasite biology. Here, we found that 38% of the lncRNAs differentially expressed after 5-AzaC treatment in *S. mansoni* females (343 out of 912 lncRNAs) are also differentially expressed between paired and unpaired females or ovaries[35], whilst 24% of them (221 out of 912 lncRNAs) belong to co-expression modules related to "gonads"[31], indicating an important involvement of lncRNAs on parasite sexual maturation and reproductive biology. In addition, 50% of the lncRNAs differentially expressed after 5-AzaC

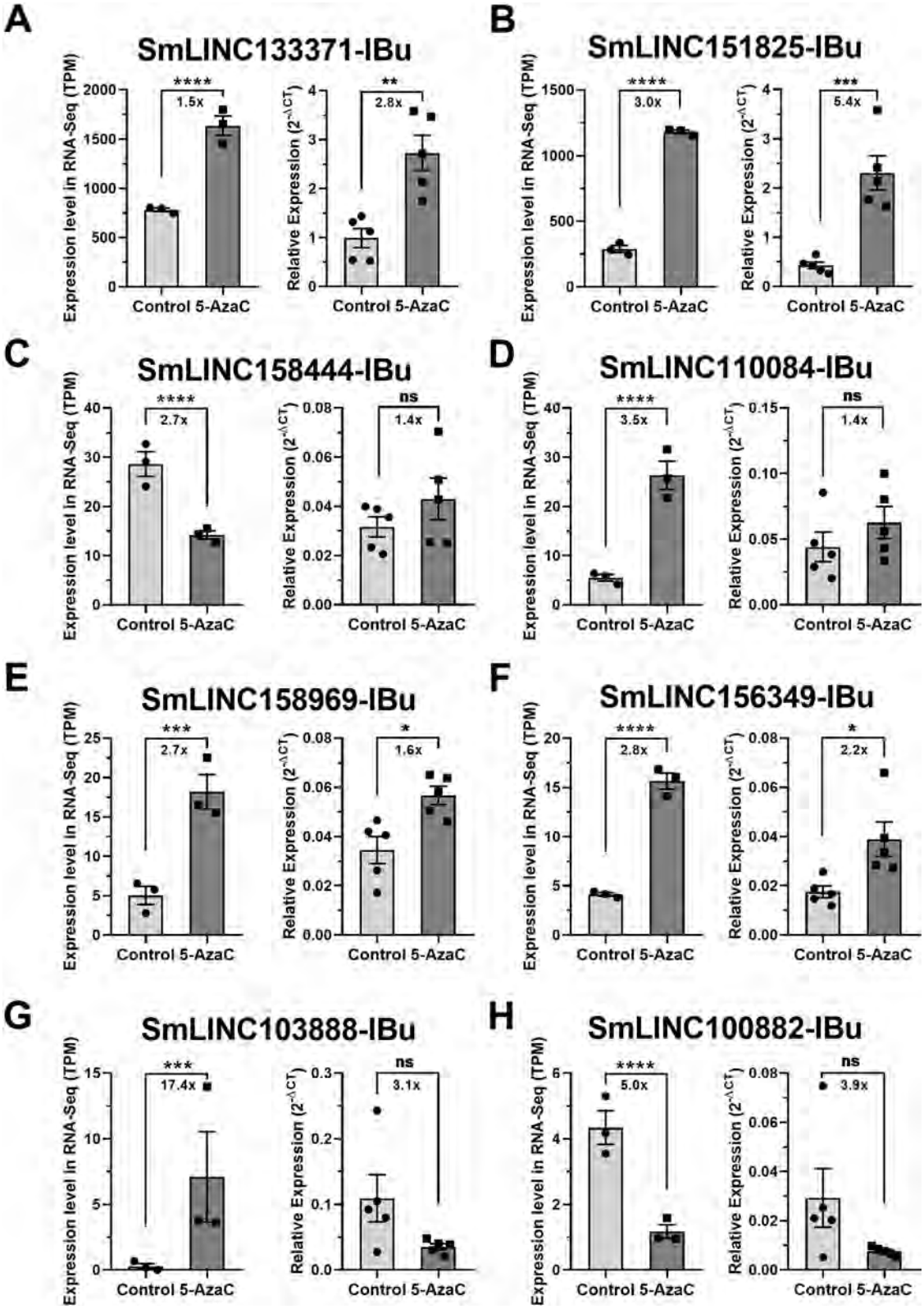


Figure 4. Expression profiles in *S. mansoni* females of selected lincRNAs differentially expressed after 5-AzaC treatment (491 µM). Eight lincRNAs were selected after re-analysis of RNA-Seq public datasets of 5-AzaC treated *S. mansoni* females[16] for validation by RT-qPCR in females. For each of the eight selected lincRNAs, the expression profiles obtained with RNA-Seq are shown on the left as TPM (transcripts per million), whereas the RT-qPCR results are shown on the right: (**A**) SmLINC133371-IBu; (**B**) SmLINC151825-IBu; (**C**) SmLINC158444-IBu; (**D**) SmLINC110084-IBu; (**E**) SmLINC158969-IBu; (**F**) SmLINC156349-IBu; (**G**) SmLINC103888-IBu; (**H**) SmLINC100882-IBu. For the RNA-Seq data, three biological replicates were analyzed; the fold-changes and p values represented by asterisks that are shown in the brackets were obtained using DESeq2. For the RT-qPCR data, mean ± SEM from five biological replicates are shown, and Student unpaired two-sided t test was applied. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$; ns: not significant.

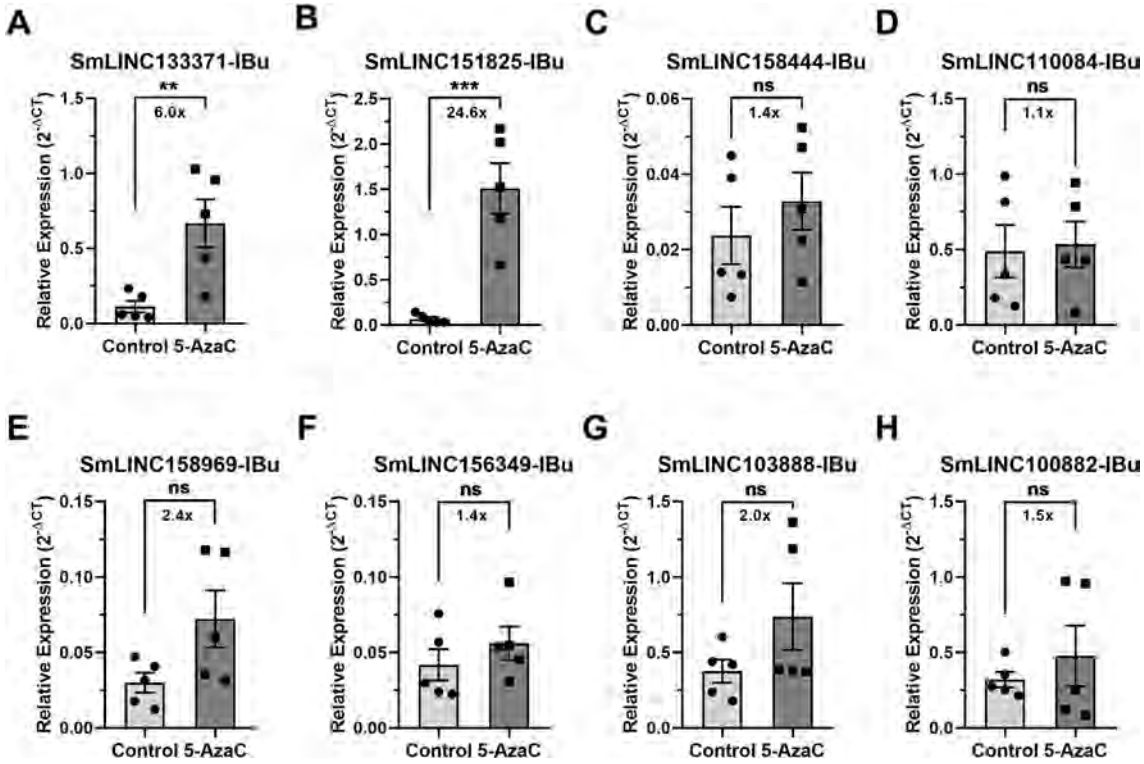


Figure 5. Expression profiles in *S. mansoni* males of selected lincRNAs differentially expressed after 5-AzaC treatment (491 μM). Eight lincRNAs were selected after re-analysis of RNA-Seq public datasets of 5-AzaC treated *S. mansoni* females[16] for evaluation of differential expression by RT-qPCR in *S. mansoni* males. For each of the eight lincRNAs, the expression profiles in controls and in 5-AzaC treated *S. mansoni* males by RT-qPCR are shown: (**A**) SmLINC133371-IBu; (**B**) SmLINC151825-IBu; (**C**) SmLINC158444-IBu; (**D**) SmLINC110084-IBu; (**E**) SmLINC158969-IBu; (**F**) SmLINC156349-IBu; (**G**) SmLINC103888-IBu; (**H**) SmLINC100882-IBu. Mean ± SEM from five biological replicates are shown, and Student unpaired two-sided t test was applied; $^{**}p < 0.01$, $^{***}p < 0.001$; *ns* not significant.

treatment in *S. mansoni* females (461 out of 912 lncRNAs) have at least one histone mark at their TSSs previously detected at *S. mansoni* life-cycle stages[31]. These lncRNAs with evidence of chromatin marks at their genomic loci could be prioritized in further functional assays to elucidate their relevance, roles and mechanisms of action in *S. mansoni* biology. Technologies for lncRNAs targeting should be considered in these studies, including cell and tissue localization, silencing by CRISPR or antisense oligonucleotides methods in vitro and in vivo, and discovery of lncRNA partners (DNA, RNA or proteins)[61–63].

The expression patterns along life-cycle stages may also be criteria for the selection of lncRNAs to be tested in functional assays. Some of the lncRNAs tested by RT-qPCR here in *S. mansoni* female and male adult worms are also expressed at high levels (TPM > 100) in other life-cycle stages, including SmLINC133371-IBu and SmLINC151825-IBu with high expression levels in miracidia, sporocysts and schistosomula. In addition, SmLINC158444-IBu and SmLINC110084-IBu are highly expressed in posterior somatic tissues and schistosomula, respectively. All these lncRNAs, except SmLINC100882-IBu are expressed in schistosomula, another life-cycle stage of interest regarding drug targeting, as praziquantel has no efficacy against schistosomula[64].

Additionally, many lncRNAs have been associated with drug resistance in human cancers[65–67]. Here, by measuring ATP levels, we confirm that 5-AzaC treatment has no effect on the viability of *S. mansoni* adult worms, as previously shown[15,16]. Moreover, we show that *S. mansoni* schistosomula viability is also not affect by 5-AzaC. It is unclear why the parasites' viability is not affected by 5-AzaC, but since schistosomes show nucleoside auxotrophy[68], precise regulation of nucleoside analogs uptake may control their toxicity. It is also possible that the lncRNAs differentially expressed after 5-AzaC exposure may be involved in a 5-AzaC drug resistance mechanism, as shown for human cancer-related lncRNAs such as *HOTAIR* and *XIST*[69–71].

Understanding the mechanisms of lncRNA expression regulation may help the selection of lncRNAs for the development of new therapeutic strategies in the future. These mechanisms, which include epigenetic regulation by histone modification[72–74] at lncRNA genomic *loci* and DNA/RNA methylation already described in human lncRNAs[75–78] are, however, less understood than those of protein-coding genes[79,80]. It is now clear that epigenetic processes play important roles on schistosomes[81–83]. In fact, epigenetic mechanisms participate in schistosome phenotypic plasticity[84,85], in egg production and adult worm viability[34,86,87] as well as in schistosomula survival[88–92]. DNA methylation, one of the most studied epigenetic mechanisms, has been detected in *S. mansoni*[15], although the significance of DNA cytosine methylation (5mC) in this parasite has been somewhat controversial[93–95]. Here, we observed that 912 lncRNAs are differentially expressed after 5-AzaC exposure in *S. mansoni* females, all of them expressed at an average TPM > 0.1 in control or 5-AzaC treated samples. As

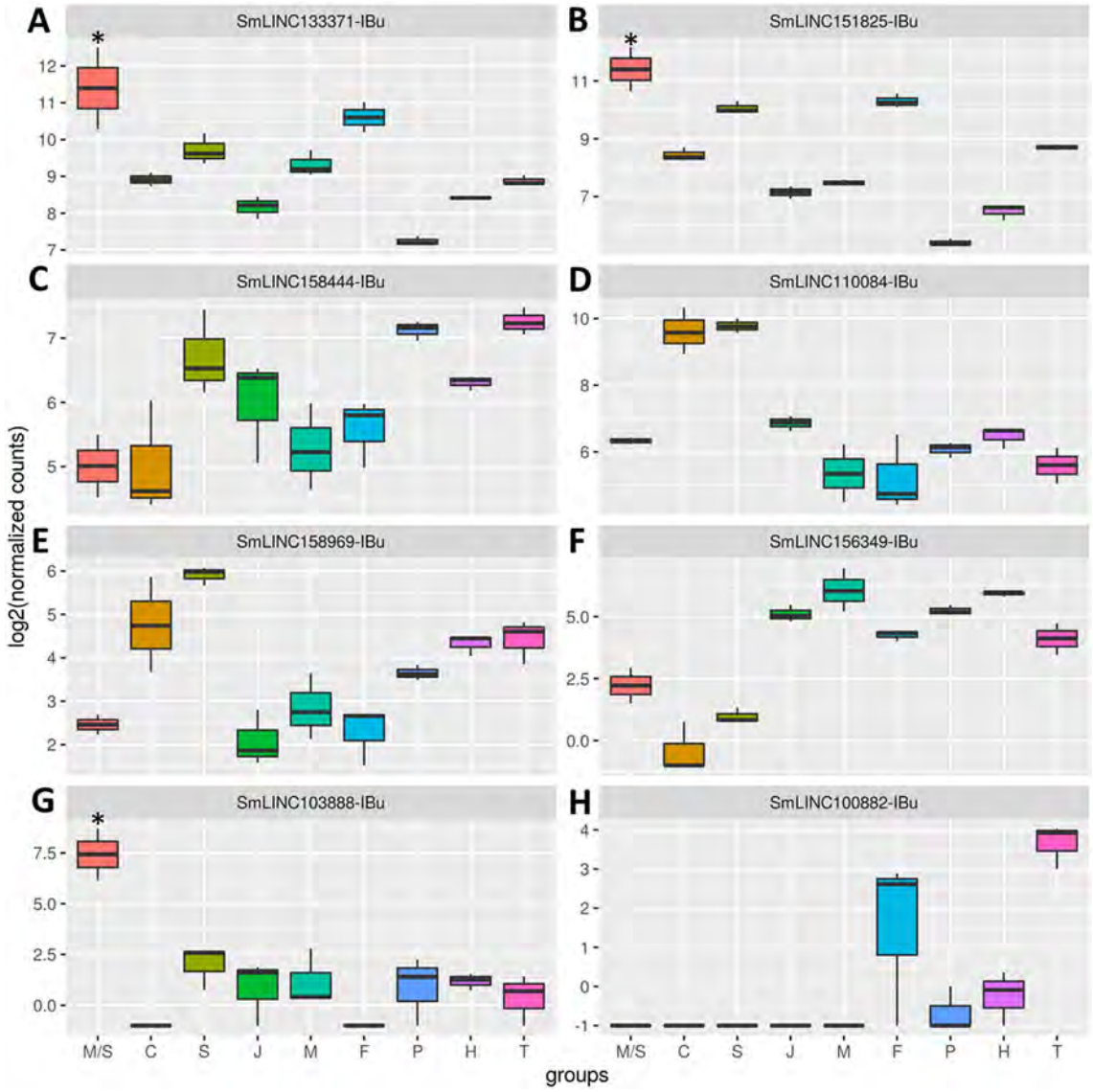


Figure 6. RNA-seq expression profiles at different *S. mansoni* stages of selected lincRNAs differentially expressed after 5-AzaC treatment (491 μM). The expression levels (shown as log2 of normalized counts) of the eight lincRNAs whose gene IDs are indicated at the top of each panel are shown. These lincRNAs were selected after re-analysis of RNA-Seq public datasets of 5-AzaC treated *S. mansoni* females[16]. The y-axis shows the expression level for each lincRNA in the RNA-seq assays (log2 of normalized counts) as determined at the stage indicated in the x-axis as follows: miracidia/sporocysts (M/S), cercariae (C), schistosomula (S), juveniles (J), adult males (M), adult females (F), posterior somatic tissues (P), heads (H) and tails (T). (**A**) SmLINC133371-IBu; (**B**) SmLINC151825-IBu; (**C**) SmLINC158444-IBu; (**D**) SmLINC110084-IBu; (**E**) SmLINC158969-IBu; (**F**) SmLINC156349-IBu; (**G**) SmLINC103888-IBu; (**H**) SmLINC100882-IBu. Only transcripts that were upregulated in one stage/tissue when compared with all others were considered as significantly more expressed in that stage/tissue and are marked with an asterisk. *p value < 0.05.

previous analysis identified 9229 lncRNAs expressed in females (TPM > 0.1) out of all 16,583 detected at any *S. mansoni* life-cycle stage[31], we estimate that 10% of all lncRNAs expressed in females are differentially expressed upon 5-AzaC exposure.

Although the presence of DNA methylation in many invertebrates has been already reported[96,97], previous work was unable to detect functional roles of DNA methylation in invertebrates[98]. 5-AzaC is an inhibitor of DNA methyltransferase that has been shown in *S. mansoni* to inhibit female specific biological processes including egg production, egg maturation and normal ovarian development[15,16], phenotypic effects confirmed in our treatments. These effects are achieved by modifications of adult female transcription and translation, with 81% inhibition in de novo protein synthesis in female schistosomes[16]. As 5-AzaC incorporates preferentially into RNAs, with only 20% being incorporated into DNA[99], it is more likely that 5-AzaC interferes preferentially in lncRNA stability through lncRNA methylation impediment than through promoter DNA methylation. In fact, many lncRNAs

were shown to be regulated by RNA methylation in humans and in *Arabidopsis*[100–102], although some human lncRNAs have been also identified as regulated by DNA methylation at their promoter regions[32,33,103].

5-AzaC also affects *S. mansoni* females' stem cells, leading to a 95% reduction in the number of proliferating stem cells[16]. Remarkably, lncRNAs actively participate in human stem cell pluripotency, maintenance and differentiation[104]. Thus, it is possible that some of the lncRNAs found here as modulated by 5-AzaC play important roles in parasite stem cells. Further analyses of lncRNA expression in spatially-distinct *S. mansoni* female stem cell populations under 5-AzaC exposure, including vitelline S1 stem cells which are vital for egg production, may uncover lncRNA functional roles on stem cell maintenance. Alternatively, 5-AzaC may modulate lncRNA expression levels by exerting pleiotropic effects similar to those reported in human cell lines such as suppressing lipid metabolism[105], inhibition of pathways that regulate DNA synthesis/repair[106] or de-repression of retroviral expression[107]. Future studies aiming to elucidate the precise mechanism of action of 5-AzaC in lncRNA regulation[108,109] could offer starting points for lncRNA targeting and manipulation in *S. mansoni*.

In summary, this study adds another layer on the understanding of the effects of 5-AzaC in *S. mansoni* and sheds light on the relevance of looking at lncRNA regulation in response to drug treatment in parasites. Although the use of 5-AzaC against schistosomiasis is unlikely as its effects are not parasite selective, the lncRNAs affected by 5-AzaC identified here, together with downstream pathways already described as affected by 5-AzaC, could represent new targets for the development of alternative chemotherapeutic strategies against schistosomiasis.

Material and methods

Analysis of 5-AzaC RNA-Seq data. Public RNA-Seq data from Geyer et al.[16] for *S. mansoni* females were downloaded from the SRA-NCBI database (project number PRJNA428470; controls #SRR6490481, #SRR6490482 and #SRR6490483; treated with 491 µM 5-AzaC #SRR6490480, #SRR6490484 and #SRR6490485). Adapters and bad quality reads were filtered out using fastp v. 0.19.5 with default parameters[110]. For transcripts expression quantitation the genome sequence v.7, and a GTF file containing the protein-coding transcriptome v 7.1 were downloaded from the WormBase ParaSite resource (version WBPS14)[111]. The latter was merged with the lncRNA transcriptome sequences identified by Maciel et al.[31] and the resulting GTF, which is available at http://schistosoma.usp.br/, was used as the reference. The filtered RNA-Seq reads were aligned with STAR v 2.7[112] and quantified with RSEM v 1.3.1[113], both using default parameters, and with the RSEM "estimate-rspd parameter on" option. Transcripts with counts lower than 10 were removed and differential expression analysis was performed using DESeq2 package[114] v. 1.24.0 with an FDR threshold of 0.05. The Volcano plot shows the −log10(p value) vs log2(fold-change) for the lncRNAs obtained in the DESeq2 analysis, using EnhancedVolcano (R package version 1.6.0), available at https://github.com/kevinblighe/EnhancedVolcano. To look for the expression patterns of lincRNAs at different *S. mansoni* life-cycle stages and tissues we re-analyzed data from the public RNA-Seq libraries indicated in Supplementary Table S7, using the same pipeline described above. Pairwise differential expression analysis was performed between each two stages and/or tissues using DESeq2 with an FDR threshold of 0.05. Only transcripts that were upregulated in one stage/tissue when compared with all others, were considered as significantly more expressed in that stage/tissue.

PCA plot was obtained after normalization using the vst function followed by the plotPCA function from DESeq2.

Analysis of the features of lncRNAs differentially expressed after 5-AzaC treatment. The lncRNAs differentially expressed in *S. mansoni* females after 5-AzaC treatment were compared with the lists of lncRNAs identified in four different datasets, namely: (1) lncRNAs belonging to one of the 15 weighted gene co-expression network analyses (WGCNA) modules previously published by Maciel et al.[31] to check to which modules the lncRNAs differentially expressed after 5-AzaC treatment belong to. In Maciel et al.[31], 90 libraries from *S. mansoni* miracidia, sporocysts, schistosomula, cercariae, and gonads (testes and ovaries) were analyzed using the unsupervised WGCNA[115] co-expression analysis approach and 15 different WGCNA modules were obtained[31], each of them representing one cluster of highly interconnected genes that are more expressed in a given stage/tissue. That analysis[31] resulted in the identification of two modules representing miracidia/sporocysts (black and purple), two modules representing juveniles (blue and midnight blue), two modules representing adult males (turquoise and yellow) and four modules representing gonads (brown, green, greenyellow and salmon). Regarding the other five modules, each of them represents only one stage/tissue: cyan (ovaries), magenta (schistosomula), pink (adult females), red (testes) and tan (cercariae); (2) lncRNAs differentially expressed between bisex females (paired, bF) and single-sex females (unpaired, sF) and between bisex ovaries (paired, bO) and single-sex ovaries (unpaired, sO) and whole worms, which we determined by a re-analysis of the transcriptomes previously obtained by Lu et al.[35] (see below); (3) lncRNAs previously published by Maciel et al.[31] as having at least one histone mark (H3K4me3 or H3K27me3) at their TSSs, to check for the presence of histone marks at the TSS of lncRNAs differentially expressed after 5-AzaC treatment; (4) the expression patterns of lncRNAs along *S. mansoni* life-cycle stages, previously published by Maciel et al.[31].

Data from Lu et al.[35] were obtained from SRA (Project number PRJEB14695). In that work, Lu et al.[35] performed RNA-Seq in bisex (paired) females (bF), single-sex females (sF), bisex (paired) males (bM), single-sex males (sM), bisex ovaries (bO), single-sex ovaries (sO), bisex testes (bT) and single-sex testes (sT), but only the protein-coding genes were analyzed. Here, a re-analysis of Lu et al.[35] raw data to detect lncRNAs was performed using the same genome, annotation files and bioinformatics tools and parameters that were used to analyze the data from Geyer et al.[16], as described above. The Venn diagram tool at http://jvenn.toulouse.inra.fr/app/index.html was used to compare the lists of lncRNAs detected as differentially expressed in the present study and in the gonad-specific and pairing-dependent study[35].

Parasite materials. All parasite materials were from a BH isolate of *S. mansoni* maintained by passage through golden hamster (*Mesocricetus auratus*) and *Biomphalaria glabrata* snails. Cercariae were collected from snails infected with 10 miracidia each. Thirty-five days after infection, the snails were placed in the dark in water and then illuminated for 2 h to induce shedding. Schistosomula were obtained by mechanical transformation of cercariae and separation of their bodies as previously described[116], with some modifications. Briefly, cercariae were collected as described above and then suspended in 15 ml of M169 medium (Vitrocell, cat number 00464) containing penicillin/streptomycin, amphotericin (Vitrocell, cat number 00148). Mechanical transformation was performed by passing the cercariae 10 times through a 23G needle. To separate schistosomula from the tails, the tail-rich supernatant was decanted and the sedimented bodies resuspended in a further 7 ml of M169 medium. The procedure was repeated until less than 1% of the tails remained. The newly transformed schistosomula were maintained for 72 h in M169 medium (Vitrocell, cat number 00464) supplemented with penicillin/streptomycin, amphotericin, gentamicin (Vitrocell, cat number 00148), 2% fetal bovine serum, 1 µM serotonin, 0.5 µM hypoxanthine, 1 µM hydrocortisone, and 0.2 µM triiodothyronine at 37 °C and 5% CO_2. Schistosomula cultivated for 72 h were used for 5-AzaC exposure. Adult *S. mansoni* worms were recovered by perfusion of golden hamsters that had been infected 7 weeks before with 250 cercariae, as previously described[116,117]. After perfusion, the adult worm pairs were kept for 3 h at 37 °C and 5% CO_2 in DMEM (Gibco, catalogue number 11995-065-500) supplemented with 10% fetal bovine serum (FBS) (Vitrocell) and 100 mg/ml penicillin/streptomycin (Vitrocell). After 3 h of incubation, the adult worm pairs were used for 5-AzaC treatment.

Parasite treatment with 5-AzaC. *Schistosoma mansoni* schistosomula and adult worms were treated with different final concentrations of 5-AzaC (Sigma, A2385) in culture medium specific to each stage: adult worms were treated with 5-AzaC at 491 µM (same treatment as in Geyer et al.[16]) and schistosomula were treated with 5-AzaC from 245 to 7.7 µM, as indicated in the Results. Adult male and female schistosome couples were cultivated in the presence (or absence) of 5-AzaC according to the methodology described in Geyer et al.[15]. 5-AzaC was added to 30 worm pairs for each of five biological replicates, while additional five replicates, lacking 5-AzaC, were included as controls. The schistosome cultures were incubated at 37 °C for 48 h in a humidified atmosphere with a 70% media exchange performed after 24 h. After 48 h, eggs were counted and schistosome worms were collected, washed three times with PBS and stored in RNA*later* (Ambion) until RNA extraction. Before the extraction of RNA from males or females, adult worm pairs were manually separated in RNA*later* (Ambion) using tweezers. Adult worm couple viability was evaluated 2 days post culture initiation using 9 worm pairs that were cultivated in 5 ml medium in 6 well tissue culture plates (n = 5 biological replicates; 5-AzaC at 491 µM treated or control worms).

Newly transformed schistosomula (NTS) were maintained in culture[34] for 72 h and then treatment with 5-AzaC was initiated. Schistosomula viability was measured after 24, 48, 72, 96 and 120 h of treatment (n = 2 biological replicates).

Viability assay. The viability of *S. mansoni* schistosomula and adult worms after treatment with 5-AzaC was determined by a cytotoxicity assay based on the CellTiter-Glo Luminescent Cell Viability Assay (G7570, Promega)[34,118]. The assay determines the amount of ATP present in freshly lysed adults or in intact schistosomula; the assay signals the presence of metabolically active cells.

RNA extraction, quantification, and quality assessment. RNA extraction, quantification, and quality assessment were performed according to Maciel et al.[31]. Male or female adult worms were first disrupted in Qiagen RLT buffer using glass potters and pestles. RNAs from males or females were then extracted and purified using the Qiagen RNeasy Mini Kit (Cat number 74104), according to the manufacturer's instructions, except for the DNase I treatment: the amount of DNase I was doubled, and the time of treatment was increased to 45 min.

The integrity of all RNAs was verified using the Agilent RNA 6000 Pico Kit (5067-1513 Agilent Technologies) in a 2100 Bioanalyzer Instrument (Agilent Technologies) and quantified using the Qubit RNA HS Assay Kit (Q32852, Thermo Fisher Scientifc). Purity was assessed by 260/280 nm and 260/230 nm ratios using Nanodrop (Thermo Fisher Scientific). Five biological replicates were assessed for 5-AzaC treated or control males or females.

Reverse transcription and quantitative PCR (RT-qPCR) assays. The reverse transcription (RT) reactions were performed with 200 ng total RNA of each control and 5-AzaC treated female samples and with 30 ng total RNA of each control and 5-AzaC treated male samples. For the RT reactions, the SuperScript IV FirstStrand Synthesis System (18091050; Life Technologies) and random hexamer primers were used in a 20 µl final volume. The obtained complementary DNAs (cDNAs) were diluted four times in DEPC water, and quantitative PCR was performed using 2.5 µl of each diluted cDNA in a total volume of 10 µl containing 1 × LightCycler 480 SYBR Green I Master Mix (04707516001, Roche Diagnostics) and 800 nM of each primer in a LightCycler 480 System (Roche Diagnostics). Primers for selected transcripts (Supplementary Table S8) were designed using the Primer 3 online tool, and each RT-qPCR was run in three technical replicates. The results were analyzed by comparative Ct method[119]. Ct values are shown in Supplementary Table S9. Real-time qPCR data were normalized in relation to the level of expression of two reference genes previously used in the literature, namely Smp_900000[120–122] and Smp_123610[117].

Statistical analyses. Two-tailed unpaired t test was used for pairwise comparisons, and GraphPad Prism software was used to perform the analyses (version 8.0). Hypergeometric test was used for enrichment calcula-

tions, using the online https://stattrek.com/online-calculator/hypergeometric.aspx tool. Quantification of data are represented as mean ± SEM and p value thresholds were * < 0.05, ** < 0.01, *** < 0.001 and **** < 0.0001.

Ethics statement. The experimental protocols were in accordance with the Ethical Principles in Animal Research adopted by the Brazilian College of Animal Experimentation (COBEA) and the protocol/experiments have been approved by the Ethics Committee for Animal Experimentation of Instituto Butantan (CEUAIB n° 1777050816).

Data availability

All data generated or analyzed during this study are included in this published article (and its Supplementary Information files).

Received: 27 July 2020; Accepted: 27 November 2020
Published online: 09 December 2020

References

1. WHO. Global Health Estimates 2016: Disease Burden by Cause, Age, Sex, by Country and by Region, 2000–2016. Geneva, World Health Organization.2016. https://www.who.int/healthinfo/global_burden_disease/estimates/en/index1.html Accessed Jan 2020. (2016).
2. Mutapi, F., Maizels, R., Fenwick, A. & Woolhouse, M. Human schistosomiasis in the post mass drug administration era. *Lancet Infect. Dis.* **17**, e42–e48. https://doi.org/10.1016/S1473-3099(16)30475-3 (2017).
3. Wilson, R. A. Schistosomiasis then and now: What has changed in the last 100 years?. *Parasitology* **147**, 507–515. https://doi.org/10.1017/S0031182020000049 (2020).
4. Colley, D. G., Bustinduy, A. L., Secor, W. E. & King, C. H. Human schistosomiasis. *Lancet* **383**, 2253–2264. https://doi.org/10.1016/S0140-6736(13)61949-2 (2014).
5. Zoni, A. C., Catala, L. & Ault, S. K. Schistosomiasis prevalence and intensity of infection in Latin America and the Caribbean Countries, 1942–2014: A systematic review in the context of a regional elimination goal. *PLoS Negl. Trop. Dis.* **10**, e0004493. https://doi.org/10.1371/journal.pntd.0004493 (2016).
6. Bergquist, R., Utzinger, J. & Keiser, J. Controlling schistosomiasis with praziquantel: How much longer without a viable alternative?. *Infect. Dis. Poverty* **6**, 74. https://doi.org/10.1186/s40249-017-0286-2 (2017).
7. Gryseels, B. *et al.* Epidemiology, immunology and chemotherapy of *Schistosoma mansoni* infections in a recently exposed community in Senegal. *Trop. Geogr. Med.* **46**, 209–219 (1994).
8. Melman, S. D. *et al.* Reduced susceptibility to praziquantel among naturally occurring Kenyan isolates of *Schistosoma mansoni*. *PLoS Negl. Trop. Dis.* **3**, e504. https://doi.org/10.1371/journal.pntd.0000504 (2009).
9. Vale, N. *et al.* Praziquantel for schistosomiasis: Single-drug metabolism revisited, mode of action, and resistance. *Antimicrob. Agents Chemother.* **61**, e02582-16. https://doi.org/10.1128/AAC.02582-16 (2017).
10. Hewitson, J. P. & Maizels, R. M. Vaccination against helminth parasite infections. *Expert Rev. Vaccines* **13**, 473–487. https://doi.org/10.1586/14760584.2014.893195 (2014).
11. Gouveia, M. J., Brindley, P. J., Gartner, F., Costa, J. & Vale, N. Drug repurposing for schistosomiasis: Combinations of drugs or biomolecules. *Pharmaceuticals (Basel)* **11**, 15. https://doi.org/10.3390/ph11010015 (2018).
12. Cataldo, V. D., Cortes, J. & Quintas-Cardama, A. Azacitidine for the treatment of myelodysplastic syndrome. *Expert Rev. Anticancer Ther.* **9**, 875–884. https://doi.org/10.1586/era.09.61 (2009).
13. Lu, L. J. & Randerath, K. Mechanism of 5-azacytidine-induced transfer RNA cytosine-5-methyltransferase deficiency. *Cancer Res.* **40**, 2701–2705 (1980).
14. Reichman, M. & Penman, S. The mechanism of inhibition of protein synthesis by 5-azacytidine in HeLa cells. *Biochim. Biophys. Acta* **324**, 282–289. https://doi.org/10.1016/0005-2787(73)90145-7 (1973).
15. Geyer, K. K. *et al.* Cytosine methylation regulates oviposition in the pathogenic blood fluke *Schistosoma mansoni*. *Nat. Commun.* **2**, 424. https://doi.org/10.1038/ncomms1433 (2011).
16. Geyer, K. K. *et al.* The anti-fecundity effect of 5-azacytidine (5-AzaC) on *Schistosoma mansoni* is linked to dis-regulated transcription, translation and stem cell activities. *Int. J. Parasitol .Drugs Drug Resist.* **8**, 213–222. https://doi.org/10.1016/j.ijpddr.2018.03.006 (2018).
17. Quinn, J. J. & Chang, H. Y. Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.* **17**, 47–62. https://doi.org/10.1038/nrg.2015.10 (2016).
18. Kopp, F. & Mendell, J. T. Functional classification and experimental dissection of long noncoding RNAs. *Cell* **172**, 393–407. https://doi.org/10.1016/j.cell.2018.01.011 (2018).
19. Ransohoff, J. D., Wei, Y. & Khavari, P. A. The functions and unique features of long intergenic non-coding RNA. *Nat. Rev. Mol. Cell Biol.* **19**, 143–157. https://doi.org/10.1038/nrm.2017.104 (2018).
20. Gil, N. & Ulitsky, I. Regulation of gene expression by cis-acting long non-coding RNAs. *Nat. Rev. Genet.* **21**, 102–117. https://doi.org/10.1038/s41576-019-0184-5 (2020).
21. Matsui, M. & Corey, D. R. Non-coding RNAs as drug targets. *Nat. Rev. Drug Discov.* **16**, 167–179. https://doi.org/10.1038/nrd.2016.117 (2017).
22. Santos, L. N. *et al.* De novo assembly and characterization of the *Trichuris trichiura* adult worm transcriptome using Ion Torrent sequencing. *Acta Trop.* **159**, 132–141. https://doi.org/10.1016/j.actatropica.2016.03.036 (2016).
23. Azlan, A., Halim, M. A. & Azzam, G. Genome-wide identification and characterization of long intergenic noncoding RNAs in the regenerative flatworm *Macrostomum lignano*. *Genomics* **112**, 1273–1281. https://doi.org/10.1016/j.ygeno.2019.07.016 (2020).
24. Ross, E., Blair, D., Guerrero-Hernandez, C. & SanchezAlvarado, A. Comparative and transcriptome analyses uncover key aspects of coding- and long noncoding RNAs in flatworm mitochondrial genomes. *G3 (Bethesda)* **6**, 1191–1200. https://doi.org/10.1534/g3.116.028175 (2016).
25. Oliveira, K. C., Carvalho, M. L., Maracaja-Coutinho, V., Kitajima, J. P. & Verjovski-Almeida, S. Non-coding RNAs in schistosomes: An unexplored world. *An. Acad. Bras. Cie.* **83**, 673–694. https://doi.org/10.1590/s0001-37652011000200026 (2011).
26. Vasconcelos, E. J. R. *et al.* The *Schistosoma mansoni* genome encodes thousands of long non-coding RNAs predicted to be functional at different parasite life-cycle stages. *Sci. Rep.* **7**, 10508. https://doi.org/10.1038/s41598-017-10853-6 (2017).
27. Liao, Q. *et al.* Identification of long noncoding RNAs in *Schistosoma mansoni* and *Schistosoma japonicum*. *Exp. Parasitol.* **191**, 82–87. https://doi.org/10.1016/j.exppara.2018.07.001 (2018).
28. Oliveira, V. F. *et al.* Identification of 170 new long noncoding RNAs in *Schistosoma mansoni*. *Biomed. Res. Int.* **2018**, 1264697. https://doi.org/10.1155/2018/1264697 (2018).

29. Vasconcelos, E. J. R. *et al.* Atlas of *Schistosoma mansoni* long non-coding RNAs and their expression correlation to protein-coding genes. *Database (Oxford)* **2018**, bay068. https://doi.org/10.1093/database/bay068 (2018).
30. Kim, H. C., Khalil, A. M. & Jolly, E. R. LncRNAs in molluscan and mammalian stages of parasitic schistosomes are developmentally-regulated and coordinately expressed with protein-coding genes. *RNA Biol.* **17**, 805–815. https://doi.org/10.1080/15476286.2020.1729594 (2020).
31. Maciel, L. F. *et al.* Weighted gene co-expression analyses point to long non-coding RNA hub genes at different *Schistosoma mansoni* life-cycle stages. *Front Genet.* **10**, 823. https://doi.org/10.3389/fgene.2019.00823 (2019).
32. Kumegawa, K. *et al.* A genomic screen for long noncoding RNA genes epigenetically silenced by aberrant DNA methylation in colorectal cancer. *Sci. Rep.* **6**, 26699. https://doi.org/10.1038/srep26699 (2016).
33. Diaz-Lagares, A. *et al.* Epigenetic inactivation of the p53-induced long noncoding RNA TP53 target 1 in human cancer. *Proc. Natl. Acad. Sci. USA* **113**, E7535–E7544. https://doi.org/10.1073/pnas.1608585113 (2016).
34. Pereira, A. S. A. *et al.* Inhibition of histone methyltransferase EZH2 in *Schistosoma mansoni* in vitro by GSK343 reduces egg laying and decreases the expression of genes implicated in DNA replication and noncoding RNA metabolism. *PLoS. Negl. Trop. Dis.* **12**, e0006873. https://doi.org/10.1371/journal.pntd.0006873 (2018).
35. Lu, Z. *et al.* Schistosome sex matters: A deep view into gonad-specific and pairing-dependent transcriptomes reveals a complex gender interplay. *Sci. Rep.* **6**, 31150. https://doi.org/10.1038/srep31150 (2016).
36. Du, J., Johnson, L. M., Jacobsen, S. E. & Patel, D. J. DNA methylation pathways and their crosstalk with histone methylation. *Nat. Rev. Mol. Cell Biol.* **16**, 519–532. https://doi.org/10.1038/nrm4043 (2015).
37. Jeltsch, A., Broche, J. & Bashtrykov, P. Molecular processes connecting DNA methylation patterns with DNA methyltransferases and histone modifications in mammalian genomes. *Genes (Basel)* **9**, 566. https://doi.org/10.3390/genes9110566 (2018).
38. Parker-Manuel, S. J., Ivens, A. C., Dillon, G. P. & Wilson, R. A. Gene expression patterns in larval *Schistosoma mansoni* associated with infection of the mammalian host. *PLoS Negl. Trop. Dis.* **5**, e1274. https://doi.org/10.1371/journal.pntd.0001274 (2011).
39. Anderson, L. *et al. Schistosoma mansoni* egg, adult male and female comparative gene expression analysis and identification of novel genes by RNA-Seq. *PLoS Negl. Trop. Dis.* **9**, e0004334. https://doi.org/10.1371/journal.pntd.0004334 (2015).
40. Pan, D., Rampal, R. & Mascarenhas, J. Clinical developments in epigenetic-directed therapies in acute myeloid leukemia. *Blood Adv.* **4**, 970–982. https://doi.org/10.1182/bloodadvances.2019001245 (2020).
41. Shao, C. C. *et al.* Comparative analysis of microRNA profiles between adult *Ascaris lumbricoides* and *Ascaris suum*. *BMC Vet. Res.* **10**, 99. https://doi.org/10.1186/1746-6148-10-99 (2014).
42. Fontenla, S., Rinaldi, G., Smircich, P. & Tort, J. F. Conservation and diversification of small RNA pathways within flatworms. *BMC Evol. Biol.* **17**, 215. https://doi.org/10.1186/s12862-017-1061-5 (2017).
43. Macchiaroli, N. *et al.* Identification and expression profiling of microRNAs in Hymenolepis. *Int. J. Parasitol.* **49**, 211–223. https://doi.org/10.1016/j.ijpara.2018.07.005 (2019).
44. Holz, A. & Streit, A. Gain and loss of small RNA classes-characterization of small RNAs in the parasitic nematode family strongyloididae. *Genome Biol. Evol.* **9**, 2826–2843. https://doi.org/10.1093/gbe/evx197 (2017).
45. Rodelsperger, C., Menden, K., Serobyan, V., Witte, H. & Baskaran, P. First insights into the nature and evolution of antisense transcription in nematodes. *BMC Evol. Biol.* **16**, 165. https://doi.org/10.1186/s12862-016-0740-y (2016).
46. Wei, S. *et al.* Systematic evaluation of *C. elegans* lincRNAs with CRISPR knockout mutants. *Genome Biol.* **20**, 7. https://doi.org/10.1186/s13059-018-1619-6 (2019).
47. Zheng, Y., Cai, X. & Bradley, J. E. microRNAs in parasites and parasite infection. *RNA Biol.* **10**, 371–379. https://doi.org/10.4161/rna.23716 (2013).
48. Britton, C., Winter, A. D., Gillan, V. & Devaney, E. microRNAs of parasitic helminths—identification, characterization and potential as drug targets. *Int. J. Parasitol. Drugs Drug Resist.* **4**, 85–94. https://doi.org/10.1016/j.ijpddr.2014.03.001 (2014).
49. Tritten, L. *et al.* Detection of circulating parasite-derived microRNAs in filarial infections. *PLoS Negl. Trop. Dis.* **8**, e2971. https://doi.org/10.1371/journal.pntd.0002971 (2014).
50. Marks, N. D. *et al.* Profiling microRNAs through development of the parasitic nematode Haemonchus identifies nematode-specific miRNAs that suppress larval development. *Sci. Rep.* **9**, 17594. https://doi.org/10.1038/s41598-019-54154-6 (2019).
51. Meningher, T. *et al.* Schistosomal extracellular vesicle-enclosed miRNAs modulate host T helper cell differentiation. *EMBO Rep.* **21**, e47882. https://doi.org/10.15252/embr.201947882 (2020).
52. Liu, J. *et al. Schistosoma japonicum* extracellular vesicle miRNA cargo regulates host macrophage functions facilitating parasitism. *PLoS Pathog.* **15**, e1007817. https://doi.org/10.1371/journal.ppat.1007817 (2019).
53. Wang, Y. *et al.* Systematic identification of non-coding pharmacogenomic landscape in cancer. *Nat. Commun.* **9**, 3192. https://doi.org/10.1038/s41467-018-05495-9 (2018).
54. Prabhakar, B., Zhong, X. B. & Rasmussen, T. P. Exploiting long noncoding RNAs as pharmacological targets to modulate epigenetic diseases. *Yale J. Biol. Med.* **90**, 73–86 (2017).
55. Blokhin, I., Khorkova, O., Hsiao, J. & Wahlestedt, C. Developments in lncRNA drug discovery: Where are we heading?. *Expert Opin. Drug Discov.* **13**, 837–849. https://doi.org/10.1080/17460441.2018.1501024 (2018).
56. Geary, T. G., Sakanari, J. A. & Caffrey, C. R. Anthelmintic drug discovery: Into the future. *J. Parasitol.* **101**, 125–133. https://doi.org/10.1645/14-703.1 (2015).
57. Fairweather, I., Brennan, G. P., Hanna, R. E. B., Robinson, M. W. & Skuce, P. J. Drug resistance in liver flukes. *Int. J. Parasitol. Drugs Drug Resist.* **12**, 39–59. https://doi.org/10.1016/j.ijpddr.2019.11.003 (2020).
58. Partridge, F. A. *et al.* Anthelmintic drug discovery: Target identification, screening methods and the role of open science. *Beilstein J. Org. Chem.* **16**, 1203–1224. https://doi.org/10.3762/bjoc.16.105 (2020).
59. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927. https://doi.org/10.1101/gad.17446611 (2011).
60. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789. https://doi.org/10.1101/gr.132159.111 (2012).
61. Warner, K. D., Hajdin, C. E. & Weeks, K. M. Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug Discov.* **17**, 547–558. https://doi.org/10.1038/nrd.2018.93 (2018).
62. Liu, S. J. & Lim, D. A. Modulating the expression of long non-coding RNAs for functional studies. *EMBO Rep.* **19**, e46955. https://doi.org/10.15252/embr.201846955 (2018).
63. Arun, G., Diermeier, S. D. & Spector, D. L. Therapeutic targeting of long non-coding RNAs in cancer. *Trends Mol. Med.* **24**, 257–277. https://doi.org/10.1016/j.molmed.2018.01.001 (2018).
64. Greenberg, R. M. New approaches for understanding mechanisms of drug resistance in schistosomes. *Parasitology* **140**, 1534–1546. https://doi.org/10.1017/S0031182013000231 (2013).
65. Dai, E. *et al.* ncDR: A comprehensive resource of non-coding RNAs involved in drug resistance. *Bioinformatics* **33**, 4010–4011. https://doi.org/10.1093/bioinformatics/btx523 (2017).
66. Merry, C. R. *et al.* Transcriptome-wide identification of mRNAs and lincRNAs associated with trastuzumab-resistance in HER2-positive breast cancer. *Oncotarget* **7**, 53230–53244. https://doi.org/10.18632/oncotarget.10637 (2016).
67. Wang, W. T., Han, C., Sun, Y. M., Chen, T. Q. & Chen, Y. Q. Noncoding RNAs in cancer therapy resistance and targeted drug development. *J. Hematol. Oncol.* **12**, 55. https://doi.org/10.1186/s13045-019-0748-z (2019).
68. Levy, M. G. & Read, C. P. Purine and pyrimidine transport in *Schistosoma mansoni*. *J. Parasitol.* **61**, 627–632 (1975).

69. Xue, X. *et al.* LncRNA HOTAIR enhances ER signaling and confers tamoxifen resistance in breast cancer. *Oncogene* **35**, 2746–2755. https://doi.org/10.1038/onc.2015.340 (2016).
70. Wang, H. *et al.* The role of long noncoding RNA HOTAIR in the acquired multidrug resistance to imatinib in chronic myeloid leukemia cells. *Hematology* **22**, 208–216. https://doi.org/10.1080/10245332.2016.1258152 (2017).
71. Zhu, J. *et al.* Knockdown of long non-coding RNA XIST inhibited doxorubicin resistance in colorectal cancer by upregulation of miR-124 and downregulation of SGK1. *Cell Physiol. Biochem.* **51**, 113–128. https://doi.org/10.1159/000495168 (2018).
72. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227. https://doi.org/10.1038/nature07672 (2009).
73. Wu, S. C., Kallin, E. M. & Zhang, Y. Role of H3K27 methylation in the regulation of lncRNA expression. *Cell Res.* **20**, 1109–1116. https://doi.org/10.1038/cr.2010.114 (2010).
74. Amin, V. *et al.* Epigenomic footprints across 111 reference epigenomes reveal tissue-specific epigenetic regulation of lincRNAs. *Nat. Commun.* **6**, 6370. https://doi.org/10.1038/ncomms7370 (2015).
75. Li, Z. *et al.* DNA methylation and gene expression profiles characterize epigenetic regulation of lncRNAs in colon adenocarcinoma. *J. Cell Biochem.* **121**, 2406–2415. https://doi.org/10.1002/jcb.29463 (2020).
76. Zhang, S. *et al.* Cancer-associated methylated lncRNAs in patients with bladder cancer. *Am. J. Transl. Res.* **11**, 3790–3800 (2019).
77. Jacob, R., Zander, S. & Gutschner, T. The dark side of the epitranscriptome: Chemical modifications in long non-coding RNAs. *Int. J. Mol. Sci.* **18**, 2387. https://doi.org/10.3390/ijms18112387 (2017).
78. Trixl, L. & Lusser, A. The dynamic RNA modification 5-methylcytosine and its emerging role as an epitranscriptomic mark. *Wiley Interdiscip. Rev. RNA* **10**, e1510. https://doi.org/10.1002/wrna.1510 (2019).
79. Wu, Z. *et al.* Regulation of lncRNA expression. *Cell Mol. Biol. Lett.* **19**, 561–575. https://doi.org/10.2478/s11658-014-0212-6 (2014).
80. Bunch, H. Gene regulation of mammalian long non-coding RNA. *Mol. Genet. Genom.* **293**, 1–15. https://doi.org/10.1007/s00438-017-1370-9 (2018).
81. Geyer, K. K. & Hoffmann, K. F. Epigenetics: A key regulator of platyhelminth developmental biology?. *Int. J Parasitol.* **42**, 221–224. https://doi.org/10.1016/j.ijpara.2012.02.003 (2012).
82. Cosseau, C. *et al.* (Epi)genetic inheritance in *Schistosoma mansoni*: A systems approach. *Trends Parasitol.* **33**, 285–294. https://doi.org/10.1016/j.pt.2016.12.002 (2017).
83. Fneich, S. *et al.* Epigenetic origin of adaptive phenotypic variants in the human blood fluke *Schistosoma mansoni*. *Epigenet. Chromatin* **9**, 27. https://doi.org/10.1186/s13072-016-0076-2 (2016).
84. Roquis, D. *et al.* Histone methylation changes are required for life cycle progression in the human parasite *Schistosoma mansoni*. *PLoS Pathog.* **14**, e1007066. https://doi.org/10.1371/journal.ppat.1007066 (2018).
85. Augusto, R. C., Duval, D. & Grunau, C. Effects of the environment on developmental plasticity and infection success of Schistosoma parasites—an epigenetic perspective. *Front Microbiol.* **10**, 1475. https://doi.org/10.3389/fmicb.2019.01475 (2019).
86. Carneiro, V. C. *et al.* Epigenetic changes modulate schistosome egg formation and are a novel target for reducing transmission of schistosomiasis. *PLoS Pathog.* **10**, e1004116. https://doi.org/10.1371/journal.ppat.1004116 (2014).
87. Padalino, G., Ferla, S., Brancale, A., Chalmers, I. W. & Hoffmann, K. F. Combining bioinformatics, cheminformatics, functional genomics and whole organism approaches for identifying epigenetic drug targets in *Schistosoma mansoni*. *Int. J. Parasitol. Drugs Drug Resist.* **8**, 559–570. https://doi.org/10.1016/j.ijpddr.2018.10.005 (2018).
88. Anderson, L. *et al.* Histone deacetylase inhibition modulates histone acetylation at gene promoter regions and affects genome-wide gene transcription in *Schistosoma mansoni*. *PLoS Negl. Trop. Dis.* **11**, e0005539. https://doi.org/10.1371/journal.pntd.0005539 (2017).
89. Dubois, F. *et al.* Histone deacetylase inhibitors induce apoptosis, histone hyperacetylation and up-regulation of gene transcription in *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* **168**, 7–15. https://doi.org/10.1016/j.molbiopara.2009.06.001 (2009).
90. Heimburg, T. *et al.* Structure-based design and synthesis of novel inhibitors targeting HDAC8 from *Schistosoma mansoni* for the treatment of schistosomiasis. *J. Med. Chem.* **59**, 2423–2435. https://doi.org/10.1021/acs.jmedchem.5b01478 (2016).
91. Lobo-Silva, J. *et al.* The antischistosomal potential of GSK-J4, an H3K27 demethylase inhibitor: Insights from molecular modeling, transcriptomics and in vitro assays. *Parasit. Vectors* **13**, 140. https://doi.org/10.1186/s13071-020-4000-z (2020).
92. Whatley, K. C. L. *et al.* The repositioning of epigenetic probes/inhibitors identifies new anti-schistosomal lead compounds and chemotherapeutic targets. *PLoS Negl. Trop. Dis.* **13**, e0007693. https://doi.org/10.1371/journal.pntd.0007693 (2019).
93. Raddatz, G. *et al.* Dnmt2-dependent methylomes lack defined DNA methylation patterns. *Proc. Natl. Acad. Sci. USA* **110**, 8627–8631. https://doi.org/10.1073/pnas.1306723110 (2013).
94. Aliaga, B., Bulla, I., Mouahid, G., Duval, D. & Grunau, C. Universality of the DNA methylation codes in Eucaryotes. *Sci. Rep.* **9**, 173. https://doi.org/10.1038/s41598-018-37407-8 (2019).
95. Fantappie, M. R., Gimba, E. R. & Rumjanek, F. D. Lack of DNA methylation in *Schistosoma mansoni*. *Exp. Parasitol.* **98**, 162–166. https://doi.org/10.1006/expr.2001.4630 (2001).
96. Salzberg, A., Fisher, O., Siman-Tov, R. & Ankri, S. Identification of methylated sequences in genomic DNA of adult *Drosophila melanogaster*. *Biochem. Biophys. Res. Commun.* **322**, 465–469. https://doi.org/10.1016/j.bbrc.2004.07.134 (2004).
97. del Gaudio, R., Di Giaimo, R. & Geraci, G. Genome methylation of the marine annelid worm *Chaetopterus variopedatus*: Methylation of a CpG in an expressed H1 histone gene. *FEBS Lett.* **417**, 48–52. https://doi.org/10.1016/s0014-5793(97)01262-3 (1997).
98. Regev, A., Lamb, M. J. & Jablonka, E. The role of DNA methylation in invertebrates: Developmental regulation or genome defense?. *Mol. Biol. Evol.* **15**, 880–891. https://doi.org/10.1093/oxfordjournals.molbev.a025992 (1998).
99. Li, L. H., Olin, E. J., Buskirk, H. H. & Reineke, L. M. Cytotoxicity and mode of action of 5-azacytidine on L1210 leukemia. *Cancer Res.* **30**, 2760–2769 (1970).
100. Amort, T. *et al.* Long non-coding RNAs as targets for cytosine methylation. *RNA Biol.* **10**, 1003–1008. https://doi.org/10.4161/rna.24454 (2013).
101. David, R. *et al.* Transcriptome-wide mapping of RNA 5-methylcytosine in Arabidopsis mRNAs and noncoding RNAs. *Plant Cell* **29**, 445–460. https://doi.org/10.1105/tpc.16.00751 (2017).
102. Squires, J. E. *et al.* Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res* **40**, 5023–5033. https://doi.org/10.1093/nar/gks144 (2012).
103. Zhao, J., Dahle, D., Zhou, Y., Zhang, X. & Klibanski, A. Hypermethylation of the promoter region is associated with the loss of MEG3 gene expression in human pituitary tumors. *J. Clin. Endocrinol. Metab.* **90**, 2179–2186. https://doi.org/10.1210/jc.2004-1848 (2005).
104. Fico, A., Fiorenzano, A., Pascale, E., Patriarca, E. J. & Minchiotti, G. Long non-coding RNA in stem cell pluripotency and lineage commitment: Functions and evolutionary conservation. *Cell Mol. Life. Sci.* **76**, 1459–1471. https://doi.org/10.1007/s00018-018-3000-z (2019).
105. Poirier, S. *et al.* The epigenetic drug 5-azacytidine interferes with cholesterol and lipid metabolism. *J. Biol. Chem.* **289**, 18736–18751. https://doi.org/10.1074/jbc.M114.563650 (2014).
106. Aimiuwu, J. *et al.* RNA-dependent inhibition of ribonucleotide reductase is a major pathway for 5-azacytidine activity in acute myeloid leukemia. *Blood* **119**, 5229–5238. https://doi.org/10.1182/blood-2011-11-382226 (2012).

107. Strick, R., Strissel, P. L., Baylin, S. B. & Chiappinelli, K. B. Unraveling the molecular pathways of DNA-methylation inhibitors: Human endogenous retroviruses induce the innate immune response in tumors. *Oncoimmunology* **5**, e1122160. https://doi.org/10.1080/2162402X.2015.1122160 (2016).
108. Romano, G., Veneziano, D., Nigita, G. & Nana-Sinkam, S. P. RNA methylation in ncRNA: Classes, detection, and molecular associations. *Front Genet.* **9**, 243. https://doi.org/10.3389/fgene.2018.00243 (2018).
109. Khoddami, V. & Cairns, B. R. Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nat. Biotechnol.* **31**, 458–464. https://doi.org/10.1038/nbt.2566 (2013).
110. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890. https://doi.org/10.1093/bioinformatics/bty560 (2018).
111. Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P. & Berriman, M. WormBase ParaSite—a comprehensive resource for helminth genomics. *Mol. Biochem. Parasitol.* **215**, 2–10. https://doi.org/10.1016/j.molbiopara.2016.11.005 (2017).
112. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21. https://doi.org/10.1093/bioinformatics/bts635 (2013).
113. Li, B. & Dewey, C. N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **12**, 323. https://doi.org/10.1186/1471-2105-12-323 (2011).
114. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. https://doi.org/10.1186/s13059-014-0550-8 (2014).
115. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559. https://doi.org/10.1186/1471-2105-9-559 (2008).
116. Basch, P. F. Cultivation of *Schistosoma mansoni* in vitro. I. Establishment of cultures from cercariae and development until pairing. *J. Parasitol.* **67**, 179–185 (1981).
117. Pereira, A. S. A. *et al.* In vitro activity of aryl-thiazole derivatives against *Schistosoma mansoni* schistosomula and adult worms. *PLoS One* **14**, e0225425. https://doi.org/10.1371/journal.pone.0225425 (2019).
118. Panic, G., Flores, D., Ingram-Sieber, K. & Keiser, J. Fluorescence/luminescence-based markers for the assessment of *Schistosoma mansoni* schistosomula drug assays. *Parasit. Vectors* **8**, 624. https://doi.org/10.1186/s13071-015-1233-3 (2015).
119. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* **25**, 402–408. https://doi.org/10.1006/meth.2001.1262 (2001).
120. Avelar, L. *et al.* Smp38 MAP kinase regulation in *Schistosoma mansoni*: Roles in survival, oviposition, and protection against oxidative stress. *Front Immunol.* **10**, 21. https://doi.org/10.3389/fimmu.2019.00021 (2019).
121. Jeremias, W. J. *et al.* Comparative sequence analysis reveals regulation of genes in developing schistosomula of *Schistosoma mansoni* exposed to host portal serum. *PLoS One* **12**, e0178829. https://doi.org/10.1371/journal.pone.0178829 (2017).
122. Marek, M. *et al.* Structural basis for the inhibition of histone deacetylase 8 (HDAC8), a key epigenetic player in the blood fluke *Schistosoma mansoni*. *PLoS Pathog.* **9**, e1003645. https://doi.org/10.1371/journal.ppat.1003645 (2013).

Acknowledgements

We acknowledge Dr. David da Silva Pires and Dr. Ana Carolina Tahira for help in the generation of high-throughput data genome tracks.

Author contributions

M.S.A. and S.V.-A. conceived the project, analyzed the data and wrote the manuscript. M.S.A. carried out all experiments. M.S.A. and L.F.M. performed bioinformatics analysis. M.S.A., G.O.S., G.G.O.O. and J.V.P.L. designed and carried out RT-qPCR experiments. M.S.A. and L.K.I. performed schistosomula in vitro treatment experiments. A.S.A.P. gave support on hamster perfusion, adult worm recovery and separation of adult male from female worms after treatments. P.A.M. and E.N. provided biological material. All authors reviewed the manuscript and approved the submitted version.

Funding

This work was supported by a grant from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (Thematic grant number 2018/23693-5 to SV-A). G.O.S., LFM and A.S.A.P. received fellowships from FAPESP (18/24015-0, 18/19591-2 and 2016/10046-6, respectively); G.G.O.O. received a fellowship from CNPq (116733/2019-5); S.V-A laboratory was also supported by institutional funds from Fundação Butantan and S.V-A received an established investigator fellowship award from CNPq, Brasil.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-020-78669-5.

Correspondence and requests for materials should be addressed to S.V.-A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

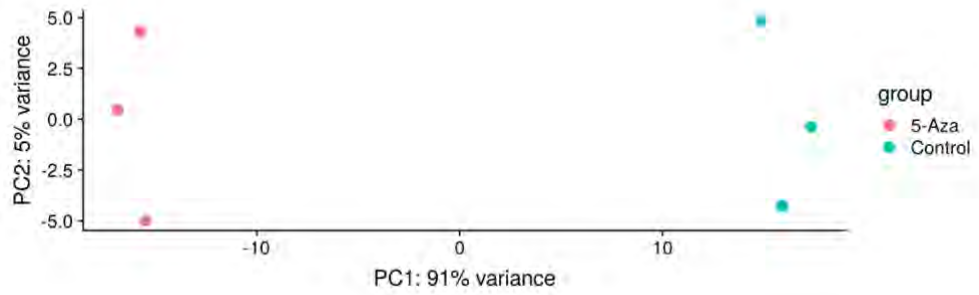
© The Author(s) 2020

Long non-coding RNA levels can be modulated by 5-Azacytidine in *Schistosoma mansoni*

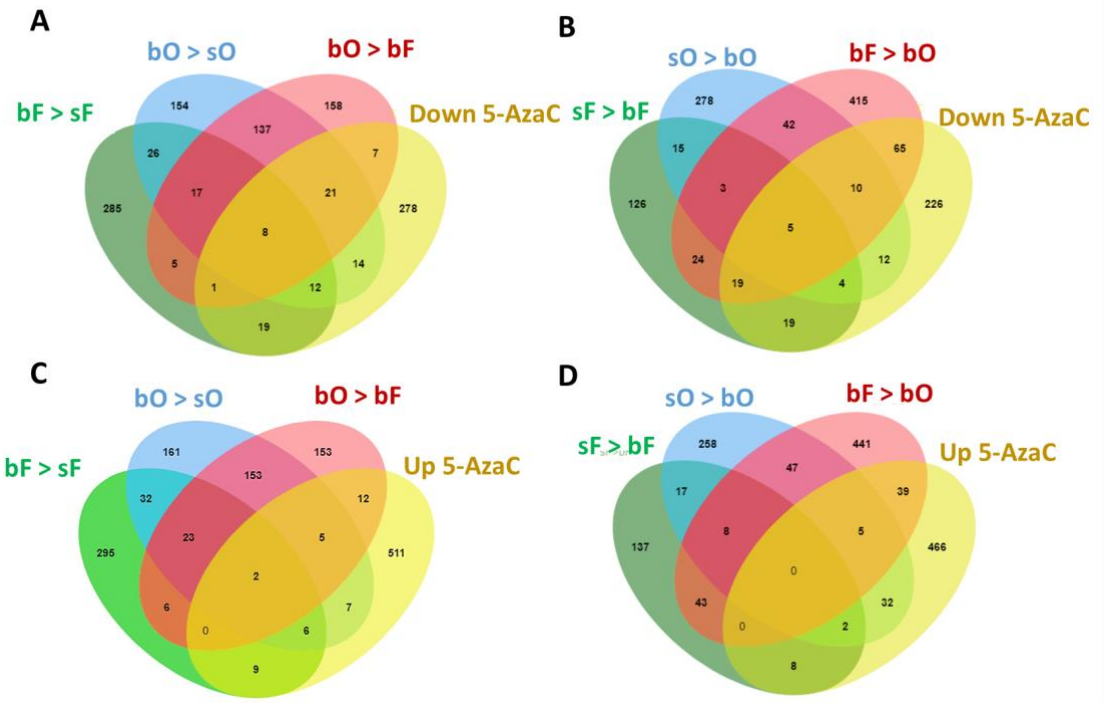
Murilo S. Amaral[1], Lucas F. Maciel[1], Gilbert O. Silveira[1,2], Giovanna G. O. Olberg[1], João V. P. Leite[1]; Lucas K. Imamura[1], Adriana S. A. Pereira[1,2], Patricia A. Miyasato[1], Eliana Nakano[1], Sergio Verjovski-Almeida[1,2*]

[1]Laboratório de Parasitologia, Instituto Butantan, São Paulo, Brazil

[2]Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

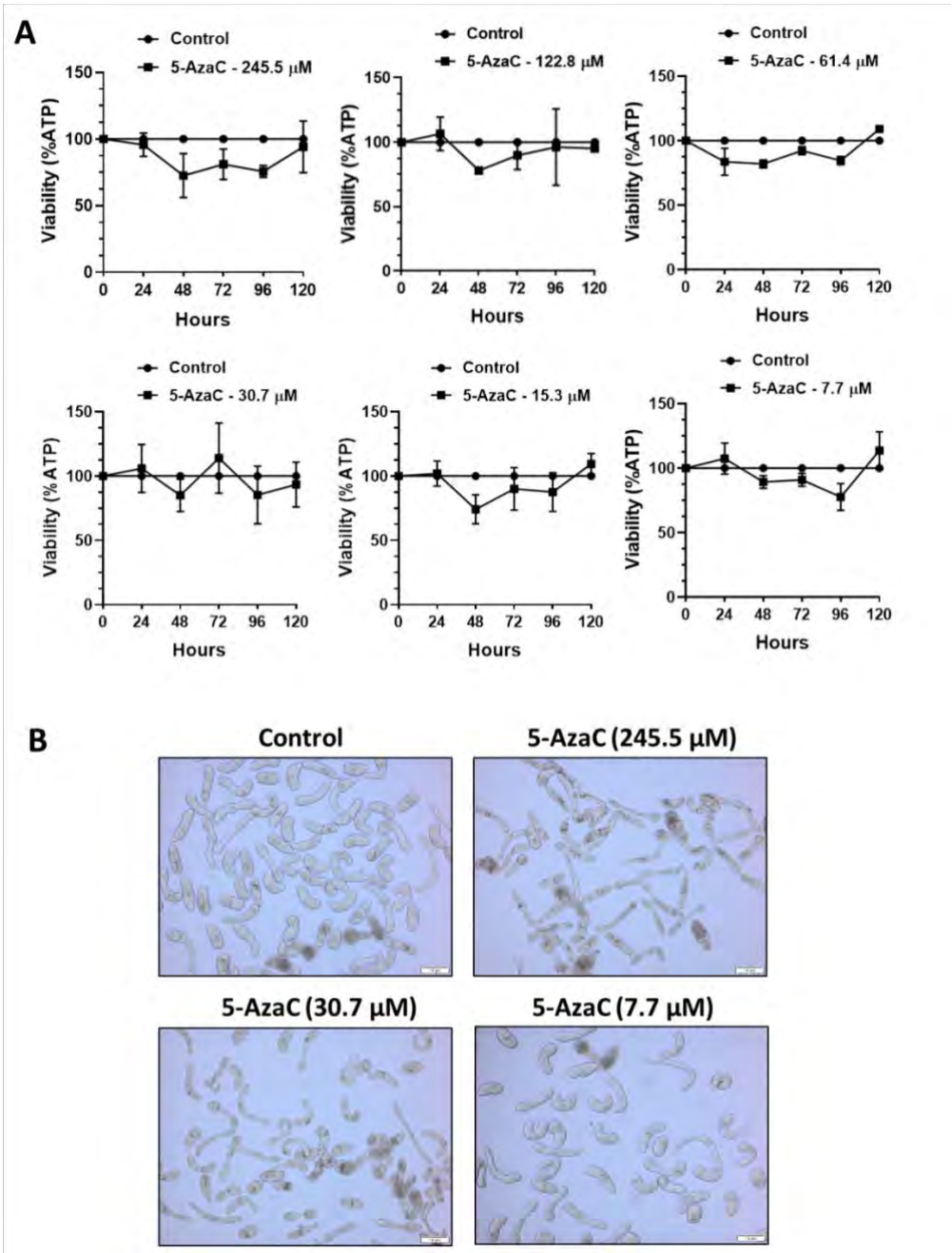


Supplementary Figure S1: Clustering of RNA-Seq biological replicates assessed by principal component analysis (PCA). RNA-Seq data from Geyer *et al*., 2018 [1] were re-analyzed using the *S. mansoni* genome PRJEA36577 (v.7) retrieved from WormBase and the recently published transcriptome that includes long non-coding RNAs [2] as reference. PCA plot was obtained after normalization using the vst function followed by the plotPCA function from DESeq2. Both control and 5-AzaC treated *S. mansoni* female samples are represented by three biological replicates each (n = 3), which are separated by their first two principal components. The control samples are represented by blue dots and the 5-AzaC treated samples by red dots.

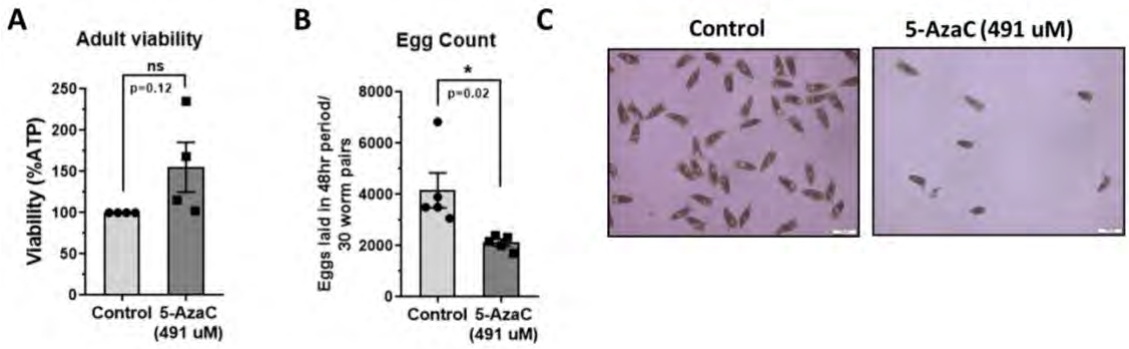


Supplementary Figure S2: Venn diagram representing the number of lncRNAs differentially expressed in different conditions of female pairing compared with females under 5-AzaC exposure. Re-analysis of RNA-seq public data from Lu *et al*., 2016 [3], focusing on lncRNAs mapping and quantification, followed by comparison with lncRNAs differentially expressed in bisex females after 5-AzaC exposure, as determined by re-analysis of data from Geyer *et al*., 2018 [1]. **A.** lncRNAs downregulated in bisex females after 5-AzaC exposure are compared with lncRNAs differentially expressed between the following conditions: lncRNAs enriched in bisex (paired) females compared with single-sex (unpaired) females (bF>sF); lncRNAs enriched in ovaries from bisex (paired) females compared with ovaries from single-sex (unpaired) females (bO>sO); lncRNAs enriched in ovaries from bisex (paired) females compared with bisex (paired) females (bO>bF). **B.** lncRNAs downregulated in bisex females after 5-AzaC exposure are compared with lncRNAs differentially expressed between the following conditions: lncRNAs enriched in single-sex (unpaired) females compared with bisex (paired) females (sF>bF); lncRNAs enriched in ovaries from single-sex (unpaired) females compared with ovaries from bisex (paired) females (sO>bO); lncRNAs enriched in bisex (paired) females compared with ovaries from bisex (paired) females (bF>bO). **C.** lncRNAs upregulated in bisex females after 5-AzaC exposure are compared with lncRNAs differentially expressed between the following conditions: lncRNAs enriched in bisex (paired) females compared with single-sex

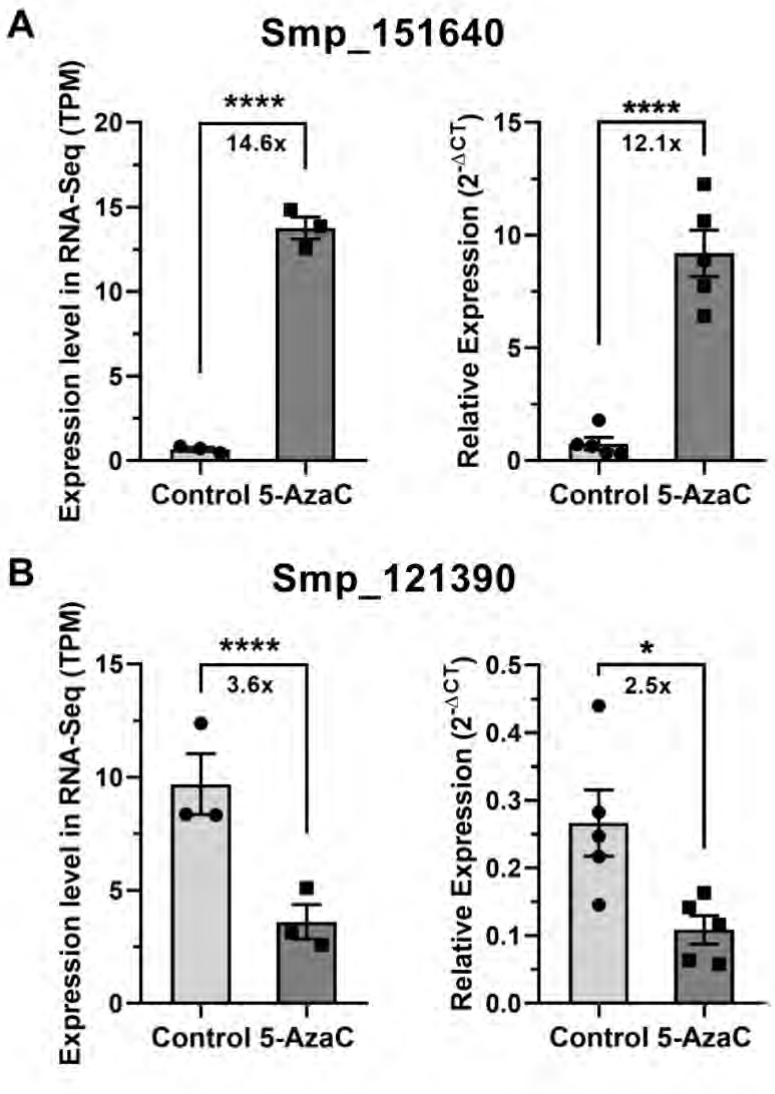
(unpaired) females (bF>sF); lncRNAs enriched in ovaries from bisex (paired) females compared with ovaries from single-sex (unpaired) females (bO>sO); lncRNAs enriched in ovaries from bisex (paired) females compared with bisex (paired) females (bO>bF). **D.** lncRNAs upregulated in bisex females after 5-AzaC exposure are compared with lncRNAs differentially expressed between the following conditions: lncRNAs enriched in single-sex (unpaired) females compared with bisex (paired) females (sF>bF); lncRNAs enriched in ovaries from single-sex (unpaired) females compared with ovaries from bisex (paired) females (sO>bO); lncRNAs enriched in bisex (paired) females compared with ovaries from bisex (paired) females (bF>bO). Samples are labeled as bF: bisex (paired) females; sF: single-sex (unpaired) females; bO: bisex (paired) ovaries; sO: single-sex (unpaired) ovaries, according to Lu *et al*., 2016 [3].



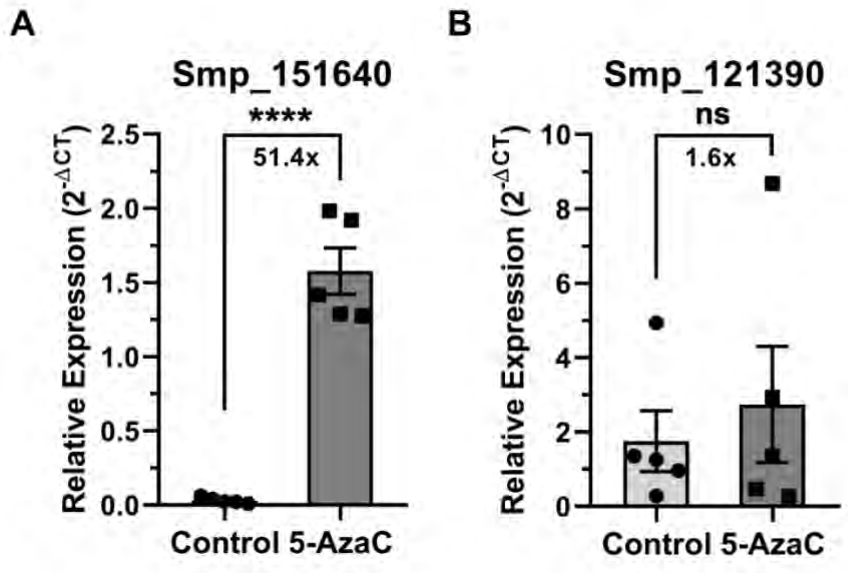
Supplementary Figure S3: Evaluation of *S. mansoni* schistosomula viability after 5-AzaC treatment at different concentrations and incubation times. (A) ATP quantitation using a luminescent assay to assess schistosomula survival under 5-AzaC exposure. *S. mansoni* schistosomula (100-110/well) were incubated with the indicated concentrations of 5-AzaC or with medium only (control) for 24, 48, 72, 96 and 120 h. Viability was expressed as % luminescence values relative to the control (medium only). Mean ± SEM from two replicate experiments. Two-way ANOVA was applied, and no statistically significant difference was found in any of the comparisons. **(B)** Light microscopy of schistosomula incubated with the indicated concentrations of 5-AzaC or with medium only (control) for 120 h. Scale bar: 100 µm.



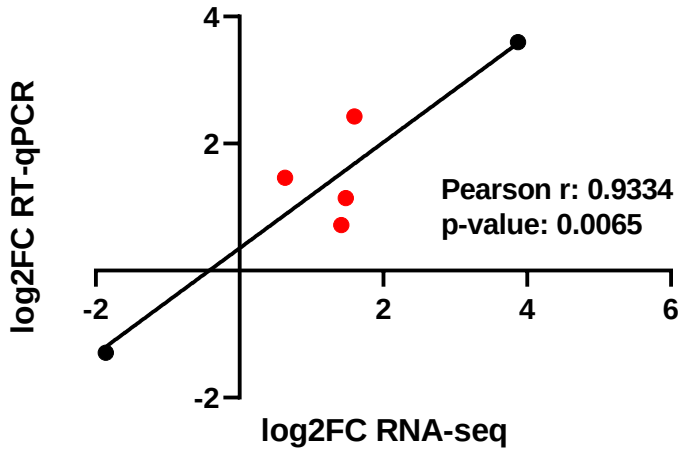
Supplementary Figure S4: Effect of 5-AzaC treatment on adult *S. mansoni* viability and egg laying. A. 5-AzaC does not affect adult *S. mansoni* viability. Nine worm pairs were cultivated in the presence (n = 5) or absence (n = 5) of 491 µM 5-AzaC for 48 h. Worms were collected and ATP levels were measured in control and treated worm couples. Student's unpaired two-sided t test; ns: not significant. **B.** 5-AzaC significantly inhibits *S. mansoni* egg production. Thirty adult worm pairs were cultured either in the presence or absence of 491 µM 5-AzaC. Each culture condition was replicated (n = 5) and eggs were collected and counted after 48 hours. Mean ± SEM are shown. Student's unpaired two-sided t test was applied. *p-value = 0.02. **C.** Light microscopy of schistosome eggs laid by control worm pairs (control) or 5-AzaC treated worm pairs for 48 h (5-AzaC, 491µM). Scale bar: 100 µm.



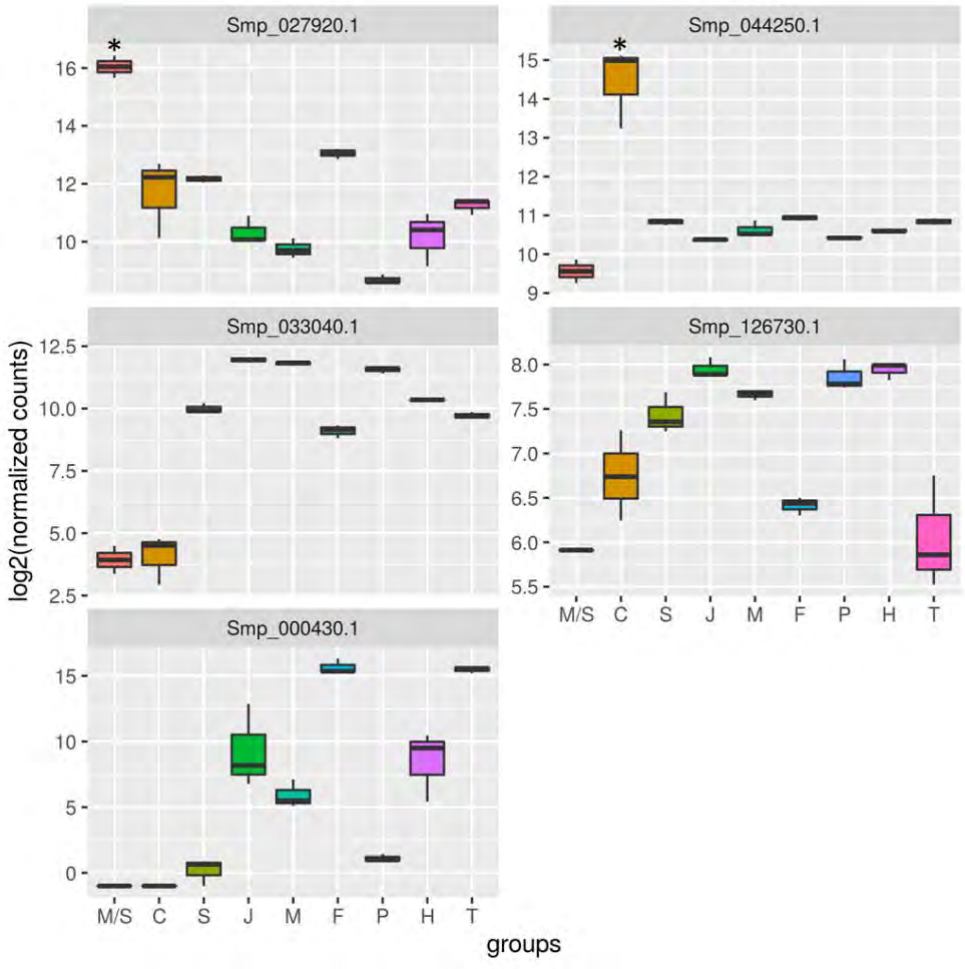
Supplementary Figure S5: Expression profiles in *S. mansoni* females of selected protein-coding genes differentially expressed after 5-AzaC treatment (491 µM). Two protein-coding genes were used as controls after re-analysis of RNA-Seq public datasets of 5-AzaC treated *S. mansoni* females from Geyer *et al*., 2018 [1] for validation by RT-qPCR in females. For each of the two protein-coding genes, their expression profiles in RNA-Seq are shown as TPM (transcripts per million) on the left, whereas the RT-qPCR results are shown on the right: **A**. Smp_151640 (*Insulin-like growth factor I*); **B**. Smp_121390 (*Genome polyprotein*). For the RNA-Seq data, three biological replicates were analyzed; the fold-changes and p-values represented by asterisks that are shown in the brackets were obtained using DESeq2. For the RT-qPCR data, mean ± SEM from five biological replicates are shown; *p<0.05, ****p<0.0001. Student's unpaired two-sided t test.



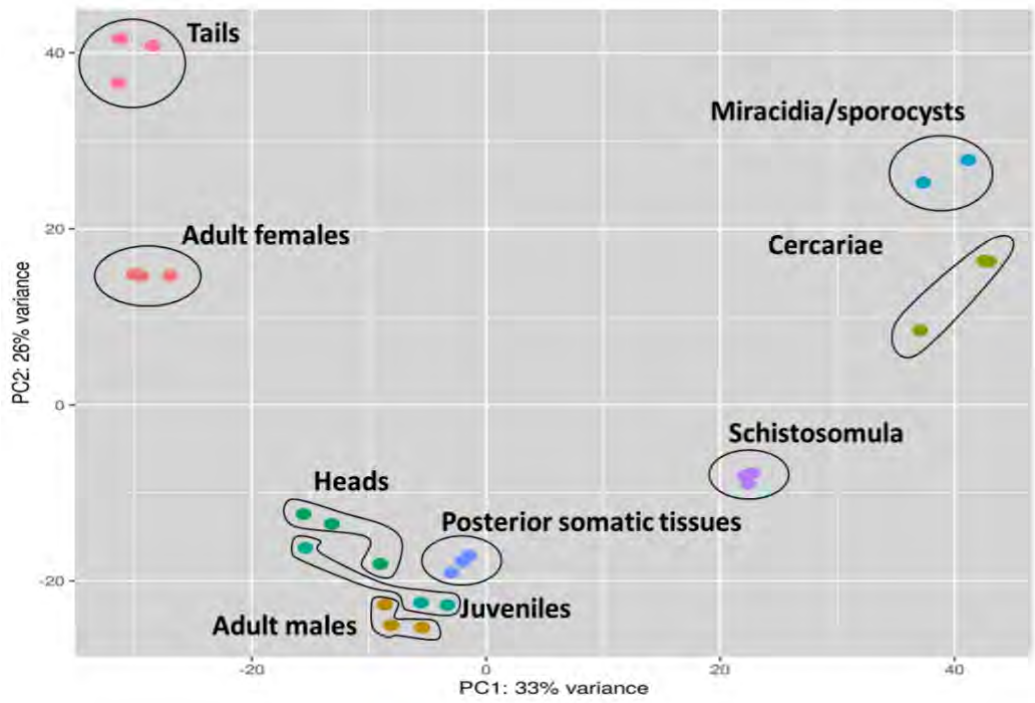
Supplementary Figure S6: Expression profiles in *S. mansoni* males of selected protein-coding genes differentially expressed after 5-AzaC treatment (491 µM). Two protein-coding genes were used as controls after re-analysis of RNA-Seq public datasets of 5-AzaC treated *S. mansoni* females from Geyer *et al*., 2018 [1] for evaluation of differential expression by RT-qPCR in males. For each of the two protein-coding genes, the expression profiles in controls and in 5-AzaC treated *S. mansoni* males by RT-qPCR are shown: **A**. Smp_151640 (*Insulin-like growth factor I*); **B**. Smp_121390 (*Genome polyprotein*). Mean ± SEM from five biological replicates are shown; Student's unpaired two-sided t test was applied. ****p<0.0001; ns: not significant.



Supplementary Figure S7: Correlation between RNA-Seq and RT-qPCR analysis. Pearson correlation between the fold changes (FC) in expression measured by RNA-seq or RT-qPCR of six selected genes (four lincRNAs – red dots – and two protein-coding genes – black dots); fold changes were obtained by measuring the expression after treatment of females with 5-AzaC and comparing with expression in the control condition. Log2FC of the six genes obtained with the RNA-Seq assay is represented in the x-axis, and log2FC of the six genes obtained with RT-qPCR is represented in the y-axis.



Supplementary Figure S8: RNA-seq expression levels in *different S. mansoni* stages of protein-coding genes used as sample markers. The expression levels (shown as log2 of normalized counts) of the five protein-coding genes whose gene IDs are indicated at the top of each panel are shown. The y-axis shows the expression level for each protein-coding gene in the RNA-seq assays (log2 of normalized counts) as determined at the stage indicated in the x-axis as follows: miracidia/sporocysts (M/S), cercariae (C), schistosomula (S), juveniles (J), adult males (M), adult females (F), posterior somatic tissues (P), heads (H) and tails (T). **A.** Smp_027920 (*Tubulin alpha-1 chain*, with high expression in eggs); **B**. Smp_044250 (*STAM-binding protein*, with high expression in cercariae); **C**. Smp_033040 (*L-lactate dehydrogenase A chain*, with high expression in schistosomula, juveniles and adult males); **D.** Smp_126730 (*5-hydroxytryptamine receptor 1A*, with high expression in juveniles and adult males); **E.** Smp_000430 (*Putative eggshell protein,* with high expression in adult females). Only transcripts that were upregulated in one stage/tissue when compared with all others were considered as significantly more expressed in that stage/tissue and are marked with an asterisk. *p-value < 0.05.



Supplementary Figure S9: Clustering of RNA-Seq biological replicates assessed by principal component analysis (PCA). RNA-Seq data from 26 public RNA-Seq libraries (listed in Supplementary Table S7) representing six life-cycle stages (miracidia/sporocysts, cercariae, schistosomula, juveniles, adult males, adult females) and adult worm heads, tails and posterior somatic tissues were re-analyzed using the *S. mansoni* genome PRJEA36577 (v.7) retrieved from WormBase and the recently published transcriptome that includes long non-coding RNAs [2] as reference. PCA plot was obtained after normalization using the vst function followed by the plotPCA function from DESeq2.

Supplementary References

1 Geyer, K. K. *et al.* The anti-fecundity effect of 5-azacytidine (5-AzaC) on Schistosoma mansoni is linked to dis-regulated transcription, translation and stem cell activities. *Int J Parasitol Drugs Drug Resist* **8**, 213-222, doi:10.1016/j.ijpddr.2018.03.006 (2018).

2 Maciel, L. F. *et al.* 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 (2019).

3 Lu, Z. *et al.* Schistosome sex matters: a deep view into gonad-specific and pairing-dependent transcriptomes reveals a complex gender interplay. *Sci Rep* **6**, 31150, doi:10.1038/srep31150 (2016).

2.3. Análises de RNA-seq de célula única mostram que RNAs longos não-codificadores são visivelmente expressos em populações de células de gametas e progenitoras do tegumento de *Schistosoma mansoni*

PREÂMBULO

Contribuições do Doutorando Gilbert de Oliveira Silveira para o manuscrito apresentado nesta sessão:

Selecionou os genes para validação por hibridização *in situ*, desenhou todos os *primers* para a clonagem dos genes selecionados. Clonou todos os genes selecionados em um plasmídeo otimizado para a preparação de sondas para hibridização *in situ*. Sintetizou todas as sondas marcadas com digoxigenina e fluoresceína para os ensaios de hibridização *in situ* por colorimetria e por fluorescência. Interpretou os resultados de hibridização *in situ*. Leu e revisou o manuscrito.

Morales-Vicente DA, Zhao L, Silveira GO, Tahira AC, Amaral MS, Collins III JJ 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

© 2022 Morales-Vicente, Zhao, Silveira, Tahira, Amaral, Collins III and Verjovski-Almeida. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). 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.

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 20 September 2022
DOI 10.3389/fgene.2022.924877

Check for updates

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

†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). 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[1,2†], Lu Zhao[3†], Gilbert O. Silveira[1,2†], Ana C. Tahira[1], Murilo S. Amaral[1], James J. Collins III[3] and Sergio Verjovski-Almeida[1,2*]

[1]Laboratório de Parasitologia, Instituto Butantan, São Paulo, Brazil, [2]Instituto de Química, Universidade de São Paulo, São Paulo, Brazil, [3]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://verjolab.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 Velocyto 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 synthetized 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 re-mapped 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^{+}, 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 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+ 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

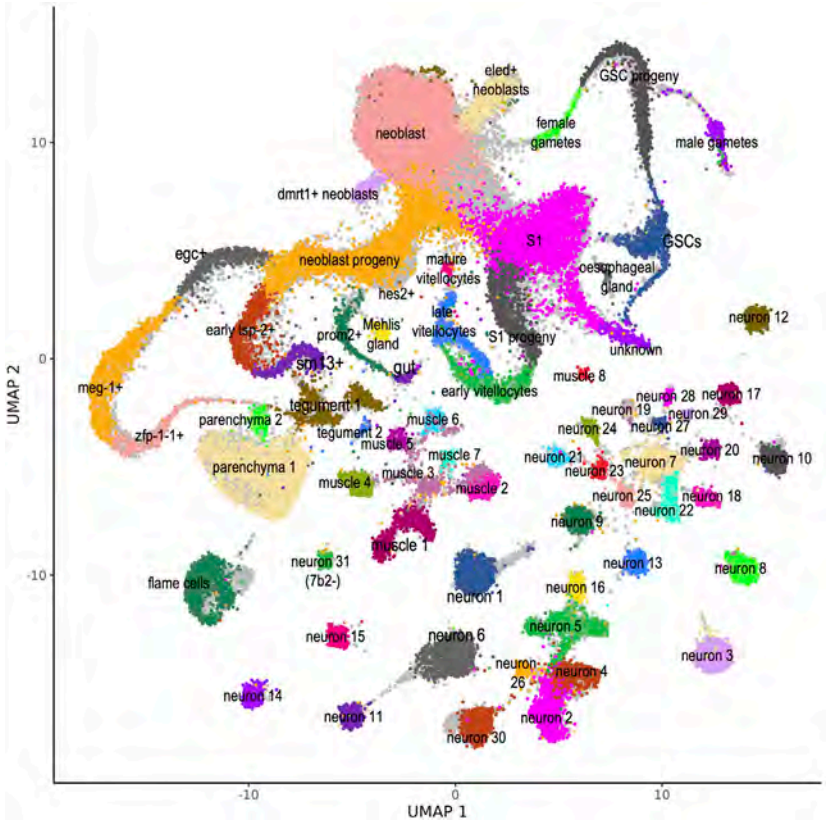


FIGURE 1
S. mansoni atlas of single-cells comprising expression data of protein-coding and lncRNA genes. UMAP plot of the 68 scRNA-seq clusters identified by Wendt et al. (2020) and projected onto our re-analyzed scRNA-seq data set. For each of the 48,094 cells recovered in our re-analysis, expression data for protein-coding and lncRNA genes was used, and all cells were assigned to one of the 68 clusters (see Methods). Cells are colored according to the cluster where they were mapped. Original cell cluster mapping data from Wendt et al. (2020) is shown in the background, colored in light grey. This atlas is available as a public website resource http://verjolab.usp.br:8081.

transfer mislabel due to difference in the gene count strategies. A possible explanation for the loss of cells from one cluster to another is the overall similarity between many of the Wendt et al. (2020) clusters, as documented by the correlation coefficient between the median expression of clusters in the reference matrix (Supplementary Figure S1B). One good example of this is the hes2$^+$, a subcluster of neoblast progeny, that lost most of its cells to the neoblast progeny cluster; both clusters are highly similar as determined by the Pearson correlation (0.973) (Supplementary Figure S1B).

Our filtering and mapping strategy (see Methods) was able to recover 5039 new, previously non-identified cells, of which 884 (17.5%) were mapped to the neoblast progeny cluster, followed by 394 cells (7.8%) mapped to the parenchyma 1 cluster, and 262 cells (5.2%) mapped to the neuron 5 cluster (Supplementary Table S4, orange). The remaining previously non-identified cells were mapped at different smaller extents (4.4–0.02%) to 64 other clusters (Supplementary Table S4).

3.3 Identification of lncRNA markers of cell clusters

To find lncRNA markers of cell clusters we looked for lncRNAs which were significantly more highly expressed in one cluster compared with all other clusters. For this, we applied a variance stabilization transformation of the data (see Methods) using the "regularized negative binomial regression" statistical approach (Hafemeister and Satija, 2019) to remove from the downstream analyses the influence of scRNA-seq technical characteristics such as sequencing depth, while preserving biological heterogeneity. Subsequently, a differential expression analysis among all clusters identified 74 lncRNA genes as markers of cell clusters (Figure 3).

A set of 18 lncRNAs were conspicuously expressed in male gametes (Figure 3, dotted black box). A total of 7 lncRNAs were expressed in the meg-1$^+$ cluster (Figure 3, blue box), and 5 of them were also expressed in the zfp-1-1$^+$ cluster, two clusters that belong to the tegument lineage.

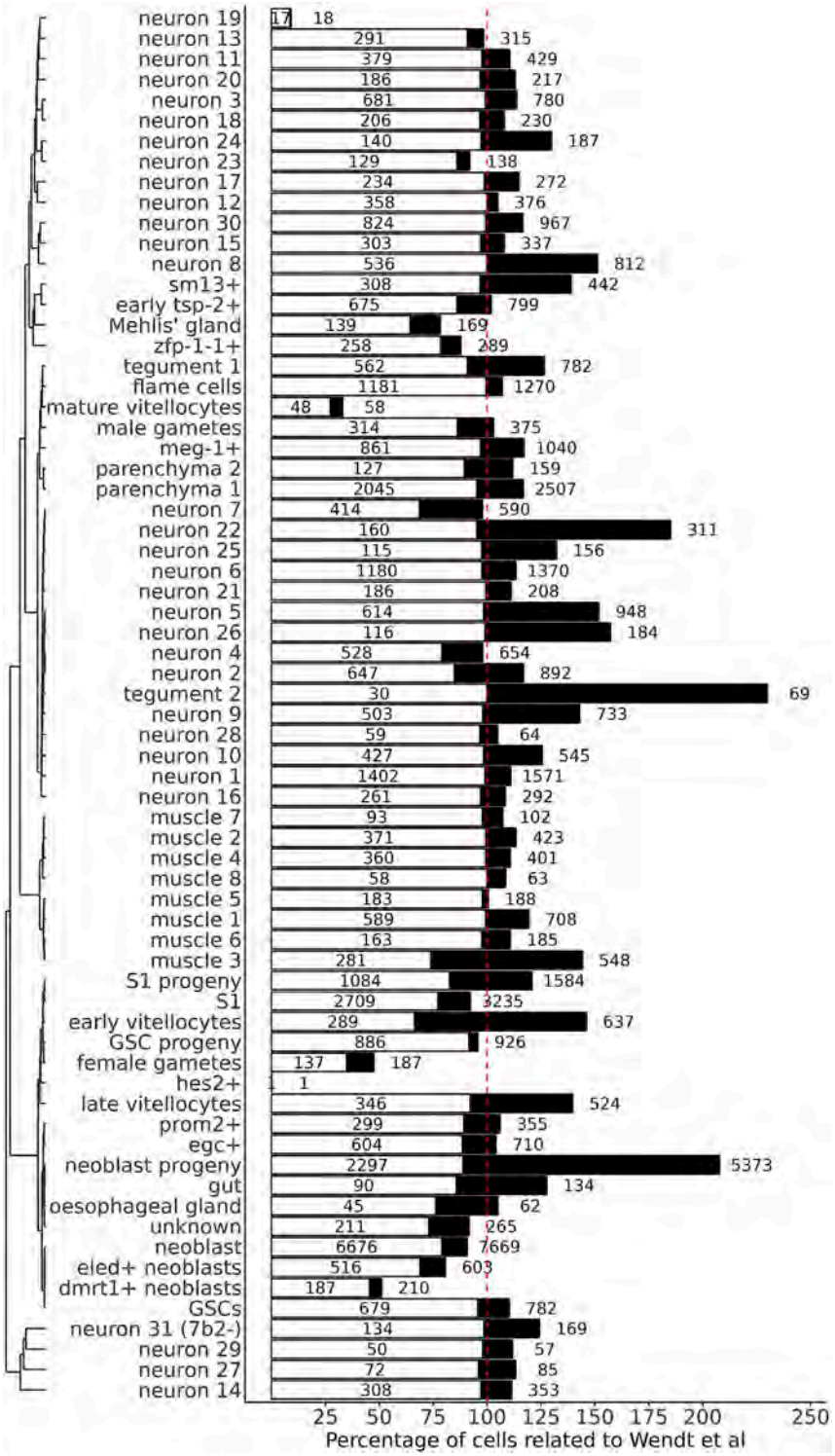


FIGURE 2
Percentage of cells mapped to each cluster in comparison to the number of cells in the original cluster annotation. The 68 clusters (indicated at left) are grouped according to the similarity of their gene expression patterns in the new, re-analyzed expression data set. The white bars indicate 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.

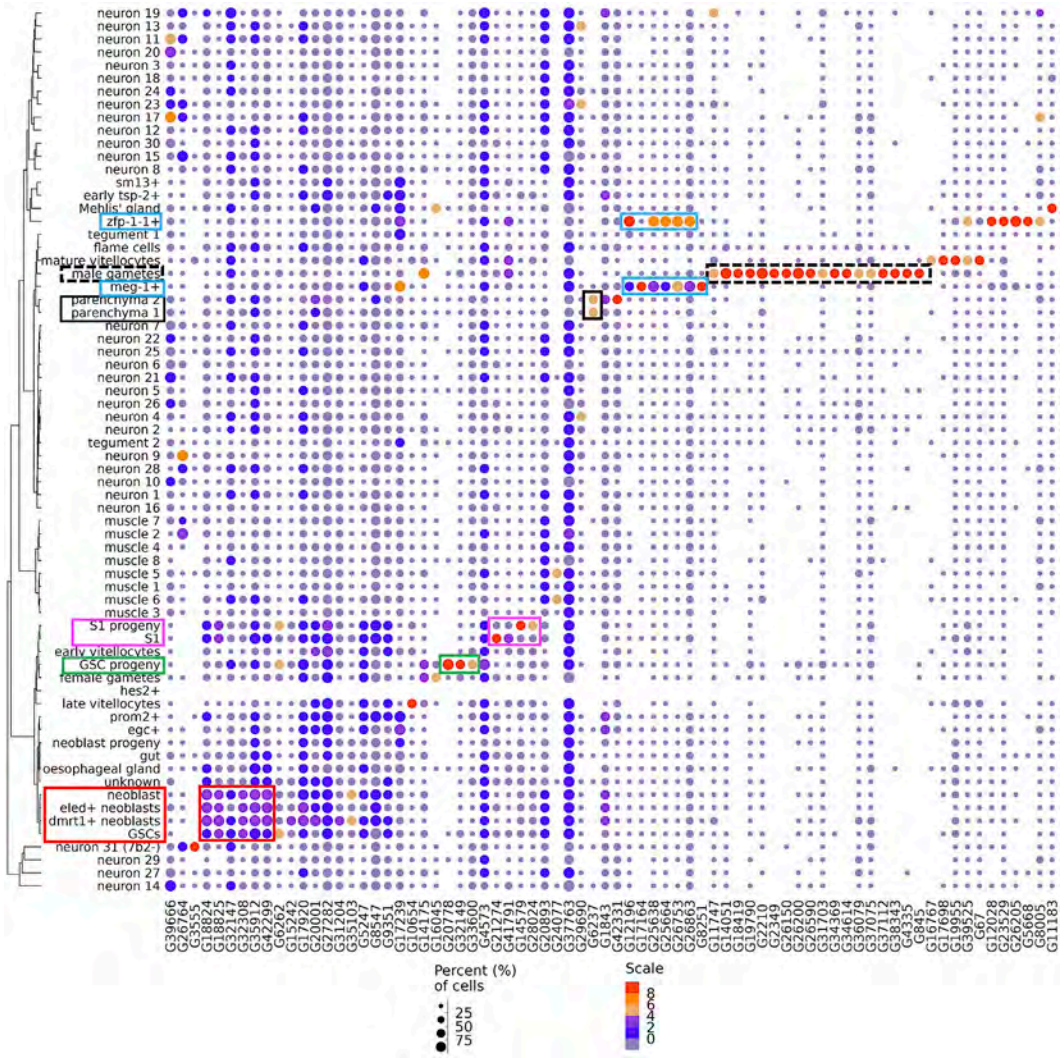


FIGURE 3
Dot-plot of 74 lncRNA genes identified as markers of single-cell clusters. A dot-plot summarizing the cluster-specific expression of each of the 74 lncRNA genes identified as markers of cell clusters in adult *S. mansoni*. Cluster IDs are on the vertical axis and gene IDs are on the horizontal 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). The colored boxes highlight the lncRNAs cited in the Results.

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

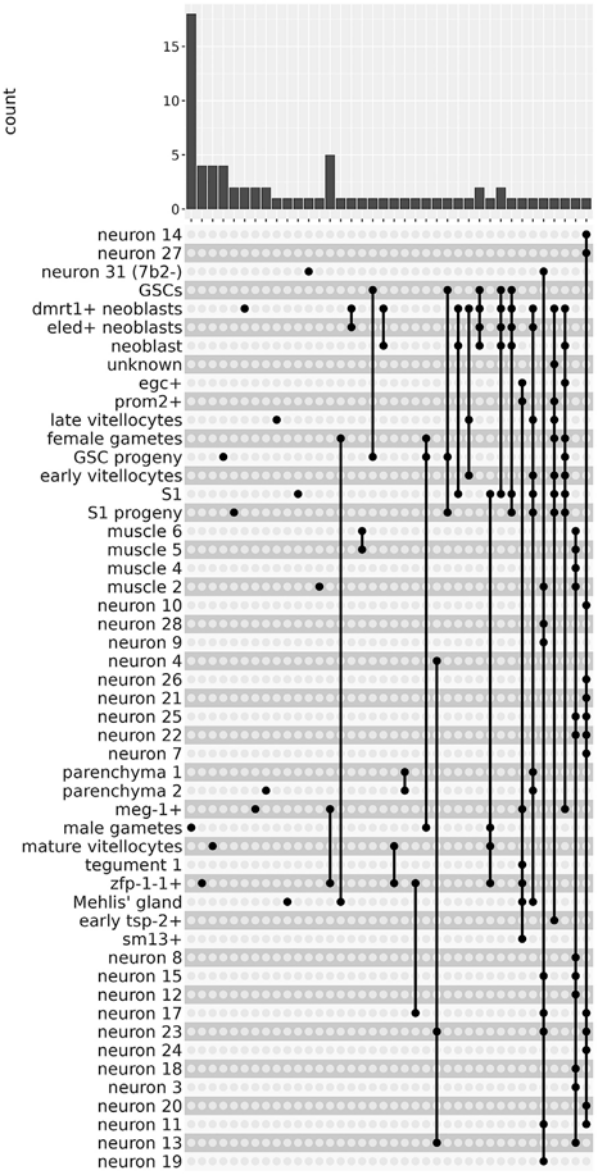


FIGURE 4
The lncRNAs are markers of 51 different single-cell clusters. The UpSet intersection diagram shows on the upper panel the number of *S. mansoni* lncRNAs (*y*-axis) that have been detected in each of the intersection sets, indicated by the connected points in the lower part of the plot, as being markers of the indicated single-cell clusters. On the left-most part of the plot are the lncRNAs that are markers of a unique single-cell cluster. On the right-most part are the lncRNAs that are markers of a group of single-cell clusters, joined with the connected dots.

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 co-localized 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).

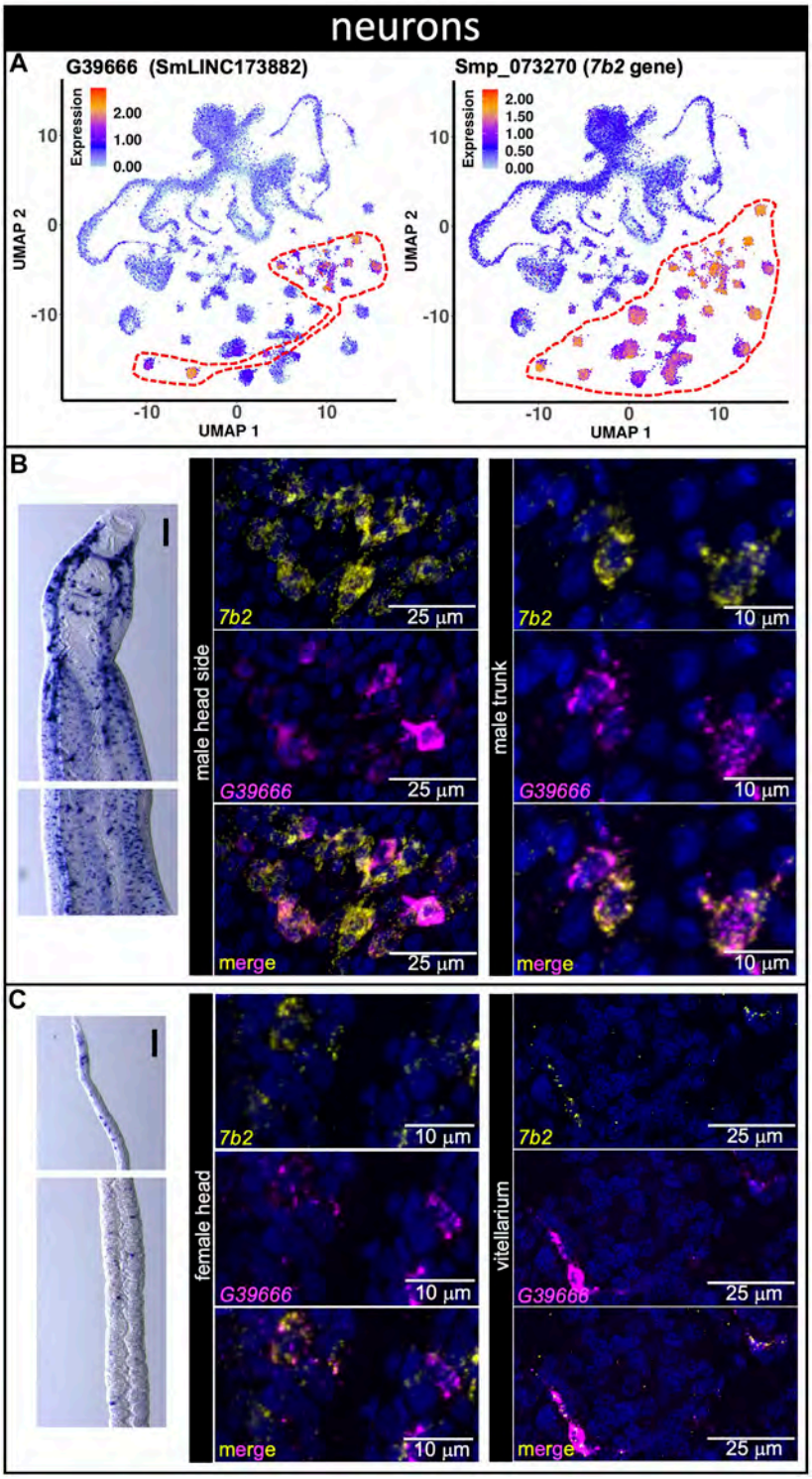


FIGURE 5
A lncRNA marker of 13 different neuron clusters co-localizes with neuroendocrine protein *7b2* message. **(A)** UMAP plot of lncRNA neurons cluster marker G39666 (left) and of general neuronal marker *7b2* (right). UMAP plots are colored by gene expression (blue = low, red = high) and the scale represents log10(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.

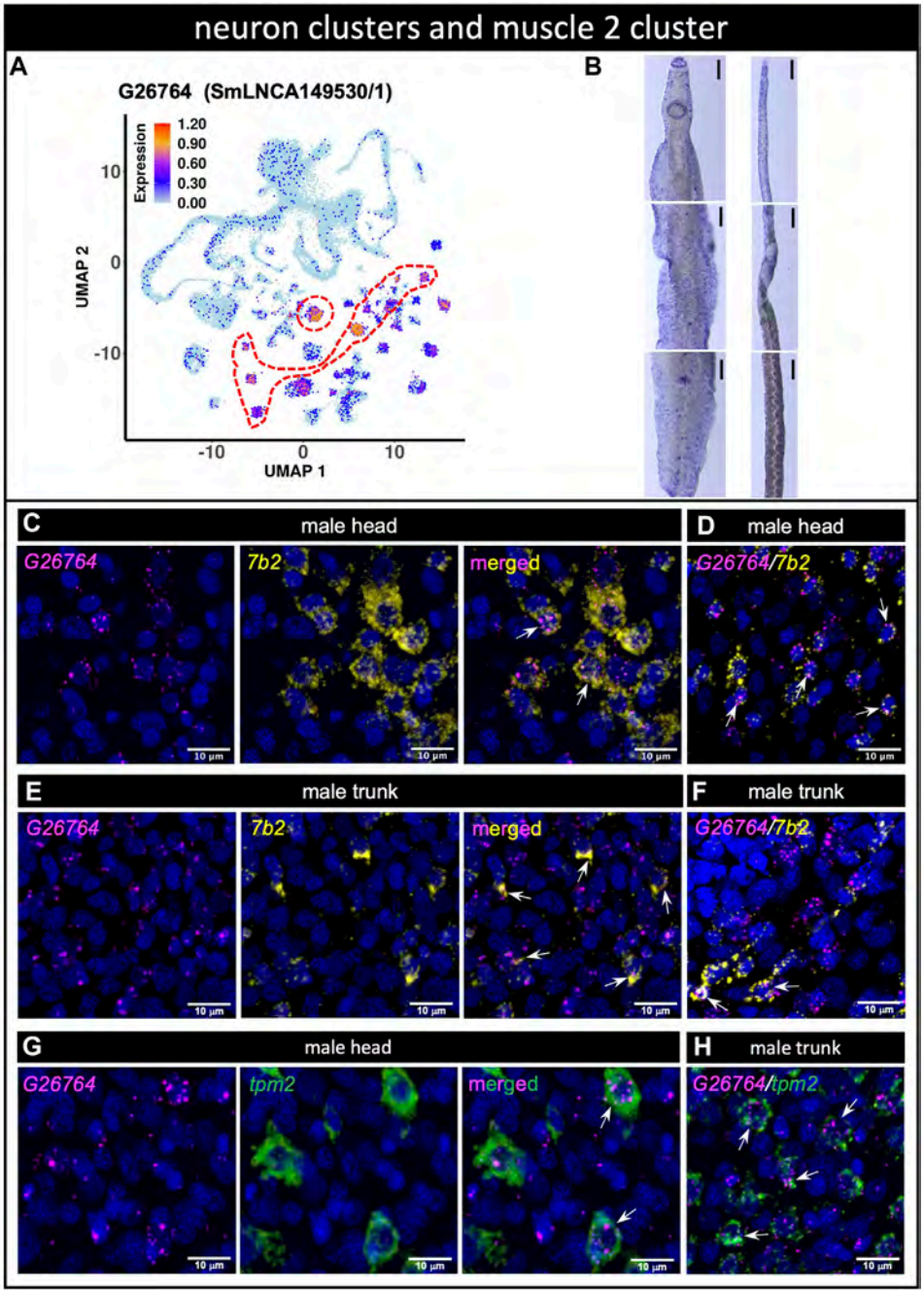


FIGURE 6
(A) lncRNA marker of 8 different neuron clusters and of muscle 2 cluster co-localizes with neuroendocrine protein *7b2* and muscle tropomyosin genes. (A) UMAP plot of lncRNA cluster marker G26764. UMAP plot is colored by gene expression (blue = low, red = high) and the scale represents log10(UMIs+1). The region enclosed by the red dashed line indicates the location of the relevant neuron cluster on the UMAP plot. **(B)** WISH of lncRNA G26764 in the head (top), trunk (middle) and tail (bottom) of a male [left] or a mature female [right]. Scale bars are 100 μm. **(C–F)** Double FISH with lncRNA 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 lncRNA G26764 and the general muscle marker gene *tpm2* tropomyosin in the male head **(G)** and trunk **(H)**. Nuclei: blue.

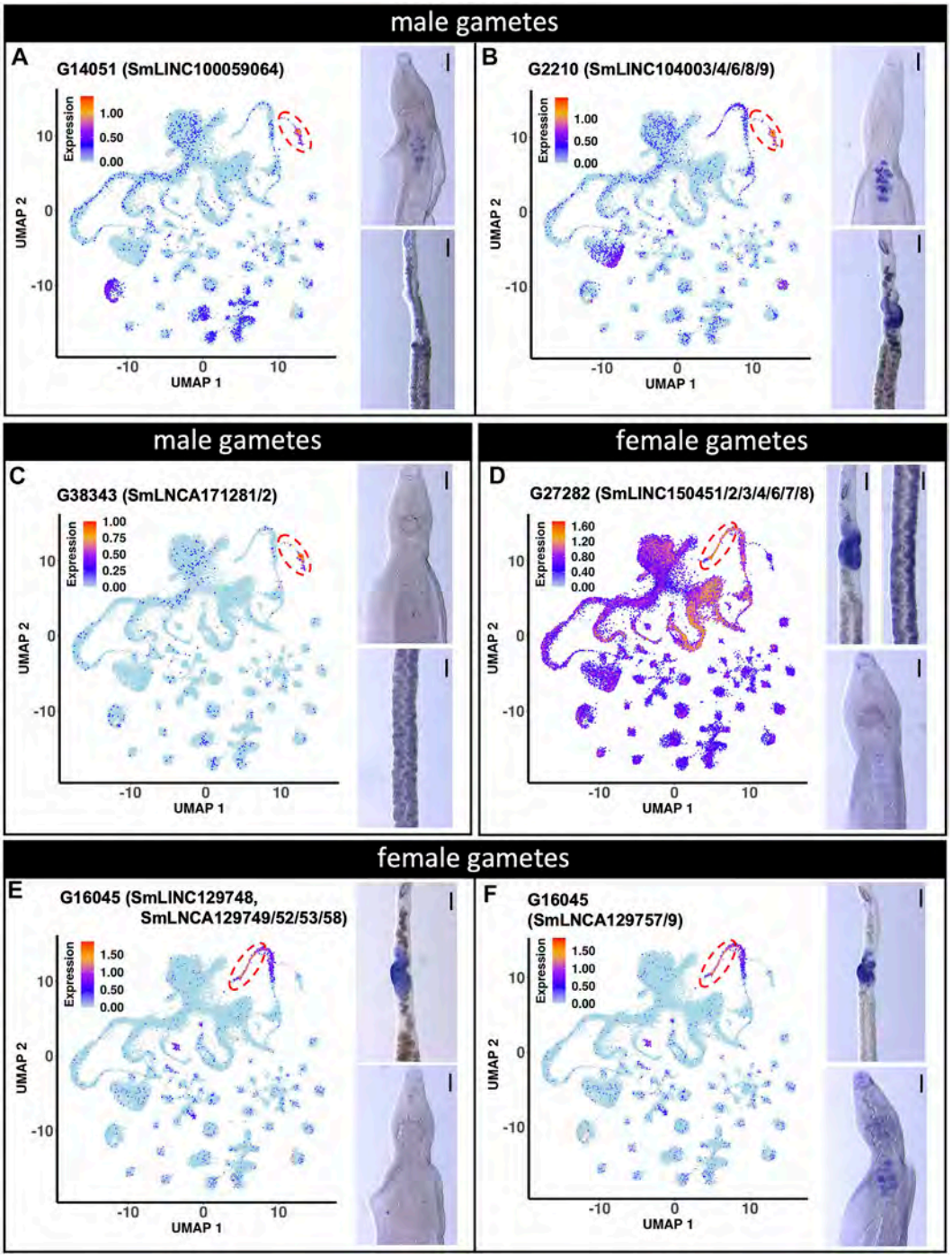


FIGURE 7
lncRNA markers of male and female gametes single-cell clusters are localized in testis and ovary. **(A** to **C)** UMAP plot (left) of the indicated lncRNA marker of male gametes cluster. WISH with the indicated lncRNA in a male head [right, top] and the ovary region of a female [right, bottom]. **(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 log10(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 μm.

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). 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 lincRNAs. G17239 (SmLINC131974) was identified with scRNA-seq as a marker of meg-1$^+$ 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$^+$ and *egc*$^+$ clusters (Figure 8A). The second tegument progenitor marker, G12028 (SmLINC122388) was identified with scRNA-seq as a marker of zfp-1-1$^+$ 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$^+$ cluster (Figure 8I) with expression also in the meg-1$^+$ 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$^+$ 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

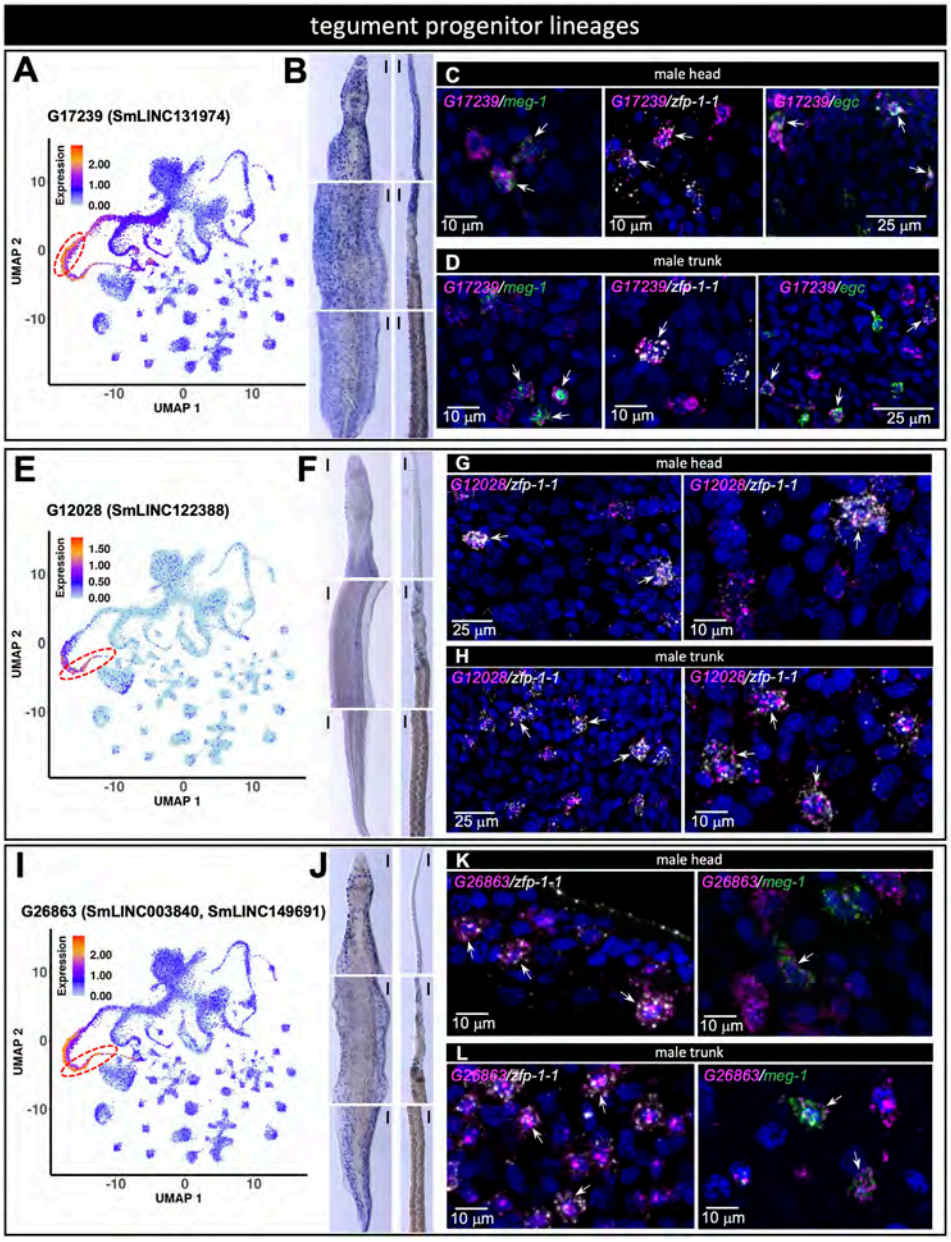


FIGURE 8
lncRNA markers of tegument progenitor lineages co-localize with protein-coding genes known to mark those tegument progenitors. **(A,E,I)** UMAP plot of the indicated lncRNA marker of tegument progenitor lineages. **(B,F,J)** WISH with the indicated lncRNA in a male [left] and a female [right] 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 log10(UMIs+1). The regions enclosed by the red dashed lines indicate the location of the relevant meg-1$^+$ **(A)** or zfp-1-1$^+$ **(E,I)** clusters on the UMAP plots. WISH scale bars are 100 µm.

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 *tgfbi* (Supplementary Figures S4C-H). LncRNA G17920 (SmLINC133371), was identified by scRNA-seq as a marker of the dmrt$^+$ and eled$^+$ 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^{+} 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) 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 2.2×10^{-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

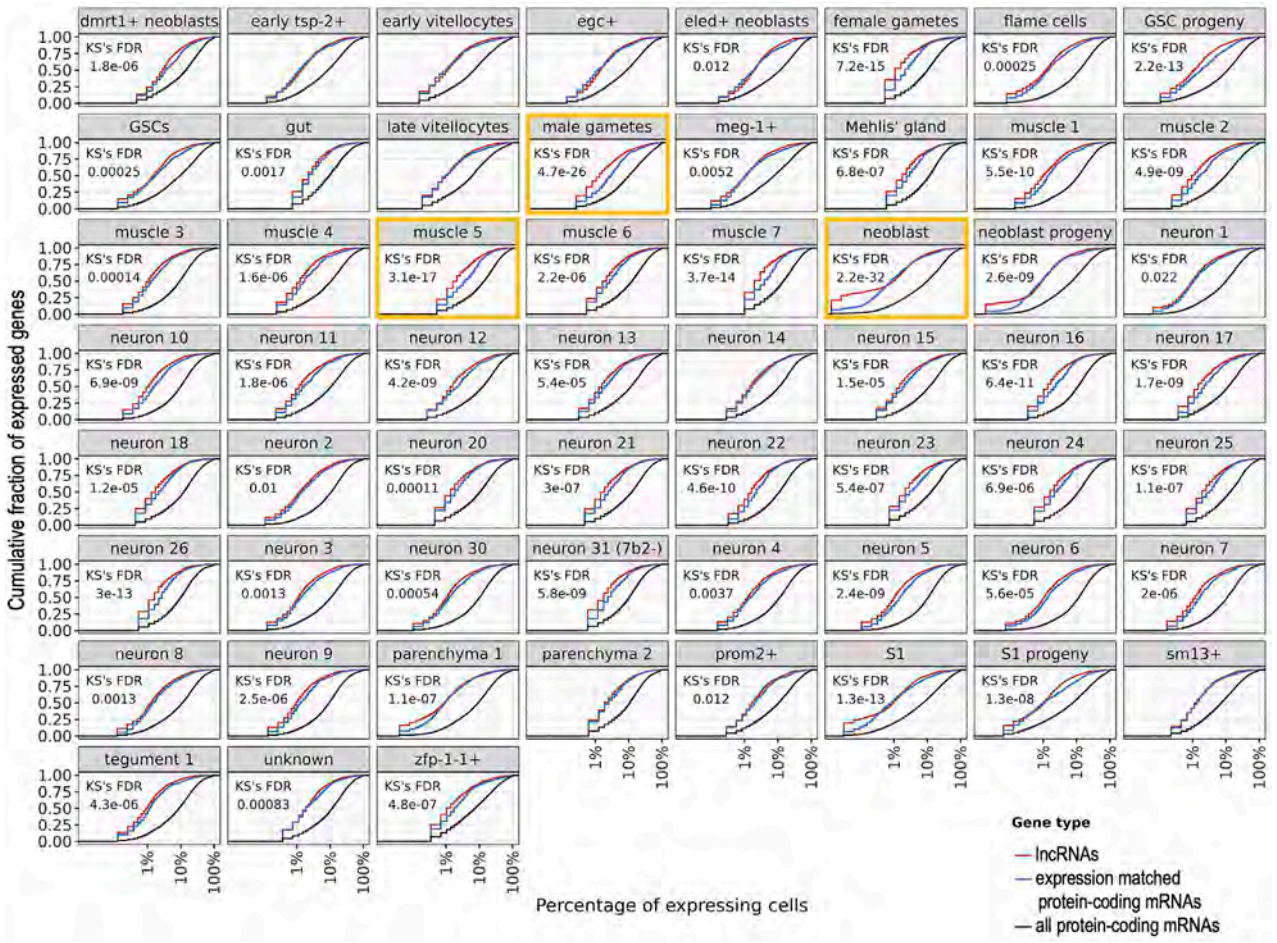


FIGURE 9
Heterogeneity of expression of all lncRNAs or mRNAs that were detected in a given cluster of cells. For each cluster named at the top of each panel, the cumulative fraction of all lncRNAs or mRNAs that were detected as expressed in the cluster (*y*-axis) is shown as a function of the percentage of cells in which the lncRNAs or the mRNAs were detected (*x*-axis). For each cluster, the red curve shows the detected lncRNAs, the blue curve shows the set of mRNAs detected with expression levels in the same range as that of the lncRNAs, and the black curve shows the complete set of mRNAs detected in the cluster. The Kolmogorov-Smirnov (KS) statistical test False Discovery Rate (FDR) is shown for the comparison between the lncRNAs and the expression-matched set of mRNAs; in the seven panels where no KS FDR is shown, no statistical difference was found (FDR >0.05). The three clusters with the most significant differences (lowest FDRs) are marked with orange rectangles. Nine clusters with less than 100 cells each were excluded from this analysis.

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 Smps 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^{+} neoblasts, male gametes, eled^{+} 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^{+}, 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^{+} and zfp-1-1^{+} clusters, two clusters that belong to the tegument lineage. With dFISH, localization of 6 lncRNA markers in the meg-1^{+}, zfp-1-1^{+} and egc^{+} 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), 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), 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.

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

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_ConsortiumBirney, 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., Strobelt, 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

Maciel, 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

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

Statello, 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

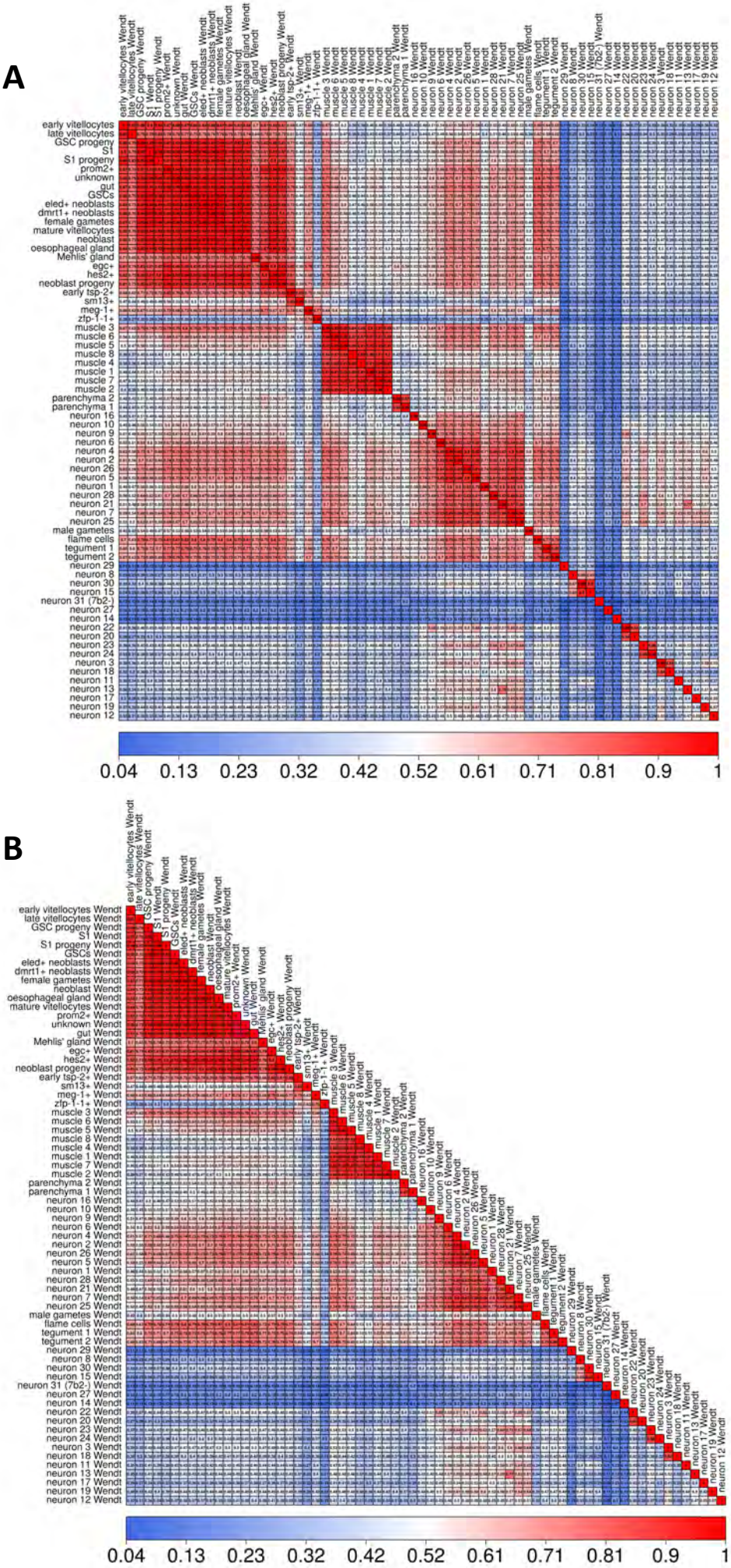
Wendt, G., Zhao, L., Chen, R., Liu, C., O'donoghue, A. J., Caffrey, C. R., et al. (2020). A single-cell RNA-seq atlas of *Schistosoma mansoni* identifies a key regulator of blood feeding. *Science* 369, 1644–1649. doi:10.1126/science.abb7709

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

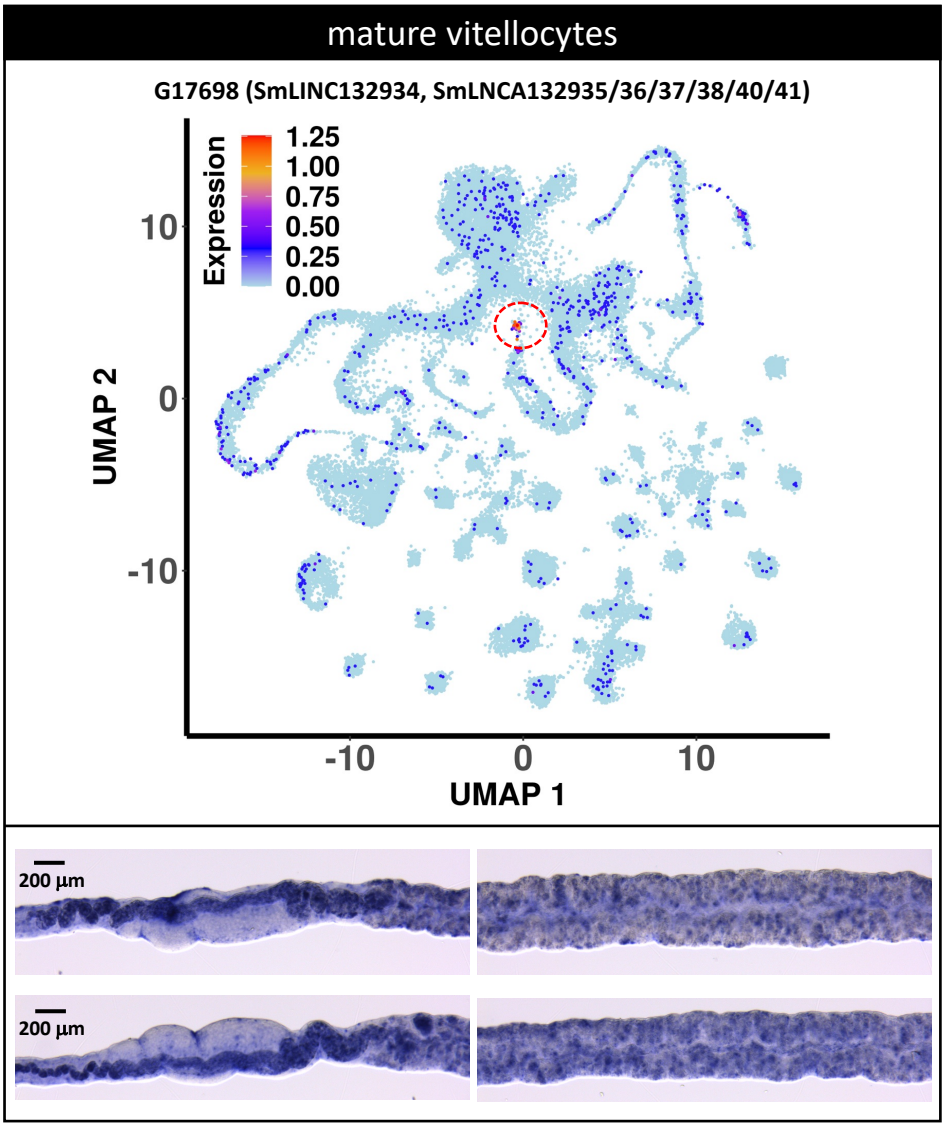
Zhou, J., Xu, J., Zhang, L., Liu, S., Ma, Y., Wen, X., et al. (2019). Combined single-cell profiling of lncRNAs and functional screening reveals that H19 is pivotal for embryonic hematopoietic stem cell development. *Cell Stem Cell* 24, 285–298. doi:10.1016/j.stem.2018.11.023

Figure S1


A
early vitellocytes Wendt
late vitellocytes Wendt
GSC progeny Wendt
S1 Wendt
S1 progeny Wendt
prom2+ Wendt
unknown Wendt
gut Wendt
GSCs Wendt
eled+ neoblasts Wendt
dmrt1+ neoblasts Wendt
female gametes Wendt
mature vitellocytes Wendt
neoblast Wendt
oesophageal gland Wendt
Mehlis' gland Wendt
egc+ Wendt
hes2+ Wendt
neoblast progeny Wendt
early tsp-2+ Wendt
sm13+ Wendt
meg-1+ Wendt
zfp-1-1+ Wendt
muscle 3 Wendt
muscle 6 Wendt
muscle 5 Wendt
muscle 8 Wendt
muscle 4 Wendt
muscle 1 Wendt
muscle 7 Wendt
muscle 2 Wendt
parenchyma 2 Wendt
parenchyma 1 Wendt
neuron 16 Wendt
neuron 10 Wendt
neuron 9 Wendt
neuron 6 Wendt
neuron 4 Wendt
neuron 2 Wendt
neuron 26 Wendt
neuron 5 Wendt
neuron 1 Wendt
neuron 28 Wendt
neuron 21 Wendt
neuron 7 Wendt
neuron 25 Wendt
male gametes Wendt
flame cells Wendt
tegument 1 Wendt
tegument 2 Wendt
neuron 29 Wendt
neuron 8 Wendt
neuron 30 Wendt
neuron 15 Wendt
neuron 31 (7b2-) Wendt
neuron 27 Wendt
neuron 14 Wendt
neuron 22 Wendt
neuron 20 Wendt
neuron 23 Wendt
neuron 24 Wendt
neuron 3 Wendt
neuron 18 Wendt
neuron 11 Wendt
neuron 13 Wendt
neuron 17 Wendt
neuron 19 Wendt
neuron 12 Wendt
0.04 0.13 0.23 0.32 0.42 0.52 0.61 0.71 0.81 0.9 1
B
0.04 0.13 0.23 0.32 0.42 0.52 0.61 0.71 0.81 0.9 1

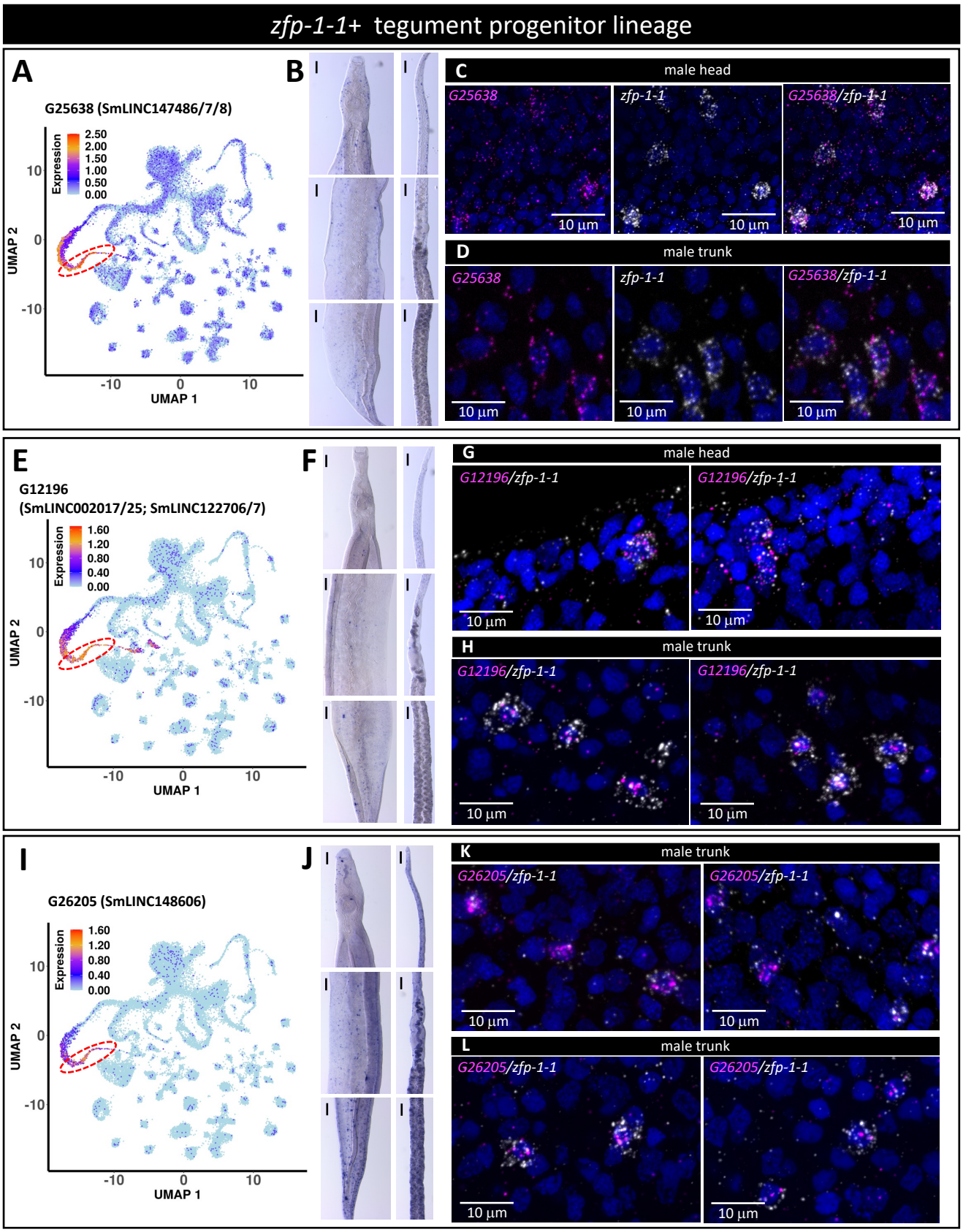
Supplementary Figure S1 – Correlation coefficient between the median expression of clusters in the reference and query matrices. Pearson correlation between the median expression of genes in the cells from the indicated clusters is shown in (**A**) for the comparison between the query matrix generated in the present work that includes lncRNAs and protein-coding genes and the Wendt et al. (2020) reference matrix with protein-coding genes only; and in (**B**) for the comparison between clusters of the Wendt et al. (2020) reference matrix.

Figure S2



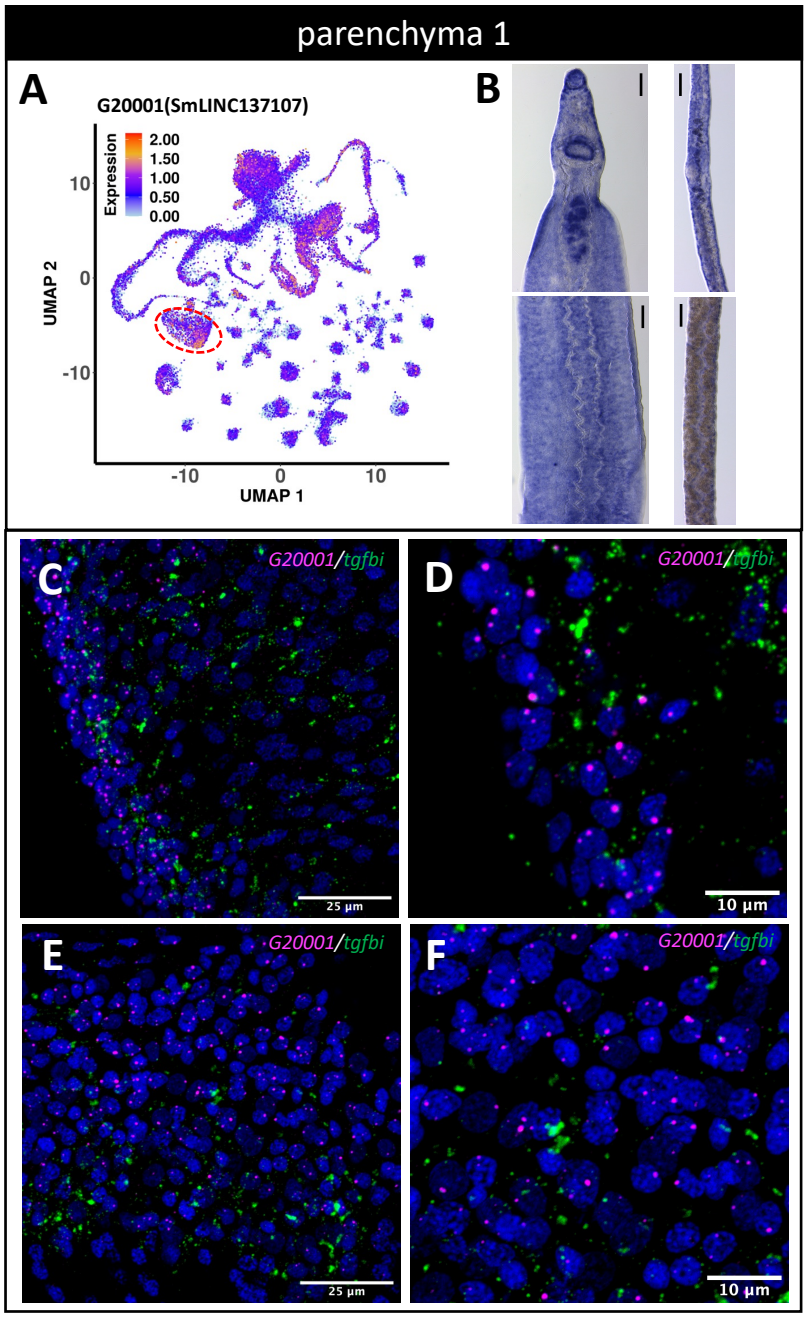
Supplementary Figure S2 – lncRNA marker of mature vitellocytes cluster. UMAP plot (top) of lncRNA G17698 marker of mature vitellocytes cluster. WISH (bottom) with lncRNA G17698 in the ovary region of two females [left, top and bottom] and in the vitellaria [right, top and bottom]. UMAP plots are colored by gene expression (blue = low, red = high) and the scale represents log10(UMIs+1). The region enclosed by the red dashed line indicates the location of the relevant mature vitellocytes cluster on the UMAP plot.

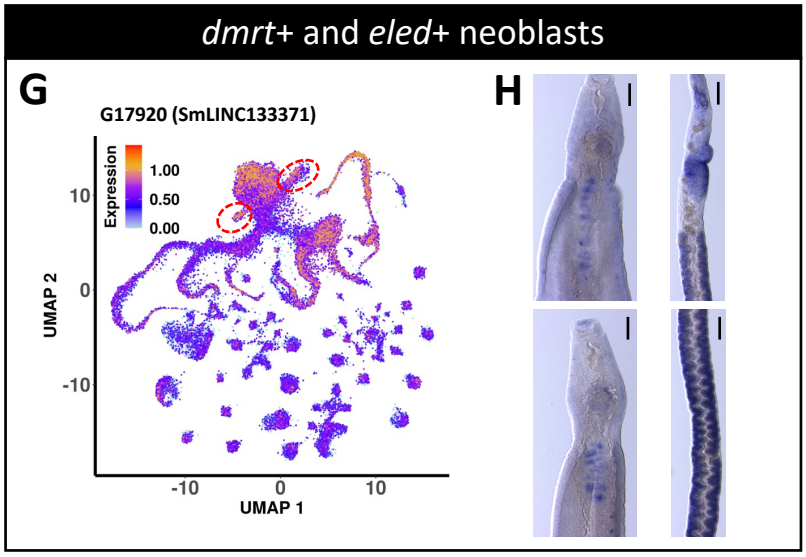
Figure S3


zfp-1-1+ tegument progenitor lineage
A
G25638 (SmLINC147486/7/8)
Expression
2.50
2.00
1.50
1.00
0.50
0.00
UMAP 2
UMAP 1
B
C
male head
G25638
zfp-1-1
G25638/zfp-1-1
10 µm
D
male trunk
E
G12196
(SmLINC002017/25; SmLINC122706/7)
1.60
1.20
0.80
0.40
0.00
F
G
male head
G12196/zfp-1-1
H
male trunk
I
G26205 (SmLINC148606)
J
K
male trunk
G26205/zfp-1-1
L
male trunk

Supplementary Figure S3 – Additional lncRNA markers of *zfp-1-1*+ tegument progenitor cluster. (**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**) Double FISH in male head with the indicated lncRNA and the general tegument progenitor marker gene *zfp-1-1*. (**D, H, K, L**) Double FISH in male trunk with the indicated lncRNA and the general tegument progenitor marker gene *zfp-1-1*. UMAP plots are colored by gene expression (blue = low, red = high) and the scale represents log10(UMIs+1). The regions enclosed by the red dashed lines indicate the location of the relevant *zfp-1-1*+ cluster on the UMAP plots. WISH scale bars are 100 μm.

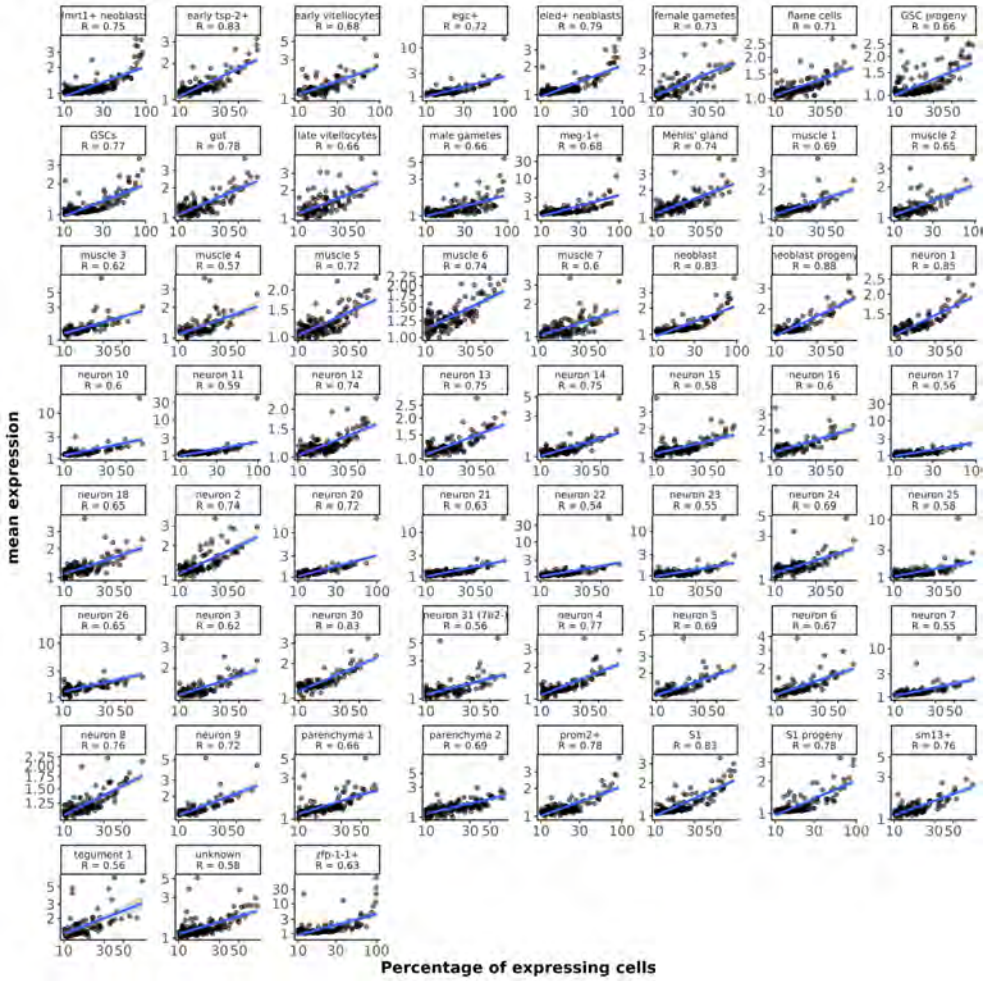
Figure S4


parenchyma 1
A
G20001(SmLINC137107)
Expression
2.00
1.50
1.00
0.50
0.00
UMAP 2
UMAP 1
10
0
-10
B
C
G20001/tgfbi
25 µm
D
10 µm
E
F

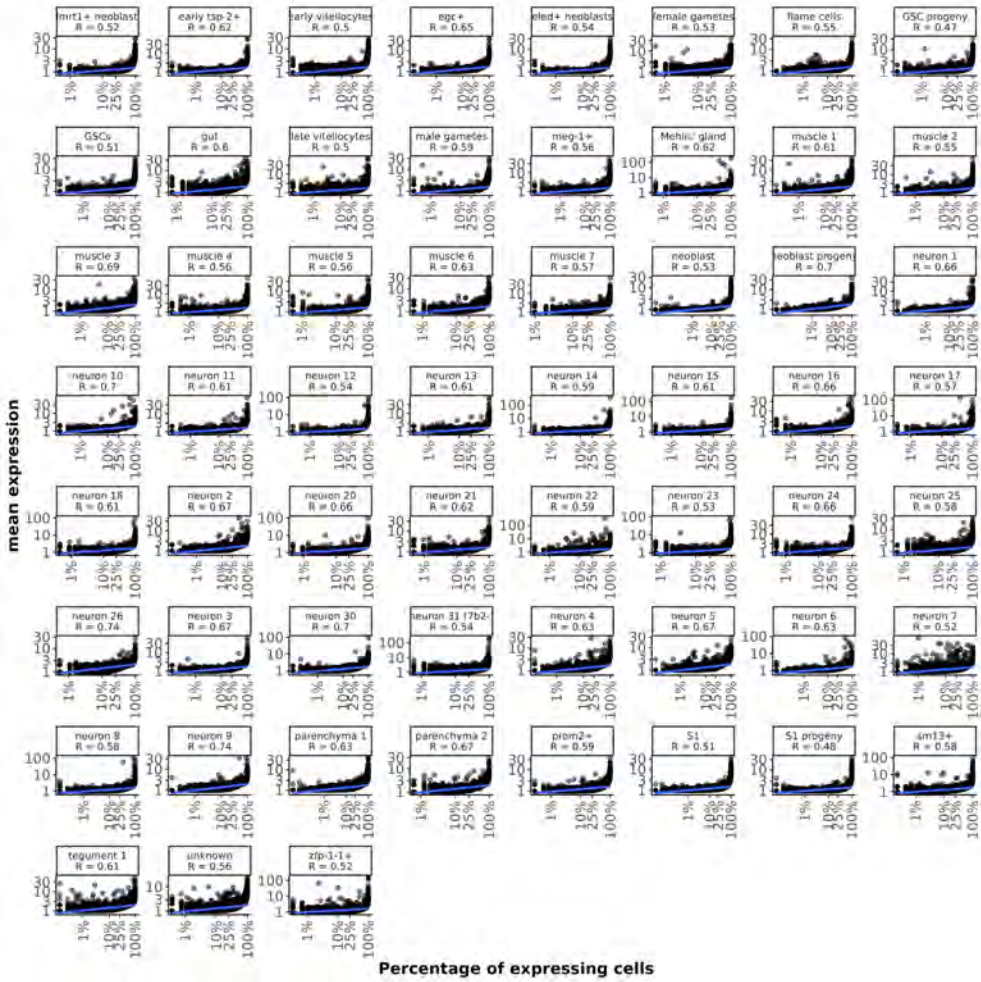

dmrt+ and eled+ neoblasts
G
G17920 (SmLINC133371)
Expression
1.00
0.50
0.00
UMAP 2
UMAP 1
10
0
-10
H

Supplementary Figure S4 – Two lncRNA candidate markers were not validated in the corresponding clusters. (**A**) UMAP plot of lncRNA G20001 marker of parenchyma 1 cluster. (**B**) WISH with lncRNA G20001 in the head [left, top] and trunk [left, bottom] of a male, and in the ovary region [right, top] and vitellaria [right, bottom] of a female. (**C to F**) Double FISH in male trunk with lncRNA G20001 and the general parenchyma 1 marker gene *tgfbi*. (**G**) UMAP plot of lncRNA G17920 marker of *dmrt*+ and *eled*+ neoblast clusters. (**H**) WISH with lncRNA G17920 in the head [left, top and bottom] of two males, and in the ovary region [right, top] and vitellaria [right, bottom] of a female. UMAP plots are colored by gene expression (blue = low, red = high) and the scale represents log10(UMIs+1). The regions enclosed by the red dashed lines indicate the location of the relevant parenchyma 1 [A] and *dmrt*+ and *eled*+ neoblast [G] clusters on the UMAP plots. WISH scale bars are 100 μm.

A

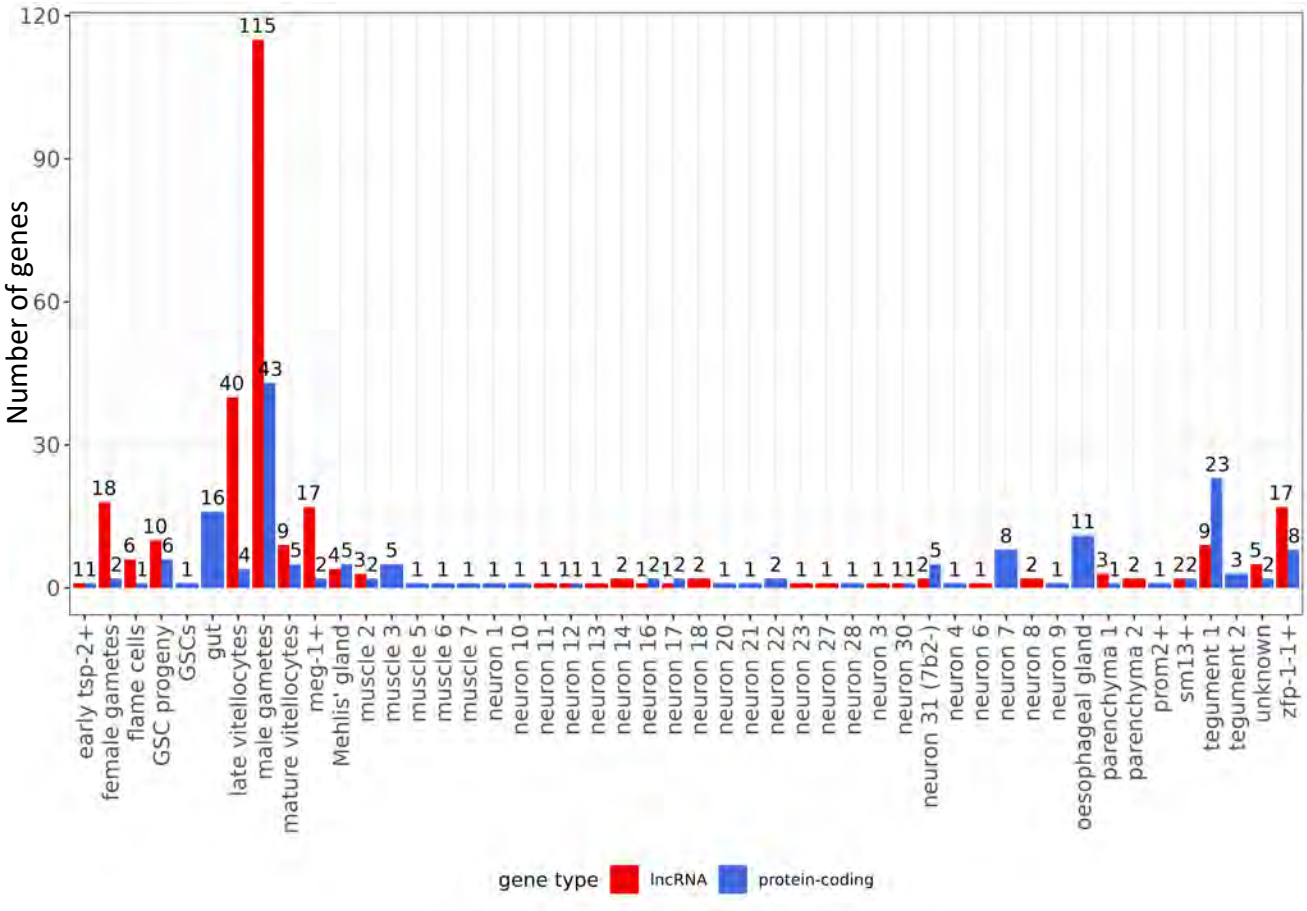

mrt1+ neoblast R = 0.75
early tsp-2+ R = 0.83
early vitellocytes R = 0.68
egc+ R = 0.72
eled+ neoblasts R = 0.79
female gametes R = 0.73
flame cells R = 0.71
GSC progeny R = 0.66
GSCs R = 0.77
gut R = 0.78
late vitellocytes R = 0.66
male gametes R = 0.66
meg-1+ R = 0.68
Mehlis' gland R = 0.74
muscle 1 R = 0.69
muscle 2 R = 0.65
muscle 3 R = 0.62
muscle 4 R = 0.57
muscle 5 R = 0.72
muscle 6 R = 0.74
muscle 7 R = 0.6
neoblast R = 0.83
neoblast progeny R = 0.88
neuron 1 R = 0.85
neuron 10 R = 0.6
neuron 11 R = 0.59
neuron 12 R = 0.74
neuron 13 R = 0.75
neuron 14 R = 0.75
neuron 15 R = 0.58
neuron 16 R = 0.6
neuron 17 R = 0.56
neuron 18 R = 0.65
neuron 2 R = 0.74
neuron 20 R = 0.72
neuron 21 R = 0.63
neuron 22 R = 0.54
neuron 23 R = 0.55
neuron 24 R = 0.69
neuron 25 R = 0.58
neuron 26 R = 0.65
neuron 3 R = 0.62
neuron 30 R = 0.83
neuron 31 (7b2+) R = 0.56
neuron 4 R = 0.77
neuron 5 R = 0.69
neuron 6 R = 0.67
neuron 7 R = 0.55
neuron 8 R = 0.76
neuron 9 R = 0.72
parenchyma 1 R = 0.66
parenchyma 2 R = 0.69
prom2+ R = 0.78
S1 R = 0.83
S1 progeny R = 0.78
sm13+ R = 0.76
tegument 1 R = 0.56
unknown R = 0.58
zfp-1-1+ R = 0.63
mean expression
Percentage of expressing cells

B


mrt1+ neoblast R = 0.52
early tsp-2+ R = 0.62
early vitellocytes R = 0.5
egc+ R = 0.65
eled+ neoblasts R = 0.54
female gametes R = 0.53
flame cells R = 0.55
GSC progeny R = 0.47
GSCs R = 0.51
gut R = 0.6
late vitellocytes R = 0.5
male gametes R = 0.59
meg-1+ R = 0.56
Mehlis' gland R = 0.62
muscle 1 R = 0.61
muscle 2 R = 0.55
muscle 3 R = 0.69
muscle 4 R = 0.56
muscle 5 R = 0.56
muscle 6 R = 0.63
muscle 7 R = 0.57
neoblast R = 0.53
neoblast progeny R = 0.7
neuron 1 R = 0.66
neuron 10 R = 0.7
neuron 11 R = 0.61
neuron 12 R = 0.54
neuron 13 R = 0.61
neuron 14 R = 0.59
neuron 15 R = 0.61
neuron 16 R = 0.66
neuron 17 R = 0.57
neuron 18 R = 0.61
neuron 2 R = 0.67
neuron 20 R = 0.66
neuron 21 R = 0.62
neuron 22 R = 0.59
neuron 23 R = 0.53
neuron 24 R = 0.66
neuron 25 R = 0.58
neuron 26 R = 0.74
neuron 3 R = 0.67
neuron 30 R = 0.7
neuron 31 (7b2+) R = 0.54
neuron 4 R = 0.63
neuron 5 R = 0.67
neuron 6 R = 0.63
neuron 7 R = 0.52
neuron 8 R = 0.58
neuron 9 R = 0.74
parenchyma 1 R = 0.63
parenchyma 2 R = 0.67
prom2+ R = 0.59
S1 R = 0.51
S1 progeny R = 0.48
sm13+ R = 0.58
tegument 1 R = 0.61
unknown R = 0.56
zfp-1-1+ R = 0.52
mean expression
Percentage of expressing cells

Supplementary Figure S5 – Correlation between the expression level of genes and the percentage of cells of a given cluster in which those genes were detected. In (**A**) the expressed lncRNAs are shown, and in (**B**) the protein-coding mRNAs. For each cluster, named at the top of each panel, the mean expression of each gene (y-axis) is plotted as a function of the percentage of cells in the cluster that are detected as expressing that gene (x-axis). The correlation coefficient R is shown below the name of the cluster. Nine clusters with less than 100 cells each were excluded from this analysis.

Figure S6



Supplementary Figure S6 – Number of lncRNAs and protein-coding mRNAs exclusively expressed in at least 1% of the cells in only one cluster. The red bars show the number of lncRNAs exclusively detected as expressed in at least 1% of the cells in the single cluster indicated in the x-axis, and not detected in more than 1% of the cells in any other cluster. For comparison, the blue bars show the number of protein-coding mRNAs with the same cluster expression patterns. Genes expressed in less than 5 cells in a cluster were excluded from this analysis.

2.4. RNAs longos não-codificadores são essenciais para a homeostase e fertilidade do parasita adulto *Schistosoma mansoni* de forma dependente do pareamento

PREÂMBULO

Contribuições do Doutorando Gilbert de Oliveira Silveira para o manuscrito apresentado nesta sessão:

Concebeu o projeto, fez a mineração dos dados da re-análise *in silico* para escolher os lncRNAs candidatos, e desenvolveu os experimentos de bancada. Analisou todos os dados e escreveu o primeiro rascunho do manuscrito. Revisou o manuscrito final.

Detailed Status Information

Manuscript #	03-10-2022-ISRA-eLife-83926
Current Revision #	0
Other Version	825e9c99-746c-4945-81ac-6536d67c6e65 (xPub)
Import Vendor ID	825e9c99-746c-4945-81ac-6536d67c6e65 (eLife/xpub)
Submission Date	3rd Oct 22
Current Stage	Under editorial assessment
Title	Long non-coding RNAs are essential for Schistosoma mansoni pairing-dependent adult worm homeostasis and fertility
Manuscript Type	Initial Submission: Research Article
Corresponding Author	Sergio Verjovski-Almeida (Instituto Butantan)
Contributing Author	N/A
Group Authorship	
Abstract	The trematode parasite Schistosoma mansoni causes schistosomiasis, which affects over 200 million people worldwide. Schistosomes are dioecious, with egg laying depending on the females' obligatory pairing with males. Long non-coding RNAs (lncRNAs) are transcripts longer than 200 nucleotides with low or no protein-coding potential that have been involved in other species with reproduction, stem cell maintenance, and drug resistance. It is totally unknown if lncRNAs are involved with the pairing status of these parasites. Here, we show that lncRNAs are essential to keep the pairing status of adult worms. First, we re-analyzed public RNA-Seq data from adult worms and their gonads, retrieved from hamsters infected with mixed-sex or single-sex cercariae, and found thousands of differentially expressed lncRNAs among the 23 biological samples that were compared. The expression levels of selected lncRNAs were validated by RT-qPCR using an in vitro unpairing model. In addition, in vitro and in vivo silencing of four selected lncRNAs showed that these lncRNAs play key roles on cell proliferation in the adult worms and their gonads and are essential for female vitellaria maintenance, adult worm reproduction, and egg development. Whole mount in situ hybridization experiments showed that these lncRNAs are expressed in tissues that correlate with the phenotypes observed upon in vitro silencing. These results show that lncRNAs are key components of adult worm pairing status in S. mansoni, presenting great potential as new therapeutic target candidates.
Reviewing Editor	Not Assigned
Suggest a Senior Editor	Arturo Casadevall (Johns Hopkins Bloomberg School of Public Health), Dominique Soldati-Favre (University of Geneva)
Suggestions from the Board of Reviewing Editors	Ashish Lal (National Cancer Institute) - BRE, Roberto Bonasio (University of Pennsylvania) - BRE, Phillip Newmark (Morgridge Institute for Research) - BRE
Author Reviewer Suggestions to Include	David Smith
Author Reviewer Suggestions to Exclude	N/A
Major Subject Area(s)	Microbiology and Infectious Disease
Research Organism(s)	
Funding	No Funders
Waiver Request Status	Click here to view if the waiver request was approved or denied
Data Availability	N/A
Ethics	
Dual-Use Research	N/a
Permissions	Have you reproduced or modified any part of an article that has been previously published or submitted to another journal? N/a

Stage	Start Date
Under editorial assessment	3rd Oct 22
Initial quality checks complete	3rd Oct 22
Initial quality checks pending	3rd Oct 22
Manuscript submitted	3rd Oct 22

Long non-coding RNAs are essential for *Schistosoma mansoni* pairing-dependent adult worm homeostasis and fertility

Gilbert O. Silveira[1,2]; Helena S. Coelho[1]; Adriana S. A. Pereira[1,2]; Patrícia A. Miyasato[3]; Daisy W. Santos[1,2]; Lucas F. Maciel[1]; Giovanna G. G. Olberg[1]; Ana C. Tahira[1]; Eliana Nakano[3]; Maria Leonor S. Oliveira[4]; Murilo S. Amaral[1*]; Sergio Verjovski-Almeida[1,2*]

[1]Laboratório de Ciclo Celular, Instituto Butantan, 05503-900 São Paulo, SP, Brazil
[2]Instituto de Química, Universidade de São Paulo, 05508-900 São Paulo, SP, Brazil
[3]Laboratório de Parasitologia, Instituto Butantan, 05503-900 São Paulo, SP, Brazil
[4]Laboratório de Bacteriologia, Instituto Butantan, São Paulo 05503-900, SP, Brazil

*Corresponding author: Murilo S. Amaral – murilo.amaral@butantan.gov.br
*Corresponding author: Sergio Verjovski-Almeida – sergio.verjovski@butantan.gov.br

Abstract (231 words)

The trematode parasite *Schistosoma mansoni* causes schistosomiasis, which affects over 200 million people worldwide. Schistosomes are dioecious, with egg laying depending on the females' obligatory pairing with males. Long non-coding RNAs (lncRNAs) are transcripts longer than 200 nucleotides with low or no protein-coding potential that have been involved in other species with reproduction, stem cell maintenance, and drug resistance. It is totally unknown if lncRNAs are involved with the pairing status of these parasites. Here, we show that lncRNAs are essential to keep the pairing status of adult worms. First, we re-analyzed public RNA-Seq data from adult worms and their gonads, retrieved from hamsters infected with mixed-sex or single-sex cercariae, and found thousands of differentially expressed lncRNAs among the 23 biological samples that were compared. The expression levels of selected lncRNAs were validated by RT-qPCR using an *in vitro* unpairing model. In addition, *in vitro* and *in vivo* silencing of four selected lncRNAs showed that these lncRNAs play key roles on cell proliferation in the adult worms and their gonads and are essential for female vitellaria maintenance, adult worm reproduction, and egg development. Whole mount *in situ* hybridization experiments showed that these lncRNAs are expressed in tissues that correlate with the phenotypes observed upon *in vitro* silencing. These results show that lncRNAs are key components of adult worm pairing status in *S. mansoni*, presenting great potential as new therapeutic target candidates.

Introduction

Schistosomiasis is a neglected tropical disease that affects more than 200 million people worldwide (Colley et al., 2014; Mutapi et al., 2017; Wilson, 2020). No vaccine has been developed so far and controlling the disease is still a challenge, with treatment using a single drug, praziquantel (Bergquist et al., 2017; Wilson et al., 2017). Furthermore, praziquantel-resistant strains have been reported and the drug is effective only against the adult stage of the parasite (Melman et al., 2009; Vale et al., 2017).

Thus, understanding schistosome biology on a molecular level is needed and could suggest new therapeutic alternatives (Gouveia et al., 2018; Hewitson and Maizels, 2014).

Schistosoma mansoni is the species present in the Americas and Africa, with adult worms living in the mesenteric veins of the mammalian host. Schistosomes are the only trematodes that are dioecious, possessing male and female separate sexes (Cort, 1921). Once paired, female egg production is initiated and results in the daily production of 300–3,000 eggs per female, depending on the species (Cheever et al., 1994). Approximately half of the eggs reach the gut lumen (most *Schistosoma* species) or the bladder (*S. haematobium*). The remaining eggs are swept away via the blood system mainly into the liver and spleen, where they penetrate the tissues causing severe inflammation and liver cirrhosis, the main cause of mortality.

Male and female adult worms must stay paired together throughout their life for reproduction to be successful. In fact, female sexual development is governed by male pairing, meaning that females that are not paired to males have immature reproductive status and thus produce no eggs (Galanti et al., 2012; LoVerde and Chen, 1991; Lu et al., 2016; Popiel et al., 1984; Popiel and Basch, 1984). RNA-Seq analyses of adult worms and their gonads retrieved from mammalian hosts infected with mixed-sex or single-sex cercariae have paved the way for understanding the role of protein-coding genes in the maintenance of the pairing status in schistosomes, with the identification of molecular pathways involved in sexual development (Berriman et al., 2009; Grevelding et al., 2018; Lu et al., 2019, 2016). Recently, a male-derived non-ribosomal peptide pheromone that controls female schistosome sexual development and egg laying has been described (Chen et al., 2022). However, the complete set of molecular players that intervene in sexual development are not fully characterized.

Long non-coding RNAs (lncRNAs) are RNAs longer than 200 nucleotides with low or no protein-coding potential that have been implicated in different biological processes in humans and in many other species (Ransohoff et al., 2018). LncRNAs work in various cellular environments and thus can act as regulators of protein-coding gene expression (Jandura and Krause, 2017; Rinn and Chang, 2020), stem cell maintenance (Chen et al., 2020), and drug resistance (Ransohoff et al., 2018). Due to their tissue-specific expression and multifaceted functions, lncRNAs were proposed as novel therapeutic targets in human diseases (Jiang et al., 2019; Nath et al., 2019). In addition, they have been recently suggested as potential targets of antiparasitic therapies, as reviewed by Silveira et al., 2022. In fact, lncRNAs have been associated with epigenetic drug treatment in *S. mansoni* (Amaral et al., 2020), and one selected lincRNA has been shown to interfere with female vitellaria development and adult worm stem cell proliferation (Silveira et al., 2022). Nevertheless, it is unknown whether lncRNAs could be associated with pairing status and fertility in adult worms.

Because lncRNAs have been proposed as possible regulators of many biological processes we hypothesized that lncRNAs could be differentially expressed between sexually immature and mature worms. Here, we report that lncRNAs are differentially expressed between paired and unpaired adult *S. mansoni* worms. We show by *in vitro* and *in vivo* silencing experiments that lncRNAs are essential for maintaining the pairing status, worm viability, female reproductive performance, and adult worm stem cell proliferation. Using whole mount *in situ* hybridization (WISH), we show that selected lncRNAs are expressed in reproductive organs, which correlate with phenotypes associated with their silencing. We propose lncRNAs as new potential therapeutic targets against schistosomiasis since they are key components of adult worm pairing status in *S. mansoni*.

Results

Identification of lncRNAs differentially expressed between immature and mature *S. mansoni* parasites and their gonads

To search for long non-coding RNAs (lncRNAs) differentially expressed between sexually immature and mature *S. mansoni* parasites and their gonads, we re-analyzed RNA-Seq data generated by Lu et al., 2016 who studied parasites retrieved from hamsters infected with either mixed sex cercariae (b, mixed-sex infections) or cercariae of only a single sex (s, single-sex infections). Lu et al. compared eight different biological samples, namely whole males (M) and testes (T) from mixed-sex (bM and bT) or single-sex (sM and sT) infections, and whole females (F) and ovaries (O) from mixed-sex (bF and bO) or single-sex (sF and sO) infections **(Figure 1A)**. Note that we kept the prefix "b" for the mixed sex infections, to be consistent with the nomenclature of the original work (Lu et al., 2016), in which "b" was used for bisex infection.

The original RNA-Seq analysis only considered the protein-coding genes that were expressed among all processed samples (Lu et al., 2016), and our re-analyses identified the expression of 16,583 lncRNAs, which were further classified as 7936 long intergenic non-coding RNA genes (SmLINCs), 7455 long antisense non-coding RNA genes (SmLNCAs), and 1192 long sense non-coding RNA genes (SmLNCSs), in addition to 14,520 protein-coding gene isoforms (Smps). **Supplementary Table S1** shows the list of all expressed lncRNA, and mRNA genes detected in the samples, along with their TPM values obtained in our re-analyses, for each gene at each of the replicate samples.

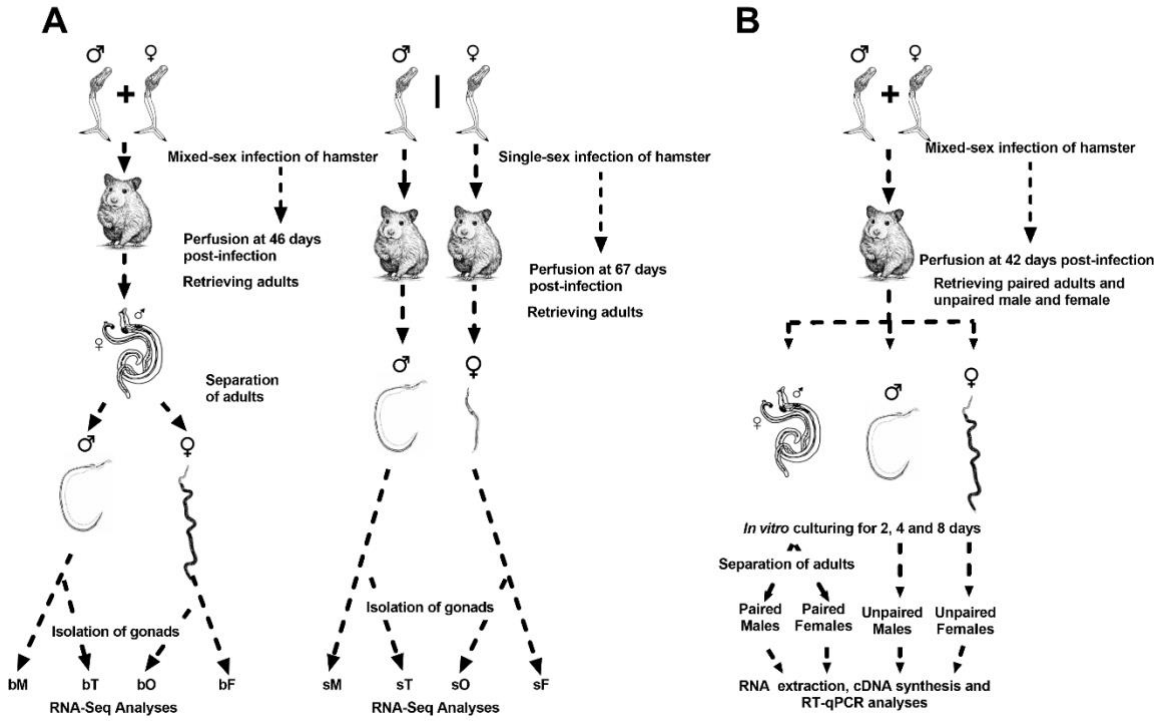


Figure 1: Schematic representation of the pairing assays performed by Lu *et al.* 2016 and in the present work. (A) Lu *et al*. exposed hamsters to a *S. mansoni* mixed-sex cercariae infection (left) or to single-sex infections (middle); 46 days after infection (left) or 67 days after infection (middle), hamsters were perfused, and the adult worms recovered. For the mixed-sex infection, parasites were manually

separated. Gonads were isolated from a fraction of the recovered worms, and RNA was extracted from mixed-sex infection whole males (bM) and testes (bT), from mixed-sex whole females (bF) and ovaries (bO), as well as from single-sex whole males (sM) and testes (sT) and single-sex whole females (sF) and ovaries (sO). Samples were submitted to RNA-Seq; in the present work, we re-analyzed the RNA-Seq data to search for differentially expressed lncRNAs. **(B)** Here, hamsters were exposed to a *S. mansoni* mixed-sex cercariae infection; adult worms were retrieved by perfusion 42 days post-infection. Worms retrieved in the perfusion as paired couples were collected separate from adult males and females that were retrieved as naturally separated worms. Worm pairs or separated worms were cultured *in vitro* for 2, 4 and 8 days in ABC media; separated worms are known to experience regression of the reproductive organs. At the end of the incubation times, the paired worm couples were manually separated. Male and female worms that were either paired or unpaired during *in vitro* culturing were submitted to RNA extraction, and RT-qPCR measurements were performed.

Upon comparing the expression levels of genes among the different worm samples, we found 3681 unique differentially expressed (DE) lncRNA genes, i.e. lncRNAs found as DE in at least one of the comparisons among the 23 analyzed samples **(Figure 2A)**, out of the 16,583 lncRNAs detected as expressed in these samples (22% DE lncRNAs). In addition, 11,109 unique protein-coding gene isoforms (out of the 14,520 isoforms, i.e. 77%) were detected as differentially expressed **(Figure 2B)**.

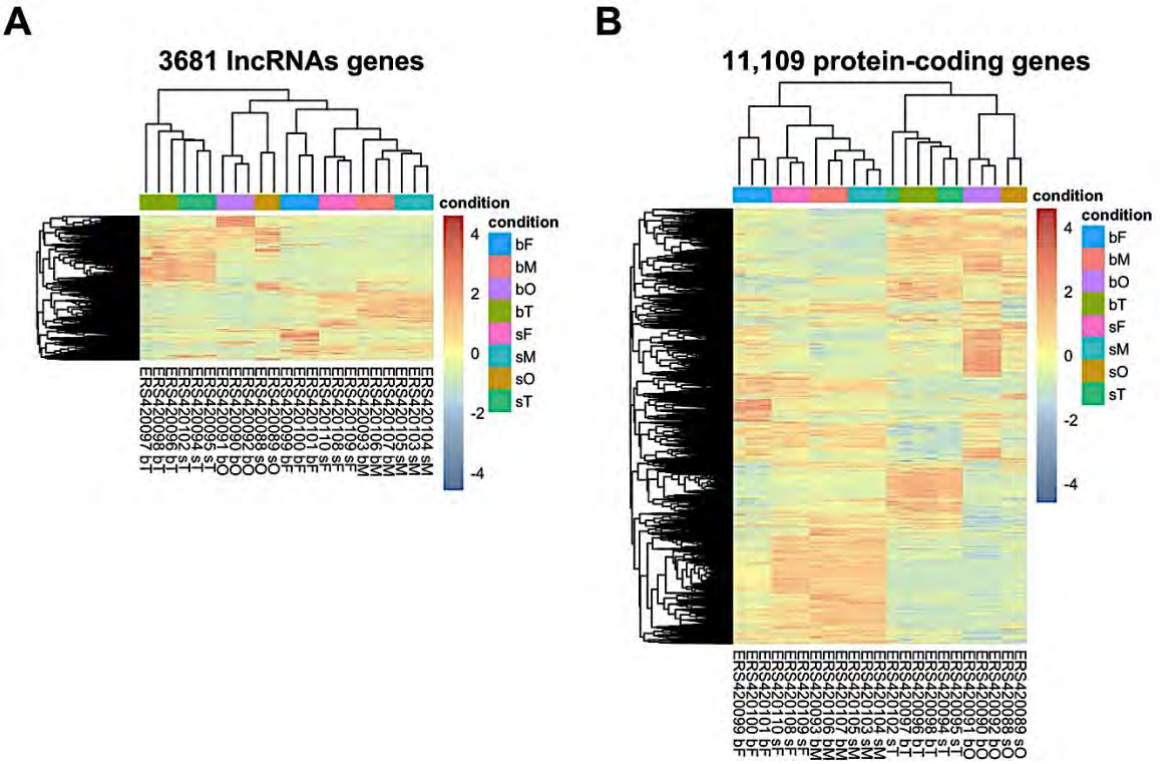


Figure 2: Heatmap of *Schistosoma mansoni* long non-coding RNAs (lncRNAs) or protein-coding genes differentially expressed (DE) between parasites from mixed-sex and single-sex infections. Non-supervised hierarchical clustering of (**A**) 3681 DE lncRNA genes (lines) or (**B**) 11,109 DE protein-coding gene isoforms (lines) in each of the biological replicates (columns) of *S. mansoni* parasites retrieved from mixed-sex (b) or single-sex (s) infections, as indicated in the sample labels at the bottom of the heatmaps. These results were obtained through a re-analysis of the RNA-Seq data from Lu *et al.,* 2016, now using as reference the lncRNA transcriptome previously published by Maciel *et al.,* 2019. Statistically significant DE genes were determined with DESeq2 (FDR < 0.05). Gene expression levels are shown as Z-scores, which are the number of standard deviations below (blue lines, downregulated) or above (red lines, up-regulated) the mean expression value of each gene among all samples, as indicated by the color scale on the right. Females from mixed-sex or single-sex infections are identified by bF or sF; males from mixed-sex or single-sex infections by bM or sM; ovaries from mixed-sex or single-sex infection females by bO or sO; testes from mixed-sex or single-sex infection males by bT or sT.

158

159 **Selection of lincRNAs to be tested for their involvement in adult worm pairing**

160 To identify a subset of lincRNA candidates to be tested for their possible
161 involvement in adult worm pairing, we devised a filtering pipeline (**Supplementary**
162 **Methods**) to single-out the lincRNAs that were up-regulated in mixed-sex females,
163 males, and their reproductive organs compared to the single-sex counterparts. We also
164 prioritized lincRNAs that were expressed at values higher than 2 transcripts per million
165 (TPM), and that had no evidence of alternatively spliced isoforms when analyzing the
166 full transcriptome. The final set of selected lincRNAs comprises 12 transcripts, being
167 eight lincRNAs enriched in females from mixed-sex infections (bF), one lincRNA in
168 males from mixed-sex infections (bM), one lincRNA in ovaries of females from mixed-
169 sex infections (bO), and two lincRNAs in testes of males from mixed-sex infections
170 (bT) (see **Supplementary Methods**).

171 **Validation by RT-qPCR of differential expression of the selected lincRNAs upon**
172 **unpairing**

173 All 12 selected lincRNAs were evaluated by RT-qPCR and 8 of them were
174 differentially expressed in our *in vitro* unpairing mimetic model (**Figure 1B**, see also
175 Methods), showing an expression pattern **(Figures 3 and 4)** similar to that found in our
176 re-analysis of the RNA-Seq data in the worms from the *in vivo* single-sex infections of
177 Lu et al., 2016. Specifically, SmLINC142881, SmLINC124324, SmLINC175062, and
178 SmLINC110998 were found in the RNA-Seq re-analysis to be more expressed in
179 females from mixed-sex infection (bF) when compared with those from single-sex
180 infection (sF) **(Figure 3A to 3D, plaid orange)**, and they were all validated by RT-
181 qPCR in the *in vitro* mimetic model **(Figure 3A to 3D, orange)**. Of note,
182 SmLINC142881 showed the highest expression levels in bF **(**TPM average of 25,
183 **Figure 3A, plaid orange)**, while SmLINC110998 showed the largest fold-change
184 difference between samples from paired and unpaired females cultured *in vitro* for 8
185 days (40X) **(Figure 3D, orange)**.
186

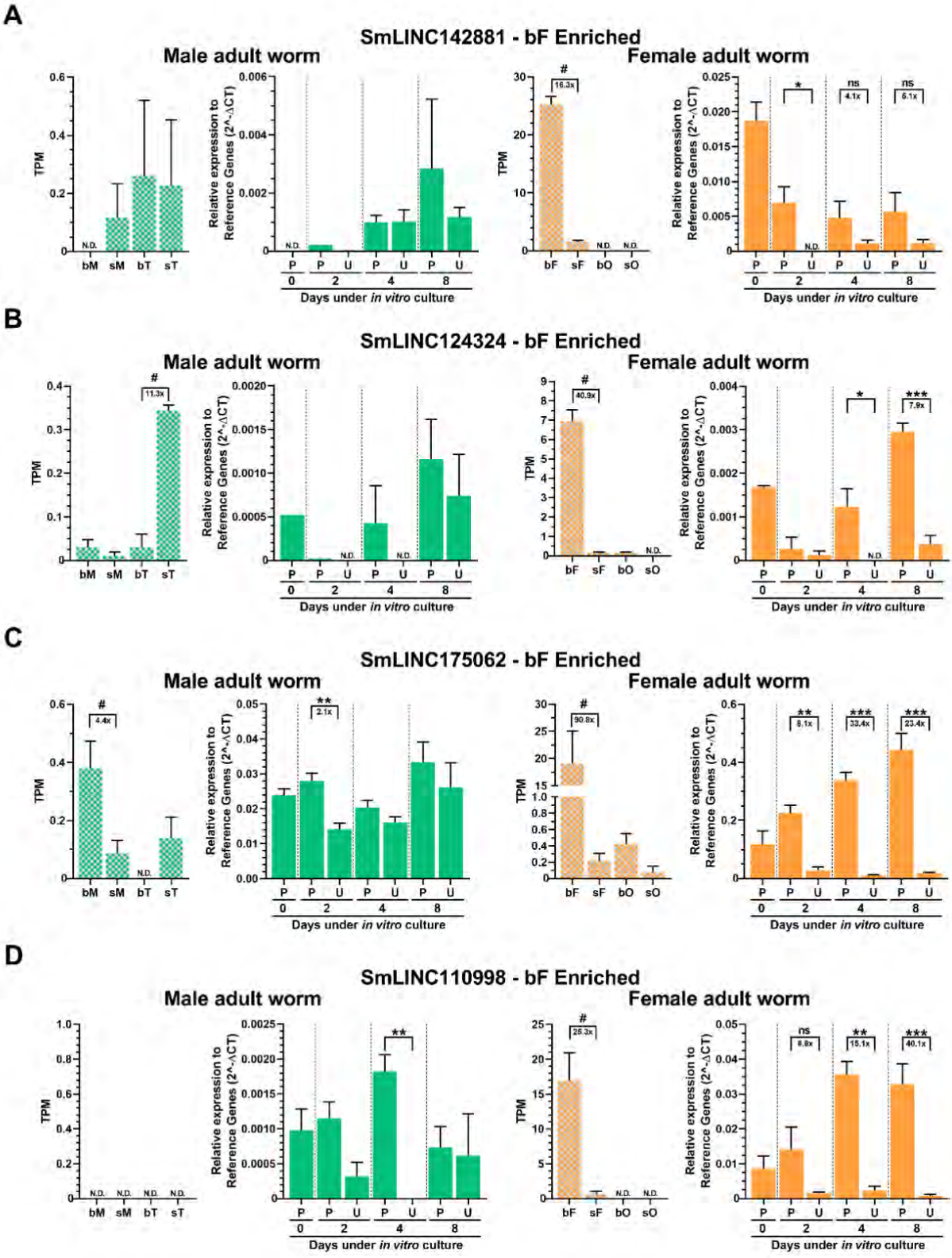


Figure 3: Expression of lincRNAs enriched in females from mixed-sex infections, and differential expression validation by RT-qPCR in *S. mansoni* cultured *in vitro* for 2, 4 or 8 days as paired couples or unpaired males and females. Four lincRNAs detected as enriched in females from mixed-sex infections (bF) in the re-analyses of the RNA-Seq dataset of Lu et al., 2016, were selected for RT-qPCR assays, namely (**A**) SmLINC142881, (**B**) SmLINC124324, (**C**) SmLINC175062, and (**D**) SmLINC110998. Male related results are shown on the left (green) and female results on the right (orange). Paired couples (P) or

unpaired (U) parasites were obtained by perfusion of hamsters infected for 42 days with *S. mansoni* cercariae. After perfusion, males and females were cultured *in vitro* for 2, 4 or 8 days as paired (P) couples or unpaired (U) worms. RT-qPCR results (solid-colored graphs) are normalized to the geometric mean of reference genes Smp_099690 and Smp_023150. Expression values from 4 different biological replicates are shown. Standard error of the mean (SEM) is shown in the error bars. (*) = $p < 0.05$; (**) = $p < 0.01$; (***) = $p < 0.001$, Student t test. N.D.: Not detected. ns: p-value > 0.05. For comparison, RNA-Seq data from the re-analysis of Lu *et* al., 2016 is shown (plaid-colored graphs) and the expression is measured in TPM (transcripts per million); RNA-Seq data is retrieved from males (M), females (F), testes (T) or ovaries (O) from either a mixed-sex (b) or a single-sex (s) infection; (#) = FDR<0.005. The fold-change differences between the compared groups are represented under the brackets.

SmLINC133371 had been identified as an sT, sF and sO enriched lincRNA in the RNA-Seq re-analyses **(Figure 4A, plaid green and orange)**, and a higher expression was confirmed by RT-qPCR in unpaired (U) females compared with paired (P) ones in the *in vitro* mimetic model **(Figure 4A, orange)**.

SmLINC154295, a lincRNA that was found enriched in bM, bF, and bO in the RNA-Seq data **(Figure 4B, plaid green and orange)**, was validated with RT-qPCR in the *in vitro* mimetic model as more highly expressed in paired (P) males or females compared with unpaired (U) ones, especially at longer *in vitro* culturing times **(Figure 4B, green and orange)**.

SmLINC101519 is a lincRNA that presented a peculiar expression profile in the RNA-Seq dataset: its expression was higher in males from mixed-sex infections (bM) when compared with single-sex males (sM) **(Figure 4C, plaid green)**, while an opposite pattern with higher expression in single-sex females (sF) than in females from mixed-sex infections (bF) was observed **(Figure 4C, plaid orange)**. This profile was confirmed by RT-qPCR in the *in vitro* mimetic model, with males cultured as paired couples (P) having higher expression than unpaired males (U) **(Figure 4C, green)**, whereas females cultured as unpaired worms (U) showed a higher expression than paired females (P), especially at longer culturing periods (4 and 8 days) **(Figure 4C, orange)**.

Interestingly, SmLINC141426, which had been detected *in vivo* as enriched in testes of males from mixed-sex infections (bT) compared to testes from single-sex males (sT) **(Figure 4D, plaid green)**, was validated by RT-qPCR in the *in vitro* mimetic model with whole worms as more highly expressed in paired (P) males when compared to unpaired (U) males **(Figure 4D, green)**.

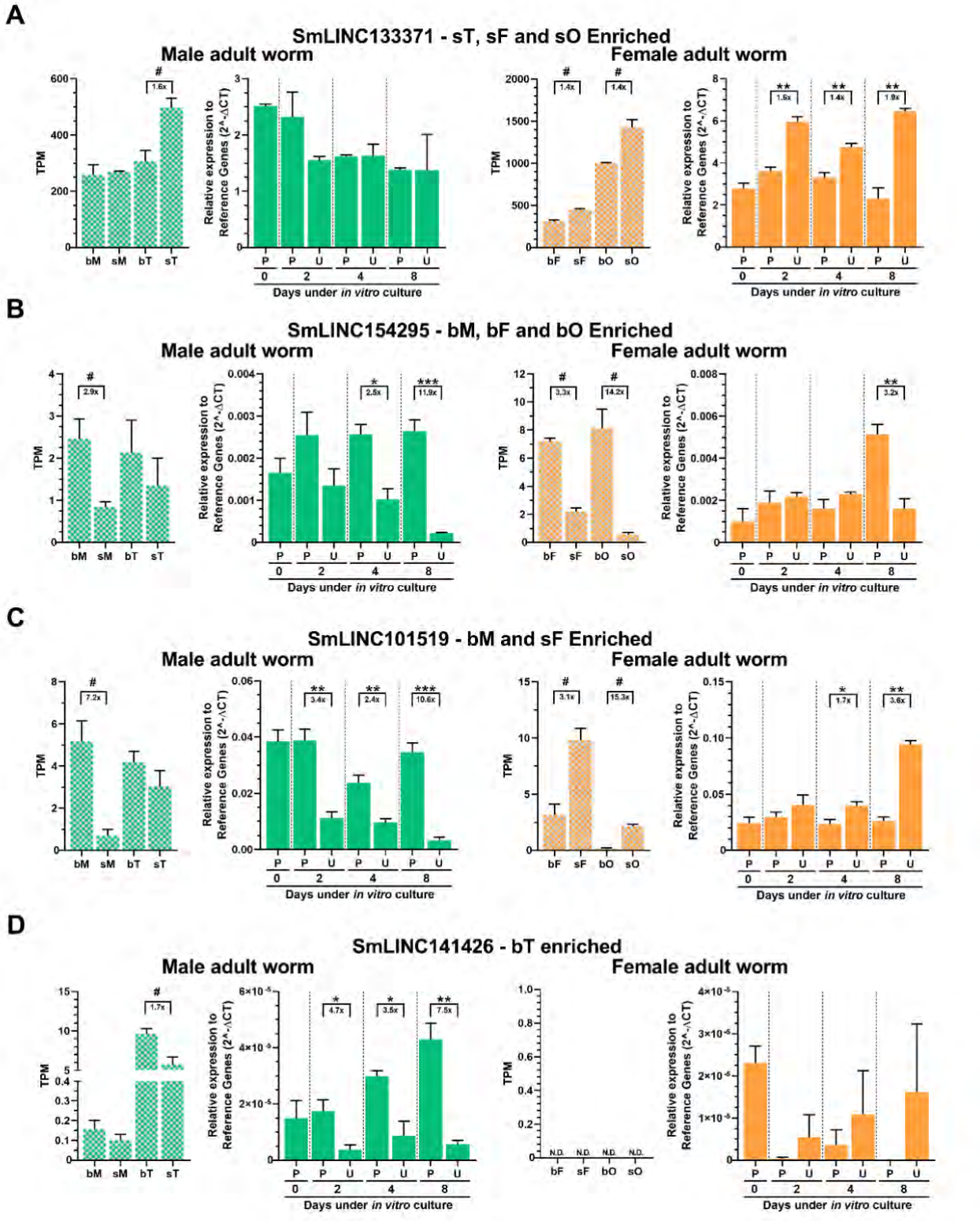


Figure 4: Expression of lincRNAs enriched in samples other than females from mixed-sex infections and differential expression validation by RT-qPCR in *S. mansoni* cultured *in vitro* for 2, 4 or 8 days as paired couples or unpaired males and females. Four lincRNAs detected as enriched in samples other than females from mixed-sex infections in the re-analyses of the RNA-Seq dataset of Lu et al., 2016, were selected for RT-qPCR assays, namely (**A**) SmLINC133371, enriched in sT, sF and sO; (**B**) SmLINC154295, enriched in bM, bF and sO; (**C**) SmLINC101519, enriched in bM and sF; and (**D**)

SmLINC141426, enriched in bT. Male related results are shown on the left (green) and female results on the right (orange). Paired couples (P) or unpaired (U) parasites were obtained by perfusion of hamsters infected for 42 days with *S. mansoni* cercariae. After perfusion, males and females were cultured *in vitro* for 2, 4 or 8 days as paired (P) couples or unpaired (U) worms. RT-qPCR results (solid-colored graphs) are normalized to the geometric mean of reference genes Smp_099690 and Smp_023150. Expression values from 4 different biological replicates are shown. Standard error of the mean (SEM) is shown in the error bars. (*) = $p < 0.05$; (**) = $p < 0.01$; (***) = $p < 0.001$, Student t test. N.D.: Not detected. ns: p-value > 0.05. For comparison, RNA-Seq data from the re-analysis of Lu *et* al., 2016 is shown (plaid-colored graphs) and the expression is measured in TPM (transcripts per million); RNA-Seq data is retrieved from males (M), females (F), testes (T) or ovaries (O) from either a single-sex (s) or mixed-sex (b) infection; (#) = FDR<0.005. The fold-change differences between the compared groups are represented under the brackets.

Four other lincRNAs have been measured by RT-qPCR and their expression profiles were not validated when comparing the *in vitro* mimetic model and the *in vivo* mixed-sex/single-sex infections. These lincRNAs are enriched in bF, bO, and sM samples in the RNA-Seq dataset **(Figure S1)**.

Taken together, these results show that a set of selected lincRNAs differentially expressed *in vivo* between worms from mixed-sex and single-sex infections are also differentially expressed upon unpairing *in vitro*, indicating potential involvement in *S. mansoni* sexual pairing and development.

In vitro silencing of pairing-dependent lincRNAs decreases worm viability, worm pairing, oviposition and egg hatching

To assess the relevance of lincRNAs in worms pairing homeostasis, we performed *in vitro* silencing by soaking adult *S. mansoni* couples for eight days with double-stranded RNAs (dsRNAs) that targeted each of the lincRNAs of interest. We were able to design and synthesize dsRNAs for 4 out of the 8 pairing-dependent lincRNAs confirmed as differentially expressed in the *in vitro* mimetic model, which target SmLINC101519, SmLINC110998, SmLINC124324, or SmLINC175062; dsRNA probes were 610, 294, 326, or 296 bases long, respectively **(Supplementary Table S2)**.

Adults were incubated *in vitro* with each of the 4 dsRNA probes targeting the pairing-dependent lincRNAs, with a control dsmCherry (a dsRNA that will activate the RNAi pathway but will not target any parasite gene), or with no dsRNA and were followed for up to eight days. RT-qPCR assays confirmed the effective knockdown of each of the 4 lincRNAs both in males and females (**Figure 5A**), with reductions in the lincRNAs expression levels of 70–75%, 85–90%, 61–64% or 68–80% for SmLINC101519, SmLINC110998, SmLINC124324, or SmLINC175062, respectively (**Figure 5A**).

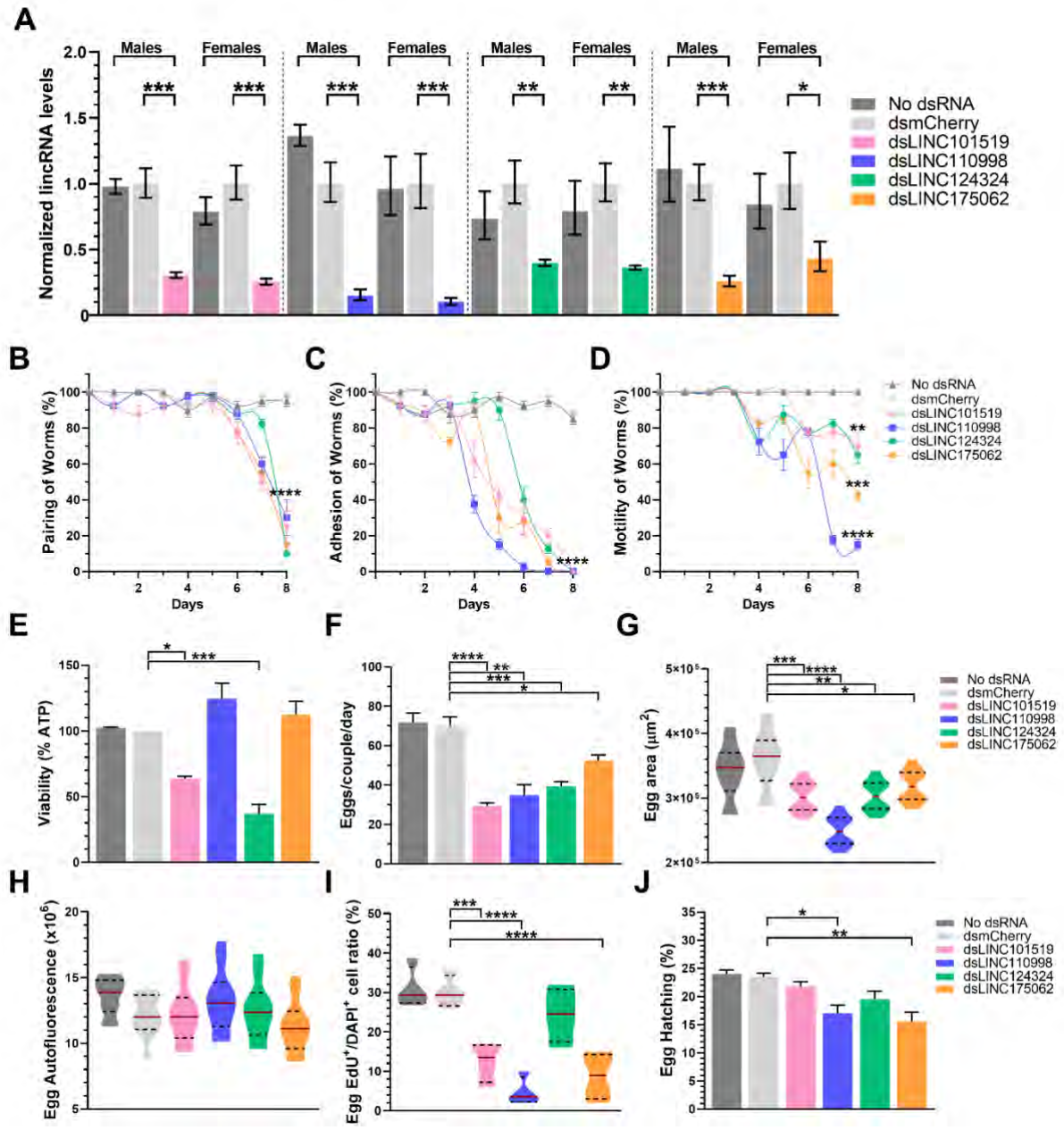


Figure 5: Phenotypic changes in *Schistosoma mansoni* adult worm couples upon *in vitro* silencing (RNAi) of each of four pairing-dependent lincRNAs. Paired couples were obtained by perfusion of hamsters infected for 42 days with *S. mansoni* cercariae. Couples were cultured *in vitro* for 8 days, in ABC media supplemented with 30 µg/mL of dsRNA targeting each of the lincRNAs indicated by the different colors, namely SmLINC101519 (pink), SmLINC110998 (blue), SmLINC124324 (green), and SmLINC175062 (orange). Medium was exchanged every other day while dsRNA was added every day. dsRNA targeting mCherry (a gene that is not present in *S. mansoni*) was assayed in parallel as a negative control (light gray). Results for parasites cultured with no dsRNA are also shown (dark gray). **(A)** RT-qPCR results for each lincRNA expression level are normalized to the geometric mean of reference genes Smp_099690 and Smp_023150. **(B-D)** Pairing status, adhesion to the plate and motility of worm couples were traced along the 8 days of the experiment. **(E)** Viability of adult worms (males+females) was monitored using the ATP-Glo Assay. **(F)** At the end of the experiment (8 days), eggs were collected and counted. **(G-J)** Collected eggs were monitored for their size (area) **(G)**, integrity of their eggshell (autofluorescence) **(H)**, proliferation status of the embryos (Egg Edu+/DAPI+ cell ratio) **(I),** and the percentage of egg hatching was measured by keeping the eggs in culture for another 7 days in ABC media for synchronization of their development, then assessing egg hatching as described in Methods, with the

percentage of hatched eggs being shown **(J)**. Violin plot representation at figures **(G-I)** with the median indicated by the red line and the quartiles represented by the dashed lines. Results from 4 different biological replicates are shown. Standard error of the mean (SEM) is shown in the error bars. (*) = $p < 0.05$; (**) = $p < 0.01$; (***) = $p < 0.001$; (****) = $p < 0.0001$, Student t test or One-Way ANOVA with multiple comparisons to dsmCherry group.

Silencing of each lincRNA caused a reduction in the pairing status of adult worms starting at day 7 and reaching a significant reduction to 10–30% pairing after 8 days in culture **(Figure 5B).** An earlier impact was observed in the adhesion of adult worms to the culture plate, starting at days 4 to 6, with a complete lack of adhesion observed at 6 to 8 days of *in vitro* culture for all silenced lincRNAs **(Figure 5C)**. Upon lincRNA knockdown, motility of adult worm couples and of unpaired worms was affected at 4 to 6 days in culture, with significant reductions in motility at 8 days of silencing of 30%, 32%, 50%, and 80% for SmLINC101519, SmLINC124324, SmLINC175062, or SmLINC110998, respectively **(Figure 5D)**. The viability of adult worm couples (male and females together) was assessed after 8 days of *in vitro* lincRNAs silencing by measuring the total ATP levels of parasites, and a significant 35% or 64% reduction in the viability of worms silenced for SmLINC101519 or for SmLINC124324 was observed, respectively **(Figure 5E)**.

Reduction in adult worm fitness reflected on egg laying and egg health: egg laying was significantly reduced by 58%, 50%, 43%, or 26% in the adult worm pairs treated for 8 days with dsRNAs targeting SmLINC101519, SmLINC110998, SmLINC124324, or SmLINC175062, respectively (**Figure 5F**). Egg size was reduced upon dsRNA treatment for all lincRNAs, with a higher egg area reduction seen in eggs from worms silenced for SmLINC110998 (**Figure 5G**). Still, no difference was seen in eggshell integrity, measured by the egg autofluorescence, in any of the lincRNAs silencing conditions (**Figure 5H**). Interestingly, egg proliferation was measured by the ratio of EdU^+ to $DAPI^+$ cell ratio (EdU, a thymidine analog, 5-ethynyl-2′-deoxyuridine), and a significant decrease in proliferation was observed with SmLINC101519, SmLINC110998, or SmLINC175062 silencing (**Figure 5I**). A significant reduction in egg hatching was only observed in eggs from worms treated with dsRNAs targeting SmLINC110998 or SmLINC175062 (**Figure 5J**), which indicates that silencing of different lincRNAs affected distinct pathways involved with egg development and maturation.

An unrelated control lincRNA, not detected as enriched in the mixed-sex/single-sex model comparison, namely SmLINC130991 was silenced *in vitro* for 8 days. Silencing was effective, and a 50–70% reduction in the level of SmLINC130991 was obtained (**Figure S2B**). Yet, none of the phenotypic changes that were observed with silencing of the four selected pairing-dependent lincRNAs was present upon silencing of SmLINC130991 (**Figure S2C to S2K**).

Cell proliferation status and vitellaria composition of adult worms are affected by the i*n vitro* silencing of pairing-dependent lincRNAs

We also investigated the proliferation status of adult worms (**Figure 6A**) and their gonads (**Figure 6B**) upon silencing of the pairing-dependent lincRNAs. This was achieved by pulse chasing with EdU for 24 hours. While mCherry dsRNA treatment did not affect the adult worm nor their gonads proliferation, SmLINC101519 silencing impacted the adult male body cells proliferation status (**Figure 6A, left**), but neither the female body cells (**Figure 6A, right**) nor the male and female gonads were affected (**Figure 6B**). SmLINC110998 silencing markedly reduced the female ovary cells

proliferation (**Figure 6B, right**) and reduced female body cells proliferation (**Figure 6A, right**). Of note, SmLINC124324 silencing showed a similar pattern of reduction in female body cells proliferation (**Figure 6A, right**) and a less pronounced effect in ovary cells proliferation (**Figure 6B, right**). As for silencing of SmLINC175062, we have only observed a decrease in male adult worm body cells proliferation (**Figure 6A, left**), whilst female adult worms (**Figure 6A, right**) and the male and female gonads proliferation were not affected (**Figure 6B**).

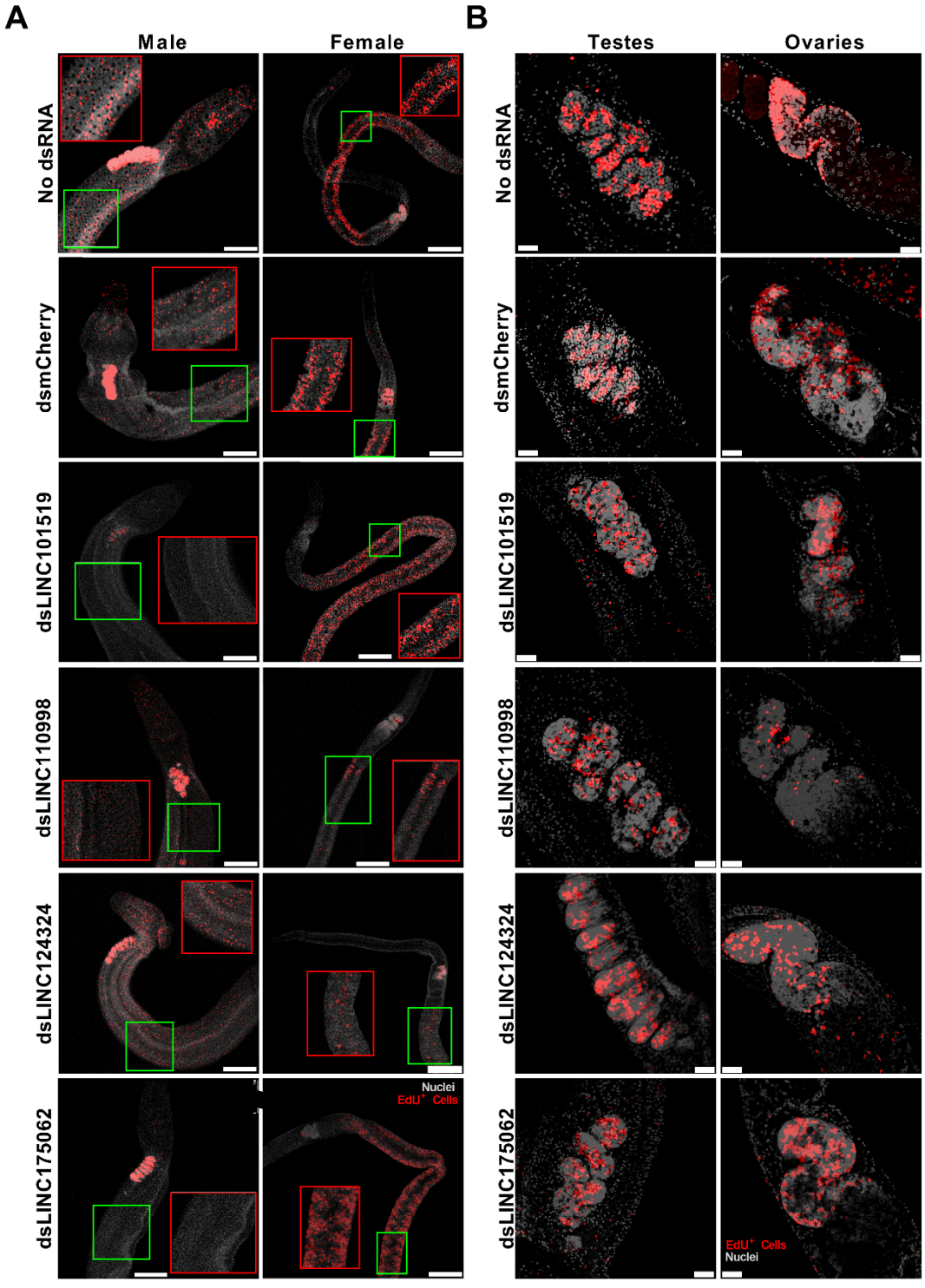


Figure 6: *In vitro* silencing of pairing-dependent lincRNAs in *S. mansoni* adult worm couples leads to an impaired proliferation status. Paired couples were obtained by perfusion of hamsters infected for 42 days with *S. mansoni* cercariae. Couples were cultured *in vitro* for 8 days in ABC media supplemented with 30 µg/mL of dsRNA targeting each of the indicated lincRNAs, namely SmLINC101519, SmLINC110998, SmLINC124324, and SmLINC175062. Medium was exchanged every other day while dsRNA was added every day. A control dsRNA targeting mCherry (a gene that is not present in *S. mansoni*) was assayed in parallel as a negative control. Results for parasites cultured in parallel with no dsRNA are also reported. Cell proliferation was assayed by labeling with thymidine analog 5-ethynyl-2′-deoxyuridine (EdU), which was added to the cultures on the 7th day of culture at a final concentration of 10 µM and incubating for 24 hours. EdU detection was performed as previously described (Collins et al. 2013). DAPI stained cells nuclei are in gray and EdU+ cells (proliferating cells) are stained in red. Scale bars: 250 µm for the adult worm images **(A)**, and 25 µm for the adult worm gonad images **(B)**. Representative images from 3 experiments with n > 10 parasites. The red borders define zoomed-in insets of interest that correspond to the regions within green borders.

Female vitellaria composition was evaluated by double staining of vitelline and lipid droplets within the vitellaria, as previously described (Wang et al., 2019). Upon SmLINC110998 silencing the female vitellaria was impacted mainly at the vitelline droplets (pink staining) **(Figure 7)**, corroborating with the egg-laying reduction **(Figure 5F)**. SmLINC124324 dsRNA treatment of worms has also impaired female vitellaria organization, with most of vitelline droplets staining being localized near the borders of the female body while the lipid droplets were seen near the vitelline duct, and no double staining was seen. This is also corroborating with egg laying results that showed a 48% decrease in this phenotype upon silencing of SmLINC124324 **(Figure 5F)**. Of note, these assays demonstrated that dsmCherry control treatment did not disturb the vitellaria of adult worm females. Taken together, these results show that lincRNAs are important for maintenance of adult worm pairing, viability, and fertility *in vitro*.

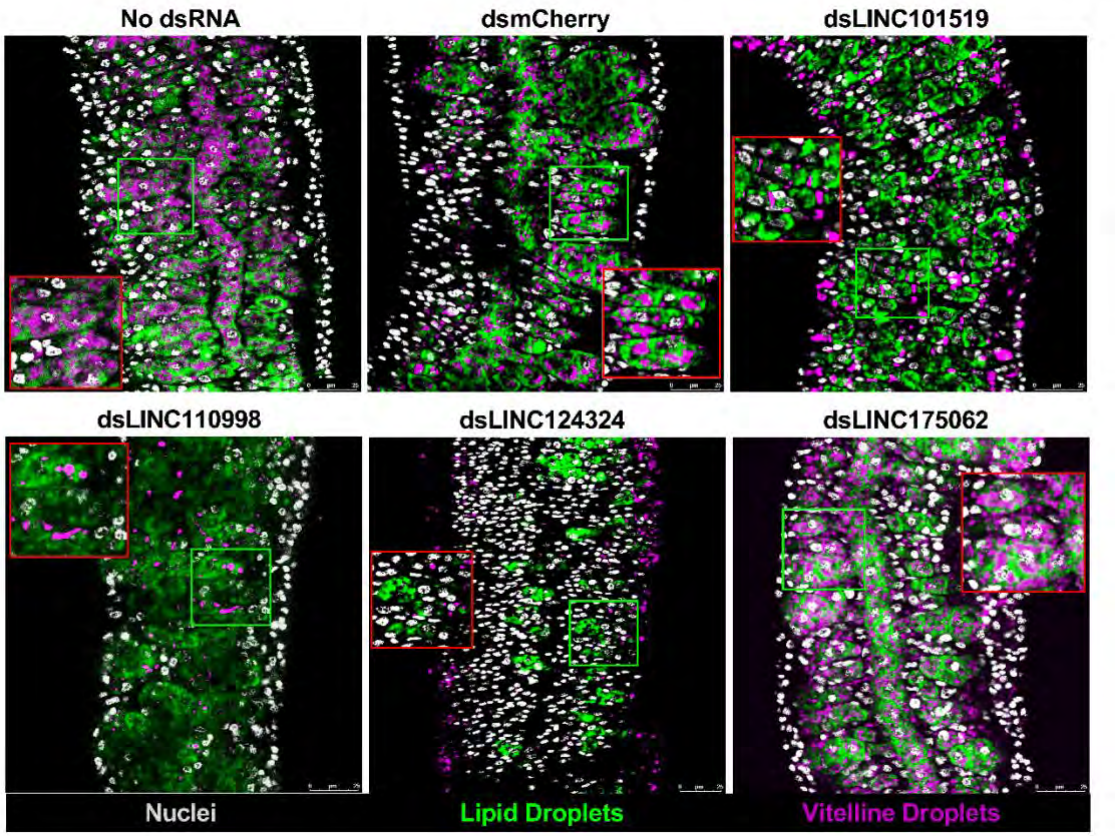


Figure 7: *In vitro* silencing of pairing-dependent lincRNAs in *S. mansoni* adult worm couples causes female vitellaria impairment. Paired couples were obtained by perfusion of hamsters infected for 42

days with *S. mansoni* cercariae. Couples were cultured *in vitro* for 8 days in ABC media supplemented with 30 µg/mL of dsRNA targeting each of the indicated lincRNAs, namely SmLINC101519, SmLINC110998, SmLINC124324, and SmLINC175062. Medium was exchanged every other day while dsRNA was added every day. dsRNA targeting mCherry (a gene that is not present in *S. mansoni*) was assayed in parallel as a negative control. Results for parasites cultured with no dsRNA are also reported. Female vitellaria were stained with Fast Blue BB (pink) and BODIPY (green), which labeled vitelline and lipid droplets in the vitellaria, respectively. DAPI staining of cells nuclei is shown in gray. Scale bars: 25 µm. Representative images from 3 experiments with n > 10 parasites. The red borders define zoomed-in insets of interest that correspond to the regions within green borders.

Silencing of the unrelated control SmLINC130991 did not affect the EdU labelling of male or female body cells, of ovaries or testes (**Figure S2L**), or the labelling of lipid and vitelline droplets (**Figure S2M**).

LincRNAs involved in pairing are expressed in worm reproductive tissues

To gain further insight into the possible functions of the selected pairing-dependent lincRNAs, we performed whole mount *in situ* hybridization (WISH) with adult female and male worms for six of the eight lincRNAs validated by RT-qPCR. SmLINC101519 was expressed in the female uterus, vitellaria and around the female ovary, while in males, SmLINC101519 was present in testes, oesophagus and along the head and trunk **(Figure 8A, right)**; curiously, this was not completely consistent with the results from the single-cell RNA-Seq (scRNA-Seq) analysis (Morales-Vicente et al., 2022), in which SmLINC101519 was detected in parenchyma, muscle and neuron cell clusters **(Figure 8A, left, cells clusters marked with 1, 2 and 3)**.

SmLINC110998 showed no relevant WISH signal in males, while in females the expression was vastly found in mature vitellocytes in both the vitellaria and the vitelline duct **(Figure 8B, right**), which was consistent with the scRNA-Seq analysis in which SmLINC110998 was detected in the late vitellocytes cluster **(Figure 8B, left, cells cluster 4**).

SmLINC124324 was detected by WISH with a strong expression signal in the female vitellaria **(Figure 8C, right),** and this agrees with the scRNA-Seq analysis which found SmLINC124324 expressed in late vitellocytes **(Figure 8C, left, cells cluster 4**). Labelling was detected as well in the ovary and spread along the whole body. Interestingly, SmLINC124324 expression in males was detected by WISH also in the margins of worms' head, body, and trunk **(Figure 8C, right)**.

Finally, SmLINC175062 *in situ* hybridization did not detect expression in male worms, while in female worms the expression appeared to localize in mature vitellocytes **(Figure 8D, right)**; SmLINC175062 was poorly detected by scRNA-Seq, with little to no expression in any of the cell clusters **(Figure 8D, left)**.

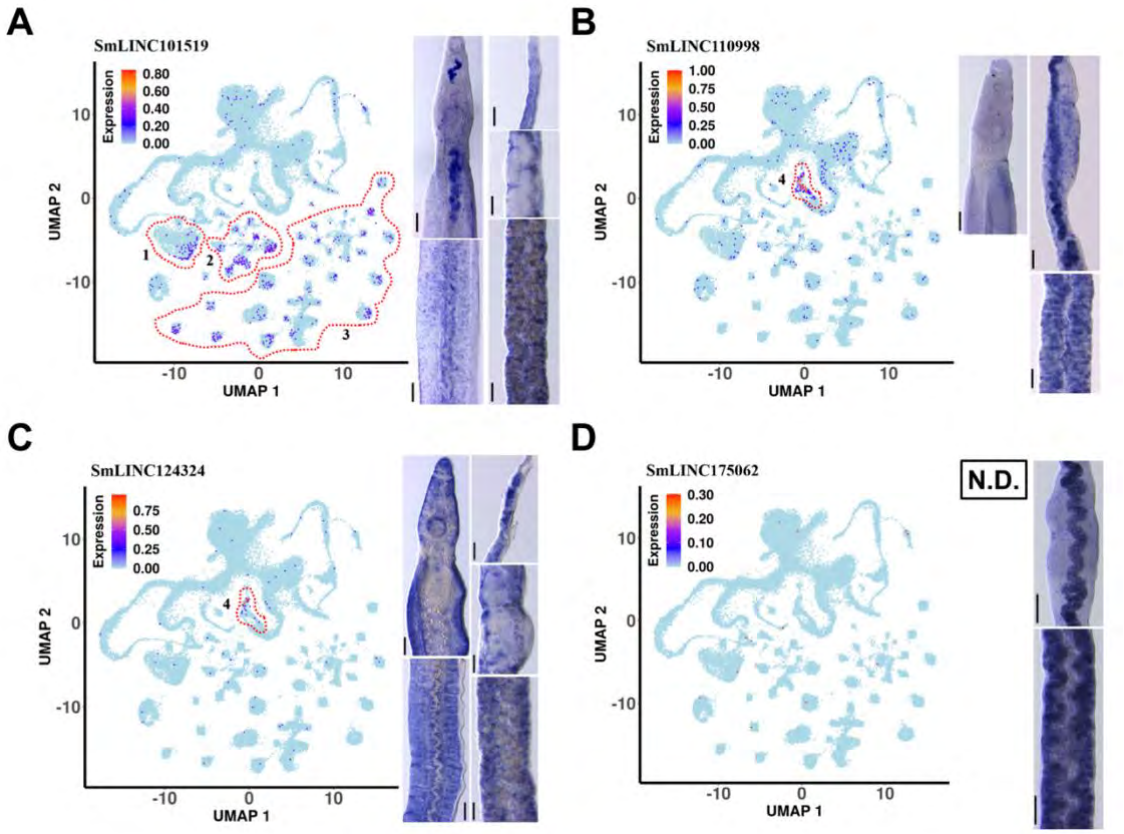


Figure 8: Localization of the selected pairing-dependent lincRNAs in adult worm tissues by whole mount *in situ* hybridization. Whole mount *in situ* hybridization (WISH) of each lincRNA is shown as the blue color stains in the male (left) or female (right) adult worm images of the heads and bodies. Scale bars are 100 µm. Results for **(A)** SmLINC101519, **(B)** SmLINC110998, **(C)** SmLINC124324, and **(D)** SmLINC175062. For comparison purposes, single-cell RNA-Seq data from Morales-Vicente et al., 2022 was retrieved (http://verjolab.usp.br:8081/). LincRNA expression patterns across the single-cell clusters are shown with UMAP plots, colored by gene expression levels (blue = low, red = high), and the scale represents log10(UMIs+1). The regions enclosed by red dashed lines indicate the relevant cell clusters: Region 1, parenchyma 1; Region 2, muscle cell clusters; Region 3, neuron cell clusters; Region 4, late vitellocytes cell cluster. N.D.: Not Detected.

Two additional pairing-dependent lincRNAs, which were not assayed for silencing, were localized by WISH in the reproductive tissues of the parasite; **Figure S3A** shows that SmLINC141426 was detected by WISH in the male head, testes, and trunk, whereas in the female, signals were detected along the head, ovary, and mature vitellocytes; with scRNA-Seq, SmLINC141426 was detected in the male gametes cell cluster (**Figure S3A, left, cluster marked as 1**). **Figure S3B** shows that SmLINC154295 was not detected by WISH in the male, while in the female, it was detected in the vitellaria and vitelline duct **(Figure S3B)**; with scRNA-Seq, SmLINC154295 was not detected in any cell cluster **(Figure S3B, left)**.

Finally, the unrelated control SmLINC130991 was assessed by WISH (**Figure S2N**). Weak signals were observed in the female ovary and vitellarium as well as in the male testes and dispersed in the trunk (**Figure S2N**). SmLINC130991 was detected by scRNA-Seq at very low levels of expression in a few cells, not being enriched in any cell cluster (**Figure S2N, left**).

In vivo silencing of pairing-dependent lincRNAs reduces worm survival

We further checked if the *in vivo* knockdown of pairing-dependent lincRNAs would cause phenotypic effects on the parasites, as was observed *in vitro*. *In vivo* silencing approaches have been used with *Schistosoma* in the past, using small interfering RNAs (Pereira et al., 2008) or dsRNAs (Li et al., 2018). For dsRNAs, Li et al. tested the intravenous injection in mice of 10–30 μg dsRNA per dose, and recommended at least four doses for the efficient silencing of *S. japonicum* protein-coding genes (Li et al., 2018).

Based on the above information we decided to *in vivo* silence the pairing-dependent lincRNAs by injecting *S. mansoni*-infected mice at 7, 21, 35, and 42 dpi with 30 μg dsRNA per dose, targeting each of the four lincRNAs. On the 49th dpi, mice were perfused, and worms were recovered. RT-qPCR assays confirmed the effective knockdown of the lincRNAs in the recovered worms, with reductions in the lincRNAs expression levels of 59–77%, 60–64%, 38–58%, or 70–78% for SmLINC101519, SmLINC110998, SmLINC124324, or SmLINC175062, respectively (**Figure 9A)**. Furthermore, the *in vivo* silencing of each lincRNA significantly decreased the number of adult worm couples recovered from the infected mice, with reductions of 55%, 57%, 38%, or 42% for SmLINC101519, SmLINC110998, SmLINC124324, or SmLINC175062, respectively **(Figure 9B).** Although no differences were found in the number of naturally unpaired males (**Figure 9C)** and females (**Figure 9D)** recovered in the perfusion in any of the silencing conditions tested, the total number of recovered worms **(Figure 9E)** decreased significantly by 31%, 35%, 22% or 26% for SmLINC101519, SmLINC110998, SmLINC124324, and SmLINC175062 silencing, respectively.

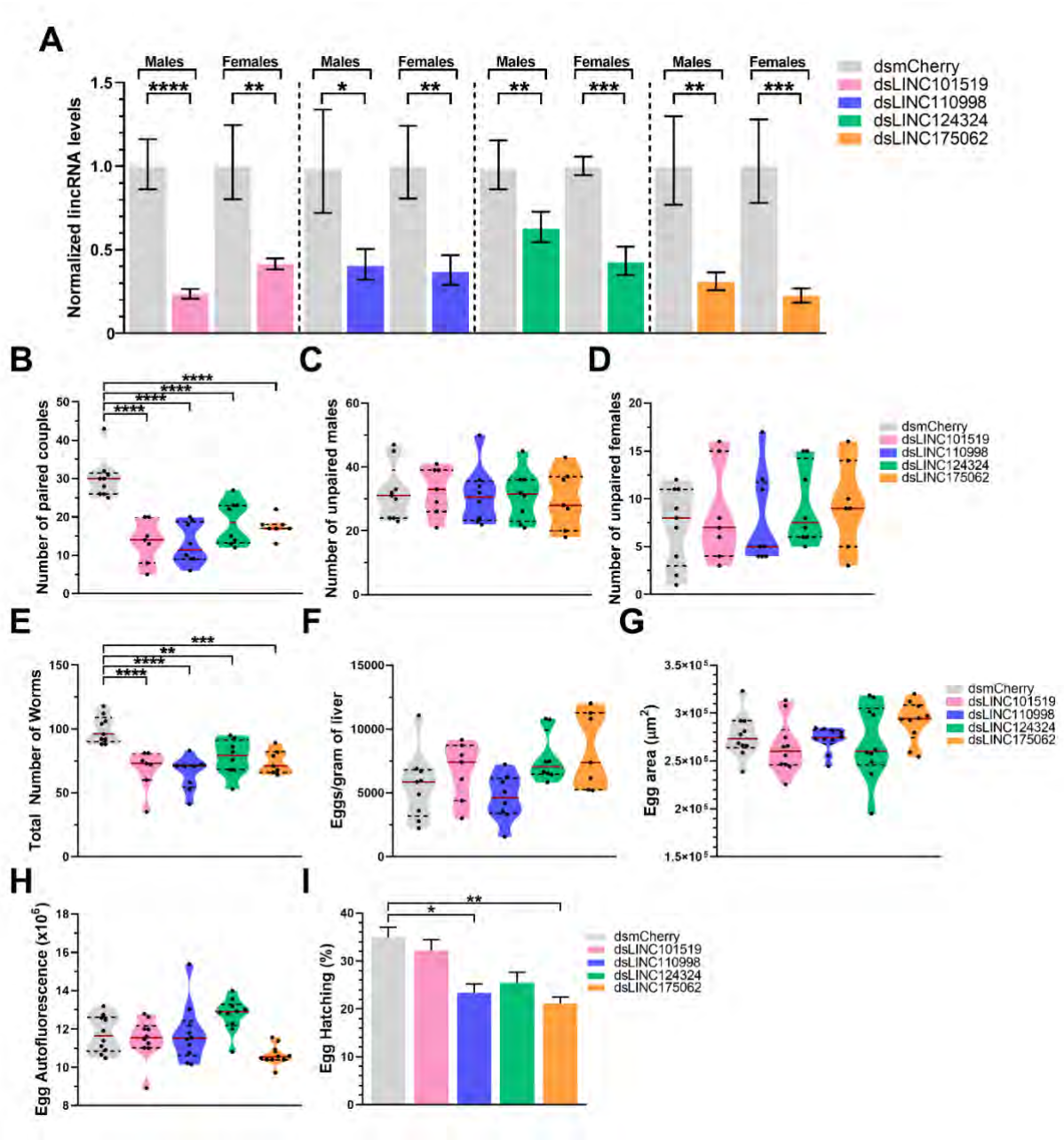


Figure 9: *In vivo* silencing (RNAi) of pairing-dependent lincRNAs decreases the number of *S. mansoni* adult worms recovered from infected mice. Five weeks old C57BL/6 female mice were infected with an average of 145 cercariae. At weeks 1, 3, 5 and 6 post infection, mice were injected in the retro orbital vein with 30 µg of dsRNA targeting the specified lincRNA (in a 100 µL solution with non-pyrogenic saline), namely SmLINC101519 (pink), SmLINC110998 (blue), SmLINC124324 (green), and SmLINC175062 (orange). A dsRNA targeting mCherry (a gene that is not present in *S. mansoni* and mice) was assayed in parallel as a negative control. At week 7, the mice were perfused, and the worms retrieved. The liver from each mouse was processed for measuring eggs. **(A)** RT-qPCR results for each lincRNA expression level are normalized to the geometric mean of reference genes Smp_099690, Smp_023150, Smp_196510, and Smp_101310. **(B-E)** Worm burden results are reported, where the number of paired worm couples **(B)**, number of unpaired male worms **(C)**, number of unpaired female worms **(D)** retrieved from perfusion, and total number of worms **(E)** are shown. The number of eggs retrieved from the livers of infected mice was measured after liver processing as mentioned in Methods. Those eggs were counted and the number of eggs per gram of liver is reported **(F)**. Egg health parameters such as size (area) **(G)** and eggshell integrity (autofluorescence) **(H)** were measured. The eggs were kept in culture for another 7 days in ABC media for synchronization of their development. Then egg hatching assay was performed as described in Methods and the percentage of hatched eggs is shown **(I)**. Standard

error of the mean (SEM) is shown in the error bars. (*) = $p < 0.05$; (**) = $p < 0.01$; (***) = $p < 0.001$; (****) = $p < 0.0001$, Student t test or One-Way ANOVA with multiple comparisons to dsmCherry group.

Surprisingly, no difference in the number of eggs per gram of liver was observed **(Figure 9F),** nor a change in the egg size (**Figure 9G)** or eggshell integrity (**Figure 9H)** in the eggs recovered from the livers of infected mice. This suggests that the worm-couples that died upon the *in vivo* silencing of the pairing-dependent lincRNAs had probably ceased laying eggs only at a late point in time in our experimental setup, near the day of perfusion on the 7th week post-infection. By this time, the eggs released by sexually mature females (Cheever et al., 2002) during weeks 5 to 7 would have already accumulated in the mice livers, thus masking the eventual impact of the decreased worm burden on the number of eggs per gram of liver detected after perfusion on week 7.

Most importantly, a significant 34% or 40% decrease in egg hatching was observed in the eggs from mice that were treated with dsRNAs targeting SmLINC110998 or SmLINC175062, respectively (**Figure 9I)**. These were the same lincRNAs whose *in vitro* silencing caused a decrease in egg hatching (**see Figure 5J**), again indicating that silencing of these particular lincRNAs affected distinct pathways involved with egg development and maturation.

Taken together, the above results show that *S. mansoni* lincRNAs can be silenced *in vitro* and *in vivo*. In summary, this is the first evidence in *S. mansoni* of different pairing related lincRNAs being involved with some critical biological events of the parasite such as cell proliferation, female vitellaria development, and female reproduction.

Discussion

Here, we show that long intergenic non-coding RNAs are differentially expressed between *S. mansoni* paired and unpaired adult worms cultured *in vitro*. Further, we provide a proof of concept that *in vitro* and *in vivo* silencing of selected pairing-dependent lincRNAs can cause a reduction of adult worm pairing or worm burden, of worm viability, of reproductive organ proliferation, and of egg laying or egg hatching. The *in vitro* unpairing mimetic model used here can provide a good tool for understanding the pairing-dependent homeostasis of the adult worms, and for further exploring the detailed mechanisms of action of lincRNAs, since these worms develop to their fullest potential before being unpaired *in vitro*. On the other hand, the mixed-sex/single-sex cercariae infection of mammalian host model (Lu et al., 2016) is a more drastic approach for understanding the role of pairing in these parasites and may also reflect the impact of single sex infection on the developmental trajectory of the parasites.

Hundreds of lncRNAs were found to be differentially expressed among the RNA-Seq samples of the mixed-sex/single-sex cercariae infections of the mammalian host, in the dataset generated by Lu et al., 2016. While short RNAs (especially miRNAs) have been more explored in helminths, lncRNAs have received little attention, with the most characterized organism being schistosomes (Silveira et al., 2022). A previous work by our group has shown a lincRNA, that is up-regulated upon 5-azacytidine treatment, to be critical for biological events of the parasite such as cell proliferation, female vitellaria development and female reproduction (Silveira et al.,

2022). Yet, no other attempt to functionally characterize the roles of lncRNAs in *S. mansoni* has been made.

A detailed investigation on the male:female pairing process has been recently published, in which Chen and collaborators showed that this process induces the GLI1-dependent expression of a non-ribosomal peptide synthetase in male worms, culminating in the production of a dipeptide formed of β-alanyl-tryptamine (BATT) (Chen et al., 2022). The BATT molecule is released from males to act as a pheromone in females, stimulating its sexual development. It remains to be tested if any lincRNA would be involved in BATT signaling.

SmLINC101519 is an 882-nt long lincRNA with two exons and an intron size of 3875 nt, mapping to chromosome 1 (SM_V7_1:11,062,339-11,067,098). *In vitro* silencing of SmLINC101519 has led to a major decrease in female oviposition, egg cell proliferation, and male adult worm cell proliferation. On chr 1, two protein-coding genes are in the neighboring genomic region of SmLINC101519, namely T-box transcription factor TBX20 (Smp_003900) upstream, and High-affinity cGMP-specific 3',5'-cyclic phosphodiesterase 9A (PDE9A) (Smp_342020) downstream of the SmLINC101519 locus (see locus in the genome browser). Inspection of the expression profiles of the two protein-coding genes and the SmLINC101519 in the adult worms scRNA-Seq data (Morales-Vicente et al., 2022) **(Figure S4A)** shows that both Smp_003900 and SmLINC101519 (G890 gene) are expressed in the same cell clusters, namely muscle 1 and neuron 17 **(Figure S4A)**; one could raise the hypothesis that SmLINC101519 could be *cis*-acting to activate the expression of the neighbor protein-coding gene, or that both Smp_003900 and SmLINC101519 are simultaneously co-regulated; however, further direct experimentation is needed to test this hypothesis. The major phenotypes observed upon SmLINC101519 silencing are in direct agreement with its expression profile and neighboring genes. TBX20 (Smp_003900) is essential for motor neuron development in invertebrates (Pocock et al., 2008) and SmLINC101519 silencing impacted adult worm adhesion and pairing by 75-100%. While PDE9A plays important roles in cell proliferation regulation in different parasites (Flueck et al., 2019; Kunz et al., 2017).

SmLINC110998 is a 322-nt long lincRNA with two exons and an intron size of 1025 nt, mapping to chr 1 (SM_V7_1:54,991,086-54,992,435). *In vitro* silencing of SmLINC11099 impacted female oviposition, egg cell proliferation, female gonad, and adult worm cell proliferation status, and female vitellaria maintenance. Those phenotypes are in concordance with the expression profile of SmLINC110998 in females, which is highly detected in the late vitellocytes cell cluster (Morales-Vicente et al., 2022) **(Figure 5B, cluster 4)**. Schistosomes produce eggs consisting of an oocyte surrounded by specialized “yolk” cells known as vitellocytes (Collins et al., 2011), which provide both nutrition for the developing zygote and constituents essential for the construction of the egg shell (Wang et al., 2019). Vitellocytes are produced by the vitellarium, which is composed of a network of thousands of follicles in which S1 stem cells differentiate to ultimately produce mature vitellocytes (S4 cells). These mature vitellocytes are fed anteriorly through the vitelline duct and are joined with fertilized oocytes in the ootype where the mature egg is formed. Disruption of mature vitellocytes production has not impacted cell proliferation status but indeed affected egg laying and egg development (Wang et al., 2019). Semaphorin-5A (Smp_159050) and Lysyl oxidase homolog 2B (Smp_159060) are the genomic upstream and downstream neighbors of SmLINC110998, respectively; however, neither of these protein-coding

genes has an expression pattern that is similar to the cell clusters' expression pattern of SmLINC110998 (**Figure S4B**), thus providing no clues to a possible regulation *in cis* eventually exerted by the lincRNA.

SmLINC124324 is a 2179-nt long transcript with 4 exons mapping to chromosome 2 (SM_V7_2:30,551,404-30,560,558). Its silencing caused a 64% reduction in adult worm viability followed by major impacts on female adult worm and ovary cell proliferation. The single-cell RNA-Seq analysis (Morales-Vicente et al., 2022) **(Figure 5C)** has pointed SmLINC124324 as enriched in late vitellocytes, agreeing with the WISH localization (**Figure 5C**). Knockdown of SmLINC124324 has impacted female vitellaria homeostasis, with vitelline and lipid droplet staining being absent in S1 to S4 vitelline cells (**Figure 8**). SmLINC124324 was found by WISH near the head and body margins of male worms (**Figure 5C**), although not detected in males in the scRNA-Seq analysis (Morales-Vicente et al., 2022). WD repeat-containing protein 6 (WDRCP6) (Smp_148220) and Biogenic amine (5HT) receptor (serotonin receptor) (Smp_245850) are the upstream and downstream genomic neighbors of SmLINC124324, and again, their cell clusters' expression patterns are quite different **(Figure S4C)**, thus providing no evidence that the lincRNA could eventually act *in cis* to regulate the protein-coding neighbor genes.

Finally, SmLINC175062 is a 669-nt long transcript with three exons mapping to chromosomes ZW (SM_V7_ZW:69,884,648-69,886,379) with Smp_096310 Protein kinase C zeta type (PKCZ) and Smp_096290 Transmembrane protein 256 homolog genes as its upstream and downstream neighbor genes; because SmLINC175062 (G40249 gene) was not detected by scRNA-Seq (**Figure S4D**), there is no evidence of a possible effect of the lincRNA on the neighbor protein-coding genes. SmLINC175062 silencing caused major effects in female egg laying and egg proliferation, with a male adult worm cell proliferation defect as well. SmLINC175062 expression was only observed by WISH in females, spread across their mature vitellocytes (**Figure 5D**), which is consistent with the female and egg phenotypes observed.

In summary, our *in vitro* and *in vivo* silencing experiments have shown that lncRNAs play pivotal roles at different aspects involved with homeostasis maintenance of schistosome adult worms. To our knowledge, this is the first report demonstrating the *in vivo* silencing of lncRNAs in parasites. We obtained evidence of major phenotypes upon *in vivo* silencing of selected lincRNAs, such as reduced worm burden and decreased egg hatching; further detailed characterization of the different mechanisms of action that were affected is warranted. In addition, we provide an extensive repository of lincRNAs differentially expressed between unpaired and paired worms, which is an important asset for future exploration of *S. mansoni* lincRNAs as possible *in vivo* targets against schistosomiasis.

Methods

Analysis of RNA-Seq data

Public RNA-Seq data from Lu et al. (Lu et al., 2016) for *S. mansoni* males, females and their gonads were downloaded from the SRA-NCBI database (project number PRJEB1237; bM: #ERS420093, #ERS420106, #ERS420107; sM: #ERS420103, #ERS420104, #ERS420105; bF: #ERS420099, #ERS420100, #ERS420101; sF: #ERS420108, #ERS420109, #ERS420110; bT: #ERS420096, #ERS420097,

#ERS420098; sT: #ERS420094, #ERS420095, #ERS420102; bO: #ERS420090, #ERS420091, #ERS420092; sO: #ERS420088, #ERS420089). Adapters and bad quality reads were filtered out using fastp v. 0.19.5 with default parameters (Chen et al., 2018). For transcripts expression quantitation the genome sequence v.7, and a GTF file containing the protein-coding transcriptome v 7.1 were downloaded from the WormBase ParaSite resource (version WBPS14) (Howe et al., 2017). The latter was merged with the lncRNA transcriptome sequences identified by Maciel et al. (Maciel et al., 2019) and the resulting GTF, which is available at http://verjolab.usp.br/public/schMan/schMan3/macielEtAl2019/files/, was used as the reference. The filtered RNA-Seq reads were aligned with STAR v 2.7 (Dobin et al., 2013) and quantified with RSEM v 1.3.1 (Li and Dewey, 2011), both using default parameters, and with the RSEM "estimate-rspd parameter on" option. Transcripts with counts lower than 10 were removed and differential expression analysis was performed using DESeq2 package (Love et al., 2014) v. 1.24.0 with an FDR threshold of 0.05. PCA plot was obtained after normalization using the vst function followed by the plotPCA function from DESeq2.

Ethics statement

The experimental protocols were in accordance with the Ethical Principles in Animal Research adopted by the Conselho Nacional de Controle da Experimentação Animal (CONCEA) and the protocol/experiments have been approved by the Comissão de Ética no Uso de Animais do Instituto Butantan (CEUAIB number 8859090919). This study was carried out in compliance with the ARRIVE guidelines (http://www.nc3rs.org.uk/page.asp?id=1357).

Parasite material

The BH strain (Belo Horizonte, Brazil) of *S. mansoni* was maintained in the intermediate snail host *Biomphalaria glabrata* and as the definitive host of the golden hamster (*Mesocricetus auratus*). Female hamsters aged 4 weeks, freshly weaned, weighing 50–60 g, were housed in cages (30 × 20 × 13 cm) containing a sterile bed of wood shavings. Female mice (C57BL/6) aged 5 weeks, weighing 17-20 g, were housed in cages (100 cm^2 in a 12,7 cm height) containing a sterile bed of wood shavings. A standard diet (Nuvilab CR-1 Irradiada, Quimtia S/A, Paraná, Brazil) and water were made available *ad libitum*. The room temperature was kept at 22 ± 2 °C, and a 12:12 h light-dark cycle was maintained.

Hamsters were infected with an *S. mansoni* cercariae suspension containing approximately 200–250 cercariae via subcutaneous injection (De Souza et al., 1979). After 42 days of infection, *S. mansoni* adult worms were recovered by perfusion of the hepatic portal system (Smithers and Terry, 1965).

Mice (C57BL/6) were infected with an *S. mansoni* cercariae suspension containing approximately 145 cercariae via a metal ring placed on the shaved abdominal skin, for 30 min, under ketamine hydrochloride (10 mg/kg body weight) and xylazine (0.5 mg/kg body weight) (Sespo, Sao Paulo, Brazil) anesthesia. After 49 days of infection, *S. mansoni* adult worms were recovered by perfusion of the hepatic portal system (Smithers and Terry, 1965). *S. mansoni* eggs were extracted from *S. mansoni* infected mice livers as previously described (Dalton et al., 1997).

Parasite *in vitro* mimetic model of paired and unpaired adult worms

To study the impact of lncRNAs on the pairing status of *S. mansoni*, we have used an *in vitro* mimetic protocol in which the worms (*S. mansoni* males or females)

retrieved from perfusion of hamsters infected with mixed sex cercariae were cultured *in vitro* for up to 8 days either as paired couples or as separated worms (Galanti et al., 2012) **(Figure 1B)**. Importantly, the naturally unpaired worms retrieved from perfusion were also cultured *in vitro* for mimicking the single-sex worms. For the pairing experiments, ten adult worm paired couples, or unpaired males or females naturally recovered from the hamster perfusion were cultured in 6-well plates containing 5 mL of ABC media (Wang et al., 2019) for 8 days and 70% of the media was exchanged every other day.

To check if this *in vitro* mimetic model has characteristics similarly to the mixed-sex/single-sex *in vivo* infection model used by Lu et al., 2016, we performed RT-qPCR assays to measure protein-coding genes known to be differentially expressed in the mixed-sex/single-sex *in vivo* infection model. We first conducted a screening on genes that could be used as reference genes for the RT-qPCR analyses, by looking at the genes whose expression was more stable in the re-analysis of: (1) all RNA-Seq libraries from Lu et al., 2016, (2) all adult worm RNA-Seq libraries from Lu et al., 2016 and (3) the three best reference genes found in a previous analysis (Haeberlein et al., 2019). The efficiencies of RT-qPCR primers for all candidate reference genes are shown in **Supplementary Table S3.** The Cq values obtained in the RT-qPCR analyses of the putative candidate reference genes measured in all paired and unpaired male and female *in vitro* cultured samples are shown in **Supplementary Table S4**. A gene expression stability survey from all candidate reference genes was taken with GeNorm (Vandesompele et al., 2002) **(Supplementary Table S5)** and NormFinder (Andersen et al., 2004) **(Supplementary Table S6)**. Our analysis found Smp_099690 (*Protein RER1*) and Smp_023150 (*Serine/threonine-protein phosphatase 6 catalytic subunit*) as the two most stable genes in our 14 different samples from adult worms cultured *in vitro* for up to 8 days either paired or not.

Then, we measured the expression of 14 protein-coding genes as controls, including genes that are differentially expressed between females and males. The efficiencies of RT-qPCR primers for all protein-coding genes used are shown in **Supplementary Table S7,** and the Cq values obtained from this analysis are shown in **Supplementary Table S8.** In **Figure S5A**, we have measured the expression of the p14 gene (Smp_316140) in males and females cultured *in vitro* either paired or unpaired and compared the expression with the data retrieved from our re-analysis of the RNA-Seq from Lu et al., 2016. Expression of the p14 gene was mainly present in females when compared with males, and mostly present in females obtained from mixed-sex infection. Our *in vitro* mimetic model analysis has shown that the p14 gene was differentially expressed when comparing paired and unpaired females *in vitro* cultured for 2, 4, and 8 days. Notably, even though the ABC medium was developed to adequately sustain females egg deposition *in vitro*, the p14 gene expression was sensitive to *in vitro* culturing of the parasites **(Figure S5A)**.

On the other hand, when looking at the expression pattern of *Calcium-binding protein CML11* gene (Smp_032990) **(Figure S5B)**, another gene mostly expressed in females, we have observed an *in vitro* culturing time-dependent differential expression in unpaired females when compared with paired females. When looking at the Smp_051920 *Nanos-type domain-containing protein* **(Figure S5C)**, a 3.2-fold higher expression in unpaired females was observed when compared to paired females cultured *in vitro* for 8 days; this is in line with its higher expression in the male and female gonads when compared to the whole worm in the re-analysis of Lu et al., 2016 data, and with the significantly higher expression in the ovaries retrieved from single-sex infections compared with mixed-sex infections **(Figure S5C)**. Further, another gene that

is mostly expressed in females has been measured, the female-specific protein 800 (fs800, Smp_307900), and the differential expression was increased in culture in a time-dependent manner in the paired females when compared with the unpaired ones **(Figure S5D)**.

Other genes such as *Egg Shell Protein* (Smp_000430), *Nucleotide pyrophosphatase: phosphodiesterase 5* (Smp_153390), *Potassium channel toxin gamma-KTx* (Smp_194830), *Heat shock 70 kDa protein homolog* (Smp_302170), and *Tumor necrosis factor receptor superfamily member 16* (Smp_332480) **(Figure S6)** have been validated in our *in vitro* mimetic model when comparing with the data retrieved from the re-analysis of Lu et al, 2016 data. When looking at the *Calcium release-activated calcium channel protein 1* (Smp_076650), *Neurocalcin homolog* (Smp_085650), and the *Vasa-like DEAD-box RNA helicase* genes (Smp_068440), their expression profiles in the Lu et al. model and our mimetic model could suggest a tendency for validation, but no significant differential expression was seen **(Figure S7)**. Finally, the *Putative Nanos RNA binding protein* (Smp_055740) and the *5-hydroxytryptamine receptor 1A* (Smp_126730) were two genes not validated when comparing the mixed-sex/single-sex cercariae infection model with our *in vitro* culturing model **(Figure S8)**.

In conclusion, the *in vivo* mimetic model has proven to be comparable to the mixed-sex/single-sex *in vivo* infection model since most protein-coding genes analyzed were differentially expressed with a similar pattern in both models.

dsRNA synthesis and *in vitro* silencing assays

To select the region of the lincRNA transcript to be targeted by the dsRNA, an *in silico* analysis of the transcript on-target and off-target bases was performed to avoid off-target effects **(Supplementary Table S2)**. Out of the 8 possible lincRNAs that were validated in our mimetic model by the RT-qPCR analysis, we selected 4 for *in vitro* silencing and for *in situ* hybridization (method described below): SmLINC101519, SmLINC110998, SmLINC124324, and SmLINC175062. Because SmLINC142881 has only 46 bases in its transcript that could be targeted by the dsRNA, it was not chosen for silencing; while SmLINC154295 and SmLINC141426 were selected only for *in situ* hybridization (see below). SmLINC133371 was discarded because it was not possible to amplify double-stranded DNA templates by PCR.

Double-stranded RNA (dsRNA) was synthesized from DNA templates amplified from cDNA of male and female adult worms, using the specific primer sequences indicated in **Supplementary Table S9**, all of them containing, in addition to the lincRNA sequence, a 17-nt T7 RNA Polymerase promoter sequence at their 5′-end. The *in vitro* dsRNA transcription reaction was adapted from a tRNA transcription protocol (Sampson and Uhlenbeck, 1988). Briefly, reactions were performed at 37 °C for 12 h in a buffer containing 40 mM Tris–HCl (pH 8.0), 22 mM MgCl2, 5 mM DTT, 2 mM spermidine, 0.05% BSA, 15 mM guanosine monophosphate, 7.5 mM of each nucleoside triphosphate, amplified template DNA (0.1 µg/µL) and 5 µM of T7 RNA polymerase. The transcribed dsRNA was treated with DNase at 37 °C for 30 min and precipitated using 1:10 (v/v) 3 M sodium acetate pH 5.2 and 1:1 (v/v) of isopropanol. The pellet was washed twice with 70% ethanol and then eluted in apyrogenic saline to reach a final concentration of 3 µg/µL. Double-stranded RNAs (30 µg/mL/day) were provided to the parasites via soaking (Krautz-Peterson et al., 2007). The mCherry gene was used as a non-related dsRNA control and its DNA template was amplified from a pPLOT-mCherry plasmid containing the mCherry gene.

For the silencing experiments, adult worms retrieved from infected Syrian hamsters were cultured for eight days in ABC media (Wang et al., 2019) supplemented with the dsRNA targeting each of the lincRNAs of interest, with mCherry dsRNA (a control dsRNA that will activate the RNAi pathway but will not target any parasite gene) or with no dsRNA. While dsRNA was added every day to the culture (to a final concentration of 30 µg/mL), the medium was exchanged every two days (70% of medium exchange). Worm pairing, adhesion, and motility were observed every day, while worm gene silencing, worm viability, worm proliferative cell status, female worm vitellaria status, and egg-related phenotypes were observed only after the eight-day treatment. At the end of the experiment, the adult worm couples were collected and stored in RNAlater (Thermo Fisher).

In vivo silencing

Infected mice (C57BL/6) were injected intravenously via the retro-orbital plexus with 0.1 mL of the indicated dsRNA at the concentration of 0.3 mg/mL (meaning 30 µg of dsRNA per injection), following the application of anesthetic collyrium (proxymetacaine hydrochloride Eye Drops 0.5%). The dsRNA injections occurred at weeks 1, 3, 5, and 6 post-infection.

RNA extraction and cDNA synthesis

Total RNA from adult males and adult females was extracted using the Qiagen RNeasy Plus Micro Kit (Cat number 74034). Briefly, 20-40 adult paired and unpaired males and females for each of four biological replicates were grounded with glass beads in buffer RLT supplemented with 2-mercaptoethanol, according to Qiagen recommendation, for 2 min and then frozen in liquid nitrogen. After freezing, male and female adult worms were ground only once and frozen and thawed three times. The protocol was followed according to the manufacturer's instructions, including gDNA exclusion by the provided gDNA elimination column. All RNA samples were quantified using the Qubit RNA HS Assay Kit (Q32852, Thermo Fisher Scientific), and the integrity of RNAs was verified using the Agilent RNA 6000 Pico Kit (5067–1513, Agilent Technologies) in a 2100 Bioanalyzer Instrument (Agilent Technologies). For the pairing status experiments, complementary DNAs (cDNAs) were obtained by reverse transcription (RT) from 1000 ng of total RNA, and for the *in vitro* and *in vivo* silencing experiments, cDNA was obtained from 200 ng and 400 ng of total RNA, respectively, using SuperScript IV Reverse Transcriptase (18091050, Invitrogen) and random hexamer primers in a 20 µL volume, according to the manufacturer's recommendations.

Primer design, quantitative RT-qPCR assays and analyses

All primer pairs were designed using the PrimerQuest Tool provided by IDT Integrated DNA Technologies (https://www.idtdna.com/PrimerQuest/) with PCR amplicon length ranging from 50 to 300 bp and melting temperature (Tm) of approximately 60 °C. All primer sequences are reported in **Supplementary Table S9,** and the primers efficiencies are shown in **Supplementary Table S3** (reference genes), **Supplementary Table S7** (pairing-dependent protein-coding genes), and **Supplementary Table S10** (pairing-dependent lncRNAs). The Cq values from the measurements of the 12 selected lincRNA expression levels on the *in vitro* mimetic model samples (at days 2, 4 and 8 of incubation) are shown in **Supplementary Table S11**. The Cq values obtained from the *in vitro* silencing of the 4 selected lincRNAs are

described in **Supplementary Table S12**. The Cq values obtained from the *in vivo* silencing of the 4 selected lincRNAs are described in **Supplementary Table S13**.

After reverse transcription, the obtained cDNAs were diluted 8x in water. Quantitative PCR was performed using 2.5 µL of each diluted cDNA in a total volume of 10 µL containing 1× LightCycler 480 SYBR Green I Master Mix (04707516001, Roche Diagnostics), 800 nM of each primer in a LightCycler 480 System (Roche Diagnostics), and each real-time qPCR was run in three technical replicates. The PCR conditions included initial activation at 95 °C for 10 min; 45 cycles with denaturation at 95 °C for 10 s, annealing at 60 °C for 10 s, and extension at 72 °C for 20 s. A dissociation step (95 °C for 15 s, 60 °C for 1 min, 95 °C for 15 s, 60 °C for 15 s) was added at the end of the run to confirm the amplicon specificity for each gene. The quantitative RT-qPCR assays were performed following the MIQE guidelines (Andersen et al., 2004; Bustin et al., 2009; Vandesompele et al., 2002). The amplification efficiency for each primer was calculated using a diluted series of cDNA synthesized using 5 µg of RNA from *S. mansoni* male and female adult worms, as previously described (Silveira et al., 2021).

Two different tools were used to evaluate gene expression stability of candidate reference genes using RT-qPCR data: geNorm (Vandesompele et al., 2002) and NormFinder (Andersen et al., 2004), and they were all processed as previously described (Silveira et al., 2021). The ΔCt method (Livak and Schmittgen, 2001) was used to determine the expression of the genes in all conditions tested, except for the silencing experiments. For the silencing experiments, the expression was measured in comparison to the control group (dsmCherry) and was conducted following a previously reported work (Taylor et al., 2019). The two best reference genes found in the present work (Smp_099630.1 and Smp_023150) were used to normalize the expression of the genes of interest in the *in vitro* mimetic model samples and in the *in vitro* silencing experiments. For the *in vivo* silencing results, Smp_196510.1 and Smp_101310.1 were used as the reference genes besides Smp_099630.1 and Smp_023150, based on a previous publication (Silveira et al., 2021).

Viability, Pairing, Adhesion, and Motility measurement

The viability of *S. mansoni* adult worms after treatment with dsRNAs was determined by a cytotoxicity assay using the CellTiter-Glo Luminescent Cell Viability Assay (G7570, Promega) (Panic et al., 2015; Pereira et al., 2018). The assay determines the amount of ATP present in freshly lysed adults; the assay signals the presence of metabolically active cells. Pairing status of the parasites was evaluated daily; only parasites with female completely outside the male's gynecophore canal were considered unpaired. Adhesion status of the parasites was evaluated daily by counting the number of females or males adhered to the plate by the ventral sucker; only parasites with no adhesion of the ventral sucker for a time longer than 10 s were considered non-adherent. Motility of the parasites was evaluated daily according to a previously determined score (Horiuchi et al., 2005; Pereira et al., 2019). Briefly, parasites with full body movement were scored 3, those that had partial or no movement, but were alive, were scored 1.5 and those that were dead scored 0; the treatment was considered lethal when no parasite movement was observed for longer than 2 min.

Egg laying and egg parameters measurement

Eggs were collected from the medium in which the worm pairs (or unpaired females) were cultured. After medium collection, wells were washed with PBS to ensure complete collection of eggs. Once collected, the eggs were suspended and 40 µL

aliquots were taken for counting under a bright light with 4-x magnification using a Nikon Eclipse inverted microscope. Counting was performed after eight days of *in vitro* culturing with dsRNA targeting the genes of interest and the amount of eggs/couple/day of incubation was then estimated. Images of the collected eggs were acquired under a 10-x magnification Nikon Eclipse inverted microscope, and the egg size was measured using ImageJ.

The egg shell integrity was measured based on the egg autofluorescence (which is present because of a high concentration of phenolic proteins that form the eggshell (Ashton et al., 2001)). Autofluorescence was acquired by fluorescence microscopy with a 492 nm emission microscope filter under a 40x magnification using a Nikon Eclipse fluorescence inverted microscope. Egg autofluorescence was quantified using ImageJ.

Egg proliferation assays were conducted with synchronized eggs as previously reported (Wang et al., 2019). Briefly, collected eggs were cultured *in vitro* for 7 days in ABC media. Cell proliferation was assayed by labeling with thymidine analog 5-ethynyl-2′-deoxyuridine (EdU), which was added to the eggs on the 6th day of *in vitro* culture at a final concentration of 10 μM and incubation for 24 h. EdU detection was performed as previously described (Collins III et al., 2013). Cells nuclei were stained with 10 μM of DAPI (Sigma). The ratio of proliferating cells over total cells was measured with ImageJ.

Egg hatching assays were conducted with eggs laid from females after *in vitro* culturing or with the eggs retrieved from the livers of infected mice from the *in vivo S. mansoni* lincRNAs silencing assays. Eggs were collected from livers and incubated *in vitro* for 7 days in ABC media for synchronization, as previously reported (Wang et al., 2019). *S. mansoni* eggs extraction and hatching were conducted as previously reported (Dalton et al., 1997; Wang et al., 2019).

Adult worm staining

Cell proliferation was assayed by labeling worms with thymidine analog 5-ethynyl-2′-deoxyuridine (EdU), which was added to the cultures on the 7th day of cultivation at a final concentration of 10 μM and incubation for 24 h. EdU detection was performed as previously described (Collins III et al., 2013). Nuclei cells were stained with 10 μM of DAPI (Sigma). Female vitellaria was stained with Fast Blue BB and BODIPY for vitelline and lipid droplets visualization, respectively, as previously described (Wang et al., 2019). DAPI was used to stain the cells nuclei. All images were acquired using a Confocal Microscope Leica TCS SP8.

Whole mount in situ hybridization and imaging

To perform *in situ* hybridization experiments for lincRNA localization, 6 lncRNAs were selected based 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 lincRNA, each lincRNA 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 are described in **Supplementary Table S2**. Pairs of primers to clone all 6 lncRNA marker probes were designed using the PrimerQuest Tool provided by IDT Integrated DNA Technologies (https://www.idtdna.com/PrimerQuest/) and are shown in **Supplementary Table S9**. 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 (Wang et al., 2016; Wendt et al., 2020).

Whole mount colorimetric *in situ* hybridization analyses were performed as previously described (Wang et al., 2016; Wendt et al., 2020). All lncRNA probes were used at 10 ng/mL in hybridization buffer. Brightfield images were acquired on a Zeiss AxioZoom V16 equipped with a transmitted light base and a Zeiss AxioCam 105 Color camera.

Acknowledgments

We would like to thank the Laboratório de Biologia Celular from Instituto Butantan and the Confocal Lab Technician Alexsander Seixas de Souza for the services provided on the Confocal Microscope Leica TCS SP8 purchased through Projeto 175 FINEP - IBUINFRA grant 01.12.0175.00 to Dr. Carlos Jared. The WISH experiments were carried out by Dr. Lu Zhao in the laboratory of Dr. James J. Collins III, Department of Pharmacology, UT Southwestern Medical Center, Dallas, TX 75390, USA, and we thank them for the generous gift of their work and data to be included in the present work.

Funding

This work was supported by a grant from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) Thematic grant number 2018/23693-5 to S.V.A. G.O.S., L.F.M., and A.S.A.P. received fellowships from FAPESP (18/24015-0, 164533/2019-2, and 116733/2019-5, respectively); H.S.C. and G.G.O.O. were supported by fellowships from Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (CNPq); S.V.A. laboratory was also supported by institutional funds from Fundação Butantan and S.V.A. received an established investigator fellowship award from CNPq, Brasil.

Contributions

G.O.S., M.S.A. and S.V.A. conceived the project and designed the experiments. G.O.S., H.S.C., G.G.O.O., A.S.A.P., P.A.M., M.L.S.O., carried out the wet-lab experiments. G.O.S., L.F.M. and A.C.T. carried out the *in-silico* experiments. E.N. provided reagents and materials. G.O.S. analyzed the data and wrote the first draft of the manuscript. M.S.A. and S.V.A. revised the manuscript. M.S.A. supervised the study. S.V.A. obtained funding and coordinated the study. All authors reviewed and approved the manuscript.

Competing interests

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

Amaral MS, Maciel LF, Silveira GO, Olberg GGO, Leite JVP, Imamura LK, Pereira ASA, Miyasato PA, Nakano E, Verjovski-Almeida S. 2020. Long non-coding

RNA levels can be modulated by 5-azacytidine in Schistosoma mansoni. *Sci Rep* **10**:21565. doi:10.1038/s41598-020-78669-5

Andersen CL, Jensen JL, Ørntoft TF. 2004. Normalization of Real-Time Quantitative Reverse Transcription-PCR Data: A Model-Based Variance Estimation Approach to Identify Genes Suited for Normalization, Applied to Bladder and Colon Cancer Data Sets. *Cancer Res* **64**:5245–5250. doi:10.1158/0008-5472.CAN-04-0496

Ashton PD, Harrop R, Shah B, Wilson RA. 2001. The schistosome egg: development and secretions. *Parasitology* **122**:329–38. doi:10.1017/s0031182001007351

Basch PF. 1981. Cultivation of Schistosoma mansoni In vitro. I. Establishment of Cultures from Cercariae and Development until Pairing. *J Parasitol* **67**:179. doi:10.2307/3280632

Bergquist R, Utzinger J, Keiser J. 2017. Controlling schistosomiasis with praziquantel: How much longer without a viable alternative? *Infect Dis Poverty* **6**:74. doi:10.1186/s40249-017-0286-2

Berriman M, Haas BJ, Loverde PT, Wilson RA, Dillon GP, Cerqueira GC, Mashiyama ST, Al-Lazikani B, Andrade LF, Ashton PD, Aslett MA, Bartholomeu DC, Blandin G, Caffrey CR, Coghlan A, Coulson R, Day TA, Delcher A, Demarco R, Djikeng A, Eyre T, Gamble JA, Ghedin E, Gu Y, Hertz-Fowler C, Hirai H, Hirai Y, Houston R, Ivens A, Johnston DA, Lacerda D, MacEdo CD, McVeigh P, Ning Z, Oliveira G, Overington JP, Parkhill J, Pertea M, Pierce RJ, Protasio A V., Quail MA, Rajandream MA, Rogers J, Sajid M, Salzberg SL, Stanke M, Tivey AR, White O, Williams DL, Wortman J, Wu W, Zamanian M, Zerlotini A, Fraser-Liggett CM, Barrell BG, El-Sayed NM. 2009. The genome of the blood fluke Schistosoma mansoni. *Nature* **460**:352–358. doi:10.1038/nature08160

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT. 2009. The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* **55**:611–622. doi:10.1373/clinchem.2008.112797

Cheever AW, Lenzi JA, Lenzi HL, Andrade ZA. 2002. Experimental Models of Schistosoma mansoni Infection. *Mem Inst Oswaldo Cruz* **97**: 917–940. doi:10.1590/S0074-02762002000700002

Cheever EA, Macedonia JG, Mosimann JE, Cheever AW. 1994. Kinetics of Egg Production and Egg Excretion by Schistosoma mansoni and S. japonicum in Mice Infected with a Single Pair of Worms. *Am J Trop Med Hyg* **50**:281–295. doi:10.4269/ajtmh.1994.50.281

Chen J, Wang Y, Wang C, Hu J-F, 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

Chen R, Wang J, Gradinaru I, Vu HS, Geboers S, Naidoo J, Ready JM, Williams NS, DeBerardinis RJ, Ross EM, Collins JJ. 2022. A male-derived nonribosomal peptide pheromone controls female schistosome development. *Cell* **185**:1506-1520.e17. doi:10.1016/j.cell.2022.03.017

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**:i884–i890. doi:10.1093/bioinformatics/bty560

1009 Colley DG, Bustinduy AL, Secor WE, King CH. 2014. Human schistosomiasis. *Lancet*
1010 **383**:2253–2264. doi:10.1016/S0140-6736(13)61949-2

1011 Collins III JJ, Wang B, Lambrus BG, Tharp ME, Iyer H, Newmark PA. 2013. Adult
1012 somatic stem cells in the human parasite Schistosoma mansoni. *Nature* **494**:476–
1013 479. doi:10.1038/nature11924

1014 Collins JJ, King RS, Cogswell A, Williams DL, Newmark PA. 2011. An Atlas for
1015 Schistosoma mansoni Organs and Life-Cycle Stages Using Cell Type-Specific
1016 Markers and Confocal Microscopy. *PLoS Negl Trop Dis* **5**:e1009.
1017 doi:10.1371/journal.pntd.0001009

1018 Cort WW. 1921. Sex in the Trematode Family Schistosomidæ. *Science (80-)* **53**:226–
1019 228. doi:10.1126/science.53.1367.226

1020 Dalton JP, Day SR, Drew AC, Brindley PJ. 1997. A method for the isolation of
1021 schistosome eggs and miracidia free of contaminating host tissues. *Parasitology*
1022 **115**:29–32. doi:10.1017/S0031182097001091

1023 De Souza CP, Dias EP, De Azevedo MD, Paulini E. 1979. [Observations upon some
1024 factors which influence the laboratory maintenance of Schistosoma mansoni
1025 (author's transl)]. *Rev Bras Pesqui Med Biol* **12**:411–419.

1026 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,
1027 Gingeras TR. 2013. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*
1028 **29**:15–21. doi:10.1093/bioinformatics/bts635

1029 Flueck C, Drought LG, Jones A, Patel A, Perrin AJ, Walker EM, Nofal SD, Snijders
1030 AP, Blackman MJ, Baker DA. 2019. Phosphodiesterase beta is the master regulator
1031 of cAMP signalling during malaria parasite invasion. *PLOS Biol* **17**:e3000154.
1032 doi:10.1371/journal.pbio.3000154

1033 Galanti SE, Huang SC-C, Pearce EJ. 2012. Cell Death and Reproductive Regression in
1034 Female Schistosoma mansoni. *PLoS Negl Trop Dis* **6**:e1509.
1035 doi:10.1371/journal.pntd.0001509

1036 Gouveia M, Brindley P, Gärtner F, Costa J, Vale N. 2018. Drug Repurposing for
1037 Schistosomiasis: Combinations of Drugs or Biomolecules. *Pharmaceuticals* **11**:15.
1038 doi:10.3390/ph11010015

1039 Grevelding CG, Langner S, Dissous C. 2018. Kinases: Molecular Stage Directors for
1040 Schistosome Development and Differentiation. *Trends Parasitol* **34**:246–260.
1041 doi:10.1016/j.pt.2017.12.001

1042 Haeberlein S, Angrisano A, Quack T, Lu Z, Kellershohn J, Blohm A, Grevelding CG,
1043 Hahnel SR. 2019. Identification of a new panel of reference genes to study pairing-
1044 dependent gene expression in Schistosoma mansoni. *Int J Parasitol* **49**:615–624.
1045 doi:10.1016/j.ijpara.2019.01.006

1046 Hewitson JP, Maizels RM. 2014. Vaccination against helminth parasite infections.
1047 *Expert Rev Vaccines* **13**:473–487. doi:10.1586/14760584.2014.893195

1048 Horiuchi A, Satou T, Akao N, Koike K, Fujita K, Nikaido T. 2005. The effect of free
1049 and polyethylene glycol-liposome-entrapped albendazole on larval mobility and
1050 number in Toxocara canis infected mice. *Vet Parasitol* **129**:83–87.
1051 doi:10.1016/j.vetpar.2004.12.017

Howe KL, Bolt BJ, Shafie M, Kersey P, 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, Krause HM. 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, Liu X, Wang S. 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

Krautz-Peterson G, Radwanska M, Ndegwa D, Shoemaker CB, Skelly PJ. 2007. Optimizing gene suppression in schistosomes using RNA interference. *Mol Biochem Parasitol* **153**:194–202. doi:10.1016/j.molbiopara.2007.03.006

Kunz S, Balmer V, Sterk GJ, Pollastri MP, Leurs R, Müller N, Hemphill A, Spycher C. 2017. The single cyclic nucleotide-specific phosphodiesterase of the intestinal parasite Giardia lamblia represents a potential drug target. *PLoS Negl Trop Dis* **11**:e0005891. doi:10.1371/journal.pntd.0005891

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**:323. doi:10.1186/1471-2105-12-323

Li J, Xiang M, Zhang R, Xu B, Hu W. 2018. RNA interference in vivo in Schistosoma japonicum: Establishing and optimization of RNAi mediated suppression of gene expression by long dsRNA in the intra-mammalian life stages of worms. *Biochem Biophys Res Commun* **503**:1004–1010. doi:10.1016/j.bbrc.2018.06.109

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods* **25**:402–408. doi:10.1006/meth.2001.1262

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**:550. doi:10.1186/s13059-014-0550-8

LoVerde PT, Chen L. 1991. Schistosome female reproductive development. *Parasitol Today* **7**:303–308. doi:10.1016/0169-4758(91)90263-N

Lu Z, Sessler F, Holroyd N, Hahnel S, Quack T, Berriman M, Grevelding CG. 2016. Schistosome sex matters: A deep view into gonad-specific and pairing-dependent transcriptomes reveals a complex gender interplay. *Sci Rep* **6**:31150. doi:10.1038/srep31150

Lu Z, Spänig S, Weth O, Grevelding CG. 2019. Males, the Wrongly Neglected Partners of the Biologically Unprecedented Male–Female Interaction of Schistosomes. *Front Genet* **10**:796. doi:10.3389/fgene.2019.00796

Maciel LF, Morales-Vicente DA, Silveira GO, Ribeiro RO, Olberg GGO, Pires DS, Amaral MS, Verjovski-Almeida S. 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

Melman SD, Steinauer ML, Cunningham C, Kubatko LS, Mwangi IN, Wynn NB, Mutuku MW, Karanja DMS, Colley DG, Black CL, Secor WE, Mkoji GM, Loker ES. 2009. Reduced Susceptibility to Praziquantel among Naturally Occurring Kenyan Isolates of Schistosoma mansoni. *PLoS Negl Trop Dis* **3**:e504. doi:10.1371/journal.pntd.0000504

Morales-Vicente DA, Zhao L, Silveira GO, Tahira AC, Amaral MS, Collins III JJ, 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

Mutapi F, Maizels R, Fenwick A, Woolhouse M. 2017. Human schistosomiasis in the post mass drug administration era. *Lancet Infect Dis* **17**:e42–e48. doi:10.1016/S1473-3099(16)30475-3

Nath A, Lau EYT, Lee AM, Geeleher P, Cho WCS, Huang RS. 2019. Discovering long noncoding RNA predictors of anticancer drug sensitivity beyond protein-coding genes. *Proc Natl Acad Sci* **116**:22020–22029. doi:10.1073/pnas.1909998116

Panic G, Flores D, Ingram-Sieber K, Keiser J. 2015. Fluorescence/luminescence-based markers for the assessment of Schistosoma mansoni schistosomula drug assays. *Parasit Vectors* **8**:624. doi:10.1186/s13071-015-1233-3

Pereira ASA, Amaral MS, Vasconcelos EJR, Pires DS, Asif H, daSilva LF, Morales-Vicente DA, Carneiro VC, Angeli CB, Palmisano G, Fantappie MR, Pierce RJ, Setubal JC, Verjovski-Almeida S. 2018. Inhibition of histone methyltransferase EZH2 in Schistosoma mansoni in vitro by GSK343 reduces egg laying and decreases the expression of genes implicated in DNA replication and noncoding RNA metabolism. *PLoS Negl Trop Dis* **12**:e0006873. doi:10.1371/journal.pntd.0006873

Pereira ASA, Silveira GO, Amaral MS, Almeida SM V., Oliveira JF, Lima MCA, Verjovski-Almeida S. 2019. In vitro activity of aryl-thiazole derivatives against Schistosoma mansoni schistosomula and adult worms. *PLoS One* **14**:e0225425. doi:10.1371/journal.pone.0225425

Pereira TC, Pascoal VDB, Marchesini RB, Maia IG, Magalhães LA, Zanotti-Magalhães EM, Lopes-Cendes I. 2008. Schistosoma mansoni: Evaluation of an RNAi-based treatment targeting HGPRTase gene. *Exp Parasitol* **118**:619–623. doi:10.1016/j.exppara.2007.11.017

Pocock R, Mione M, Hussain S, Maxwell S, Pontecorvi M, Aslam S, Gerrelli D, Sowden JC, Woollard A. 2008. Neuronal function of Tbx20 conserved from nematodes to vertebrates. *Dev Biol* **317**:671–685. doi:10.1016/j.ydbio.2008.02.015

Popiel I, Basch PF. 1984. Reproductive development of femaleSchistosoma mansoni (Digenea: Schistosomatidae) following bisexual pairing of worms and worm segments. *J Exp Zool* **232**:141–150. doi:10.1002/jez.1402320117

Popiel I, Cioli D, Erasmus DA. 1984. The morphology and reproductive status of female Schistosoma mansoni following separation from male worms. *Int J Parasitol* **14**:183–190. doi:10.1016/0020-7519(84)90047-X

Ransohoff JD, Wei Y, Khavari PA. 2018. The functions and unique features of long

1139 intergenic non-coding RNA. *Nat Rev Mol Cell Biol* **19**:143–157.
1140 doi:10.1038/nrm.2017.104

1141 Rinn JL, Chang HY. 2020. Long Noncoding RNAs: Molecular Modalities to
1142 Organismal Functions. *Annu Rev Biochem* **89**:283–308. doi:10.1146/annurev-
1143 biochem-062917-012708

1144 Sampson JR, Uhlenbeck OC. 1988. Biochemical and physical characterization of an
1145 unmodified yeast phenylalanine transfer RNA transcribed in vitro. *Proc Natl Acad*
1146 *Sci U S A* **85**:1033–1037. doi:10.1073/pnas.85.4.1033

1147 Silveira GO, Amaral MS, Coelho HS, Maciel LF, Pereira ASA, Olberg GGO, Miyasato
1148 PA, Nakano E, Verjovski-Almeida S. 2021. Assessment of reference genes at six
1149 different developmental stages of Schistosoma mansoni for quantitative RT-PCR.
1150 *Sci Rep* **11**:16816. doi:10.1038/s41598-021-96055-7

1151 Silveira GO, Coelho HS, Amaral MS, Verjovski-Almeida S. 2022. Long non-coding
1152 RNAs as possible therapeutic targets in protozoa, and in Schistosoma and other
1153 helminths. *Parasitol Res* **121**:1091–1115. doi:10.1007/s00436-021-07384-5

1154 Smithers SR, Terry RJ. 1965. The infection of laboratory hosts with cercariae of
1155 Schistosoma mansoni and the recovery of the adult worms. *Parasitology* **55**:695–
1156 700. doi:10.1017/S0031182000086248

1157 Taylor SC, Nadeau K, Abbasi M, Lachance C, Nguyen M, Fenrich J. 2019. The
1158 Ultimate qPCR Experiment: Producing Publication Quality, Reproducible Data the
1159 First Time. *Trends Biotechnol* **37**:761–774. doi:10.1016/j.tibtech.2018.12.002

1160 Vale N, Gouveia MJ, Rinaldi G, Brindley PJ, Gärtner F, Correia da Costa JM. 2017.
1161 Praziquantel for Schistosomiasis: Single-Drug Metabolism Revisited, Mode of
1162 Action, and Resistance. *Antimicrob Agents Chemother* **61**:e02582-16.
1163 doi:10.1128/AAC.02582-16

1164 Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, Van Roy A, Speleman F.
1165 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric
1166 averaging of multiple internal control genes. *Genome Biol* **3**:34–1. doi:10.1186/gb-
1167 2002-3-7-research0034

1168 Wang J, Chen R, Collins JJ. 2019. Systematically improved in vitro culture conditions
1169 reveal new insights into the reproductive biology of the human parasite
1170 Schistosoma mansoni. *PLOS Biol* **17**:e3000254. doi:10.1371/journal.pbio.3000254

1171 Wang J, Collins JJ, Collins 3rd JJ. 2016. Identification of new markers for the
1172 Schistosoma mansoni vitelline lineage. *Int J Parasitol* **46**:405–410.
1173 doi:10.1016/j.ijpara.2016.03.004

1174 Wendt G, Zhao L, Chen R, Liu C, O'Donoghue AJ, Caffrey CR, Reese ML, Collins JJ.
1175 2020. A single-cell RNA-seq atlas of Schistosoma mansoni identifies a key
1176 regulator of blood feeding. *Science (80-)* **369**:1644–1649.
1177 doi:10.1126/science.abb7709

1178 Wilson RA. 2020. Schistosomiasis then and now: what has changed in the last 100
1179 years? *Parasitology* **147**:507–515. doi:10.1017/S0031182020000049

1180 Wilson RA, Li XH, Castro-Borges W. 2017. Schistosome vaccines: problems, pitfalls
1181 and prospects. *Emerg Top Life Sci* **1**:641–650. doi:10.1042/ETLS20170094

33

1182

Supplementary Figures and Legends

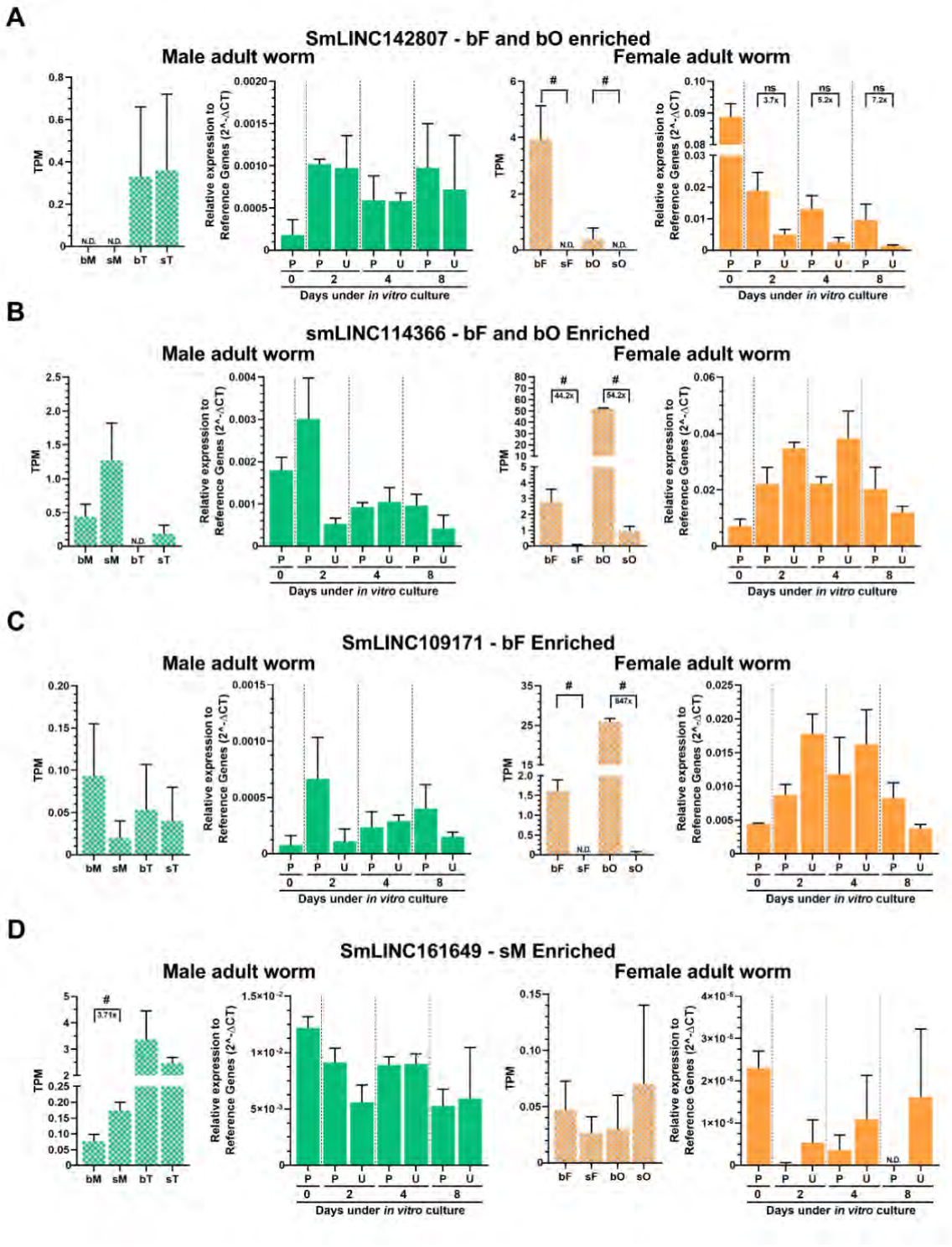


Figure S1: Expression validation of lincRNAs enriched in four different male and female samples in the RNA-Seq analyses and measured by RT-qPCR in *S. mansoni* cultured *in* vitro for 2, 4 or 8 days as paired couples or unpaired males and females. Four lincRNAs detected as enriched in the re-analyses of the RNA-Seq dataset of Lu et al., 2016, were selected for RT-qPCR assays, namely (A) SmLINC142807, enriched in bF and bO; (B) SmLINC114366, enriched in bF and bO; (C)

SmLINC109171, enriched in bF; and (D) SmLINC161649, enriched in sM. Male related results are shown on the left (green) and female results on the right (orange). Paired couples (P) or unpaired (U) parasites were obtained by perfusion of hamsters infected for 42 days with *S. mansoni* cercariae. After perfusion, males and females were cultured *in vitro* for 2, 4 or 8 days as paired (P) couples or unpaired (U) worms. RT-qPCR results (solid-colored graphs) are normalized to the geometric mean of reference genes Smp_099690 and Smp_023150. Expression values from 4 different biological replicates are shown. Standard error of the mean (SEM) is shown in the error bars. N.D.: Not detected. ns: p-value > 0.05, Student t test. For comparison, RNA-Seq data from re-analysis of Lu *et* al., 2016 is shown (plaid-colored graphs) and the expression is measured in TPM (transcripts per million); RNA-Seq data is retrieved from males (M), females (F), testes (T) or ovaries (O) from either a mixed-sex (b) or a single-sex (s) infection; (#) = FDR<0.005. The fold-change differences between the compared groups are represented under the brackets.

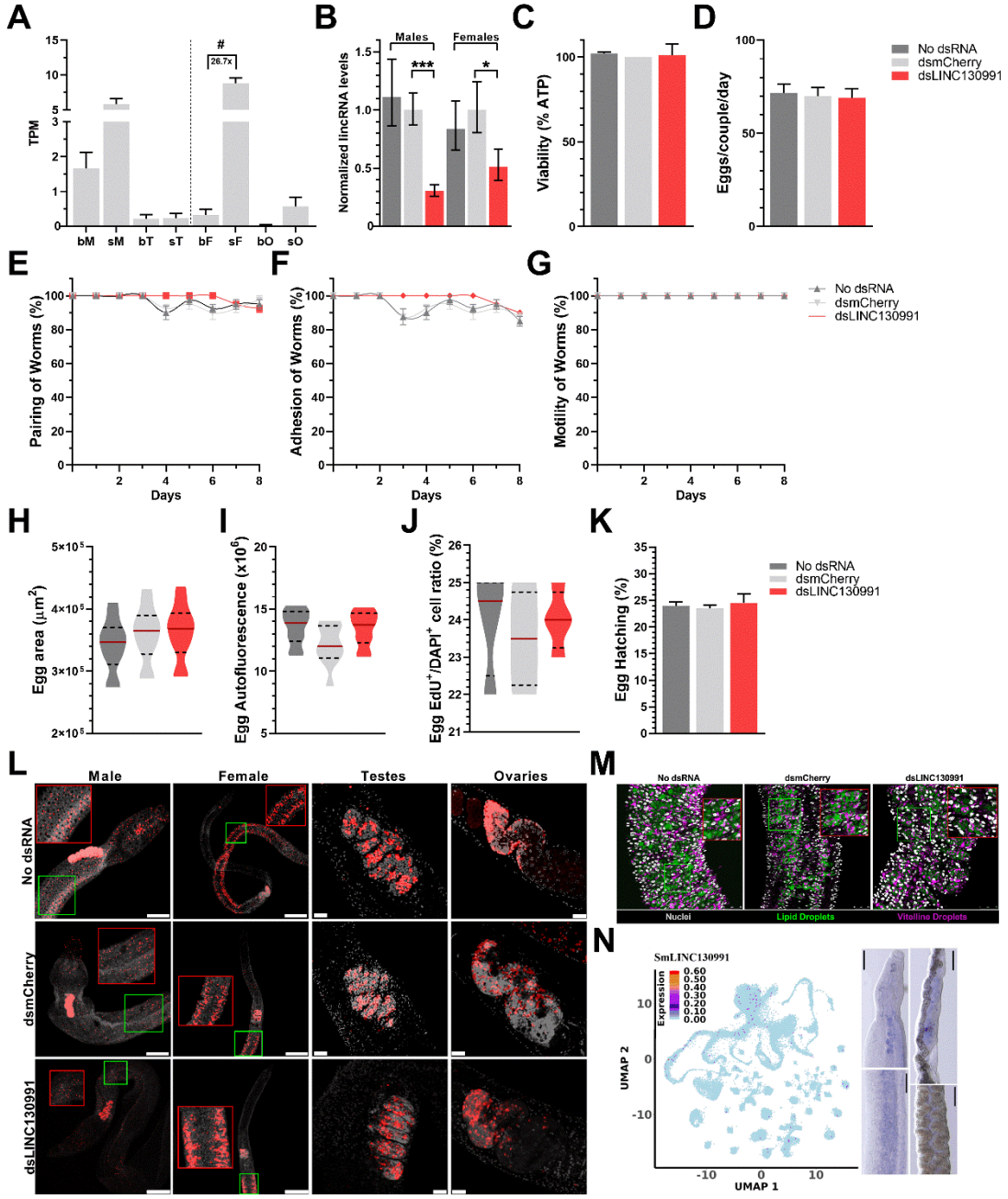


Figure S2: Lack of phenotypic changes in *Schistosoma mansoni* adult worm couples upon *in vitro* silencing of SmLINC130991, an unrelated control lincRNA. Paired couples were obtained by perfusion of hamsters infected for 42 days with *S. mansoni* cercariae. Couples were cultured *in vitro* for 8 days, in ABC media supplemented with 30 µg/mL of dsRNA targeting SmLINC130991 (red bars in panels B to K). Medium was exchanged every other day while dsRNA was added every day. dsRNA targeting mCherry (a gene that is not present in *S. mansoni*) was assayed in parallel as a negative control (light gray bars). Results for parasites cultured with no dsRNA are also shown (dark gray bars). **(A)** RNA-Seq expression data for SmLINC130991 from the re-analysis of Lu et al., 2016 is shown and the expression is measured in TPM (transcripts per million); RNA-Seq data is retrieved from males (M), females (F), testes (T) or ovaries (O) from either a mixed-sex (b) or a single-sex (s) infection; (#) = FDR<0.005. The fold-change difference between the compared groups is represented under the bracket. This lincRNA was chosen as an unrelated control because it was not enriched in any of the mixed-sex (b) samples. (**B**) RT-qPCR results for SmLINC130991 expression level are normalized to the geometric mean of reference genes Smp_099690 and Smp_023150. **(C)** Viability of adult worms (males+females) was monitored using the ATP-Glo Assay. **(D)** At the end of the experiment

(8 days), eggs were collected and counted. **(E-G)** Pairing status, adhesion to the plate and motility of worm couples were traced along the 8 days of the experiment. **(H-K)** Collected eggs were monitored for their size (area) **(H)**, integrity of their eggshell (autofluorescence) **(I)**, proliferation status of the embryos (Egg EdU+/DAPI+ cell ratio) **(J),** and the percentage of egg hatching was measured by keeping the eggs in culture for another 7 days in ABC media for synchronization of their development, then assessing egg hatching as described in Methods, with the percentage of hatched eggs being shown **(K)**. Violin plot representation at figures **(H-J)** with the median indicated by the red line and the quartiles represented by the dashed lines. (**L**) EdU detection was performed as described in the Methods. DAPI stained cells nuclei are in gray and EdU+ cells (proliferating cells) are stained in red. Scale bars: 250 μm for the adult worm images (Males and Females), and 25 μm for the adult worm gonad images (Testes and Ovaries). (**M**) Female vitellaria stained with Fast Blue BB (pink) and BODIPY (green), which labeled vitelline and lipid droplets in the vitellaria, respectively. DAPI staining of cells nuclei is shown in gray. Scale bars: 25 μm. (**N**) WISH of SmLINC130991 is shown as the blue color stains in the male (left) or female (right) adult worm images of the heads and bodies. Scale bars are 100 μm. SmLINC130991 expression patterns across single-cell clusters are shown with a UMAP plot, which is colored by gene expression level (blue = low, red = high) and the scale represents log10(UMIs+1). The red borders in all microscopy images define zoomed-in insets of interest that correspond to the regions within green borders. Representative microscopy images from 3 experiments with n > 10 parasites. Quantitative data from 4 different biological replicates is shown. Standard error of the mean (SEM) is shown in the error bars in (**B**). (*) = $p < 0.05$; (***) = $p < 0.001$, Student t test.

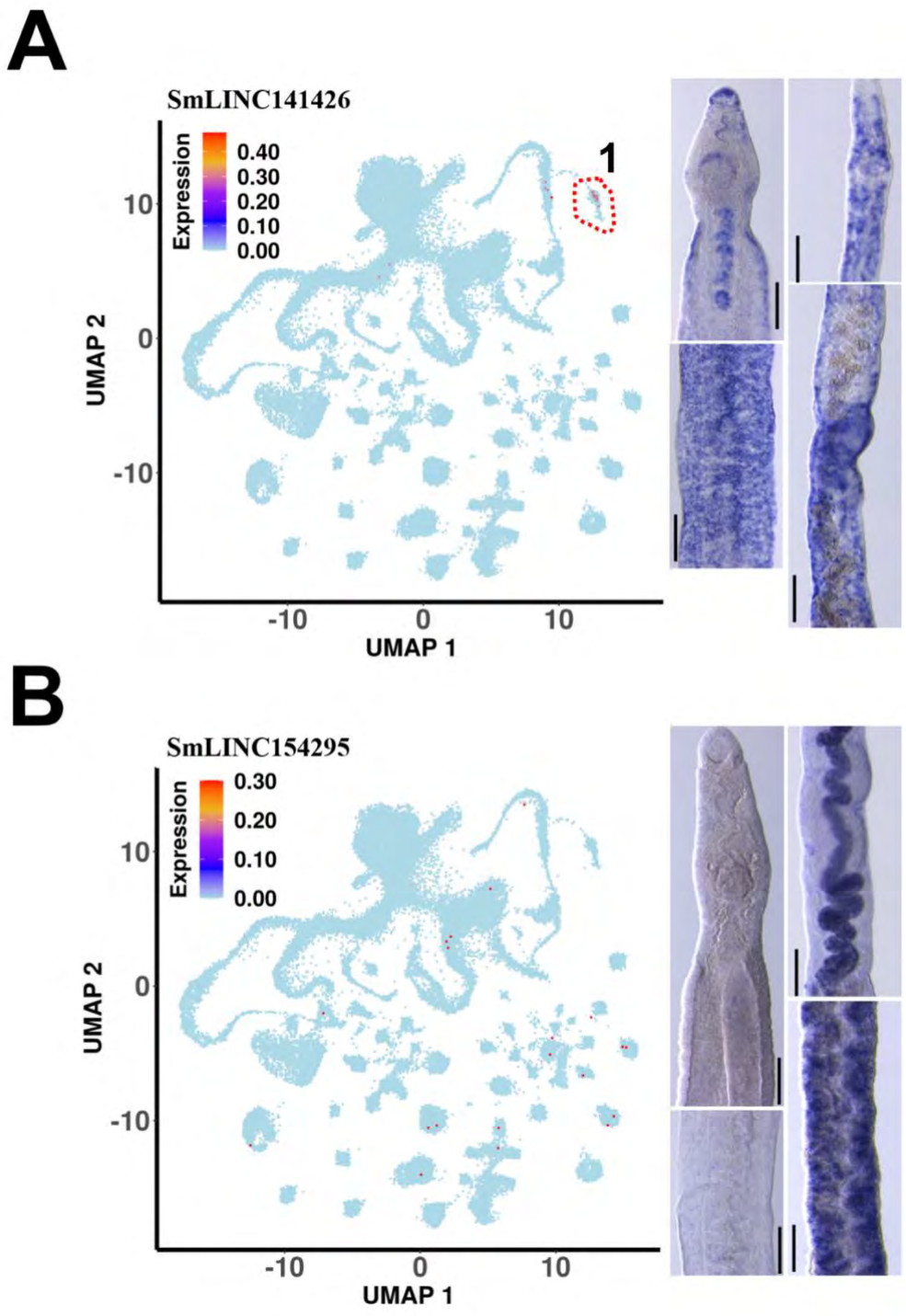


Figure S3: Localization of two selected lincRNAs in adult worm tissues by whole mount *in situ* hybridization. Whole mount *in situ* hybridization (WISH) of each lincRNA is shown as the blue color stains in the male (left) or female (right) adult worm images of the heads and bodies. Scale bars are 100 µm. Results for **(A)** SmLINC141426 and **(B)** SmLINC154295. For comparison purposes, single-cell RNA-Seq data from Morales-Vicente et al., 2022 was retrieved (http://verjolab.usp.br:8081/). LincRNA

expression patterns are shown with UMAP plots, which are colored by gene expression (blue = low, red = high) and the scale represents log10(UMIs+1). Region 1 enclosed by the red dashed line in (A) indicates the male gametes cell cluster.

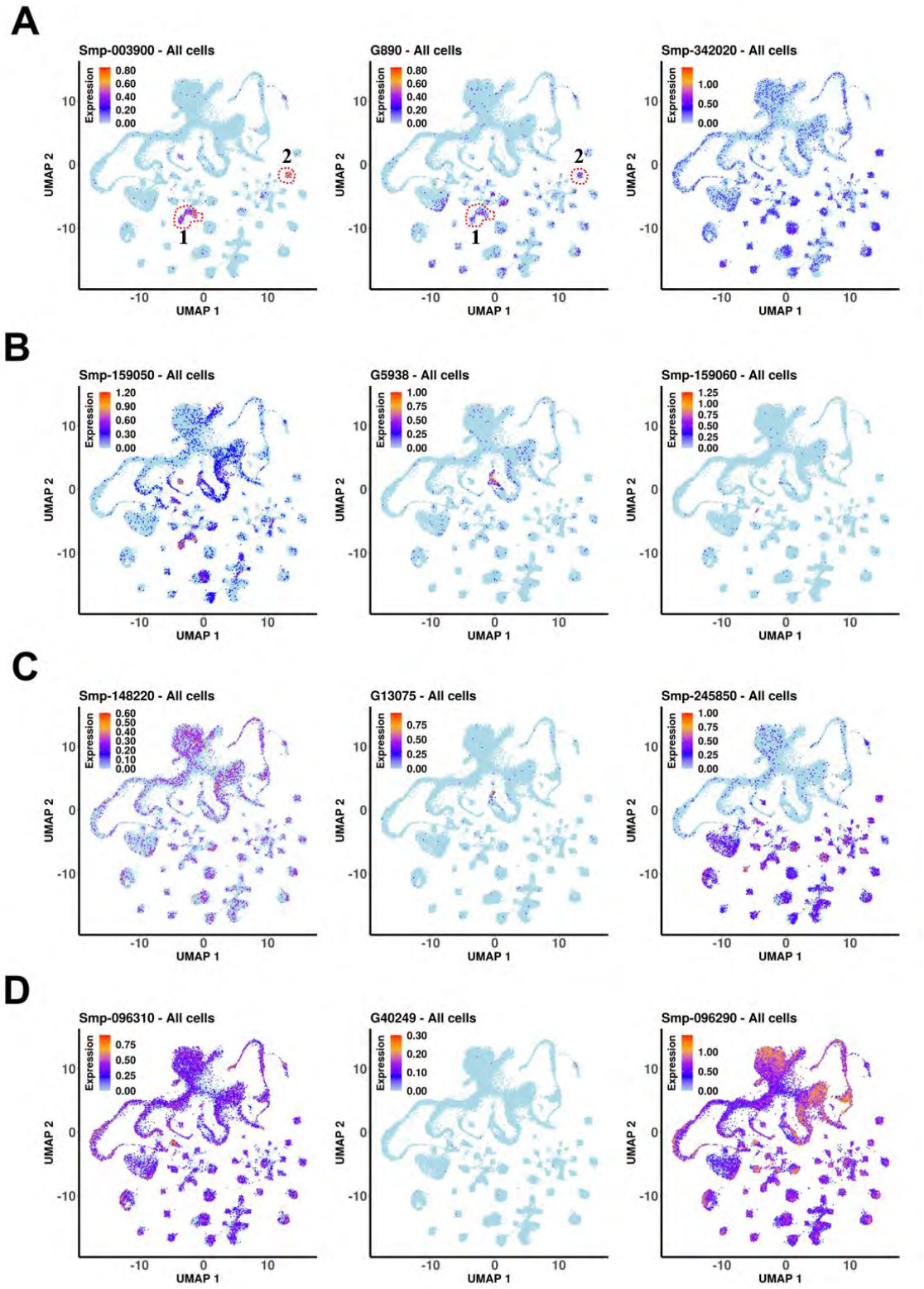


Figure S4: UMAP plots of expression levels of lincRNAs involved in pairing status, with their corresponding *cis*-neighboring protein-coding genes (up- and downstream in the lincRNA genomic locus). Single-cell RNA-Seq data was retrieved

from Morales-Vicente et al., 2022 (see http://verjolab.usp.br:8081/). The lincRNA and Smp expression patterns are represented by UMAP plots. On the left panel, the upstream neighbor protein-coding gene is shown; the lincRNA is shown in middle panel; and on the right, the downstream neighbor protein-coding gene is shown. **(A)** G890, SmLINC101519, and the red dotted lines mark (1) muscle 1 and (2) neuron 17 cell clusters; **(B)** G5938, SmLINC110998; **(C)** G13075, SmLINC124324; and **(D)** G40429, SmLINC175062.

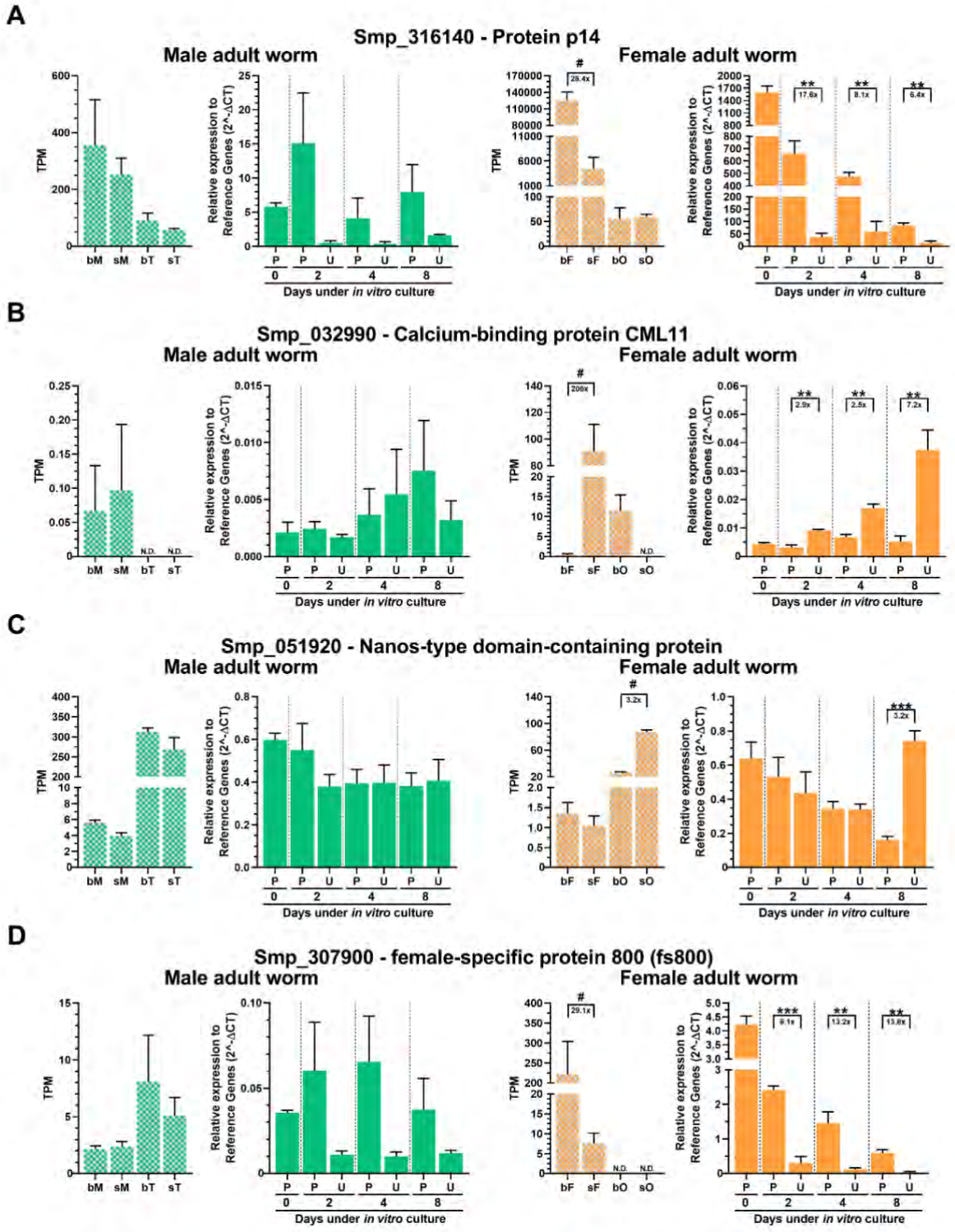


Figure S5: Expression validation of protein-coding genes known to be related to the reproductive system and measured by RT-qPCR in *S. mansoni* cultured *in* vitro for 2, 4 or 8 days as paired couples or unpaired males and females. Four protein-coding genes were selected for RT-qPCR assays, namely (**A**) Smp_316140, Protein p14; (**B**) Smp_032990, Calcium binding protein CML11; (**C**) Smp_051920, Nanos type domain-containing protein; and (**D**) Smp_307900, female specific protein 800 (fs800).

Male related results are shown on the left (green) and female results on the right (orange). Paired couples (P) or unpaired (U) parasites were obtained by perfusion of hamsters infected for 42 days with *S. mansoni* cercariae. After perfusion, males and females were cultured *in vitro* for 2, 4 or 8 days as paired (P) couples or unpaired (U) worms. RT-qPCR results (solid-colored graphs) are normalized to the geometric mean of reference genes Smp_099690 and Smp_023150. Expression values from 4 different biological replicates are shown. Standard error of the mean (SEM) is shown in the error bars. (**) = $p < 0.01$; (***) = $p < 0.001$, Student t test. N.D.: Not detected. ns: p-value > 0.05. For comparison, RNA-Seq data from re-analysis of Lu *et* al., 2016 is shown (plaid-colored graphs) and the expression is measured in TPM (transcripts per million); RNA-Seq data is retrieved from males (M), females (F), testes (T) or ovaries (O) from either a mixed-sex (b) or a single-sex (s) infection; (#) = FDR<0.005. The fold-change differences between the compared groups are represented under the brackets.

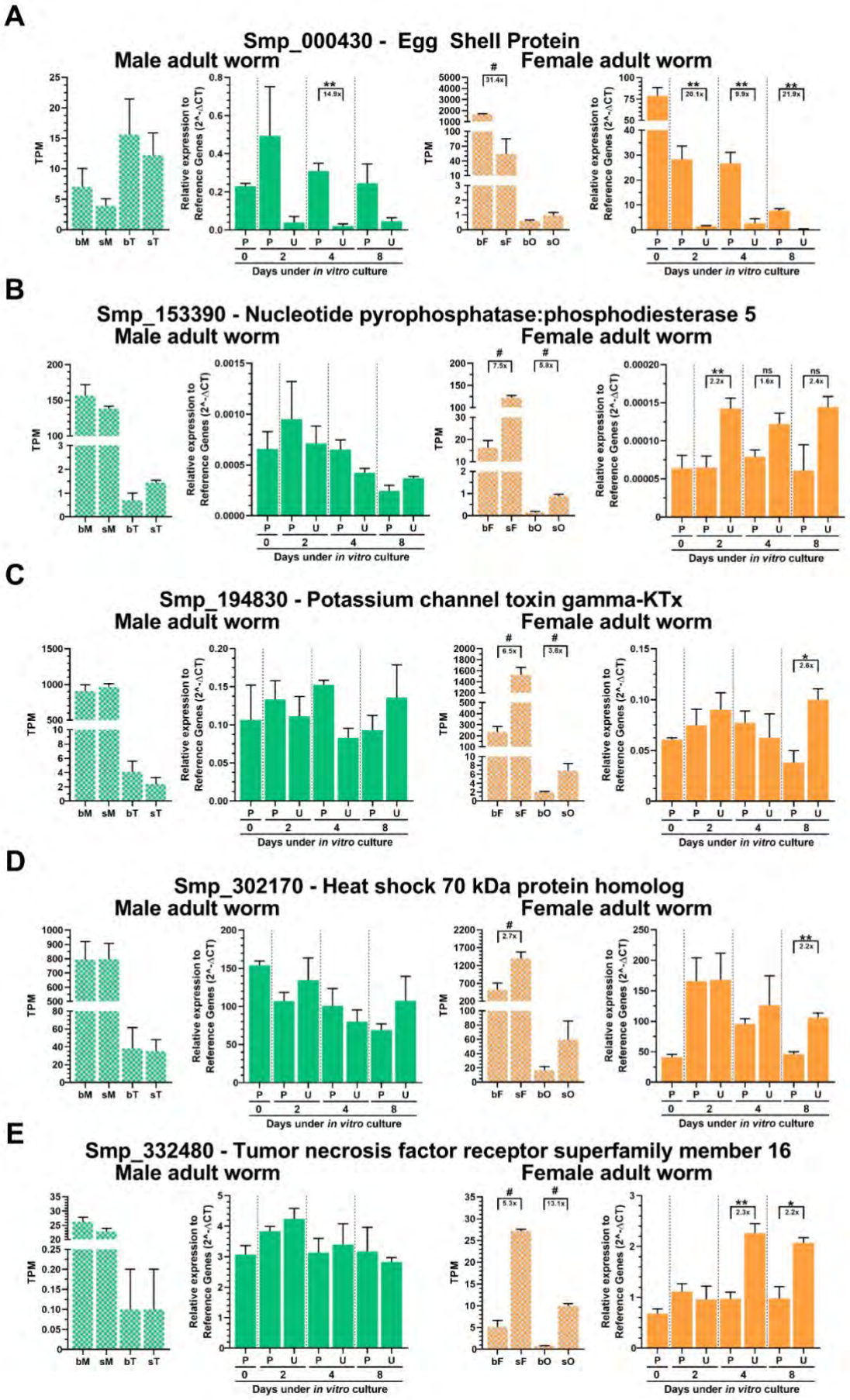

A
Smp_000430 - Egg Shell Protein
Male adult worm
Female adult worm
TPM
Relative expression to Reference Genes (2^-ΔCT)
bM sM bT sT
bF sF bO sO
P P U P U P U
0 2 4 8
Days under in vitro culture
B
Smp_153390 - Nucleotide pyrophosphatase:phosphodiesterase 5
Male adult worm
Female adult worm
C
Smp_194830 - Potassium channel toxin gamma-KTx
Male adult worm
Female adult worm
D
Smp_302170 - Heat shock 70 kDa protein homolog
Male adult worm
Female adult worm
E
Smp_332480 - Tumor necrosis factor receptor superfamily member 16
Male adult worm
Female adult worm

Figure S6: Expression validation of protein-coding genes known to be related to the reproductive system and measured by RT-qPCR in *S. mansoni* cultured *in* vitro for 2, 4 or 8 days as paired couples or unpaired males and females. Four protein-coding genes were selected for RT-qPCR assays, namely (**A**) Smp_000430, Eggshel protein; (**B**) Smp_153390, Nucleotide pyrophosphatase:phosphodiesterase 5; (**C**) Smp_194830, Potassium channel toxin gamma-KTx; and (**D**) Smp_302170, Heat shock 70 kDa protein homolog. Male related results are shown on the left (green) and female results on the right (orange). Paired couples (P) or unpaired (U) parasites were obtained by perfusion of hamsters infected for 42 days with *S. mansoni* cercariae. After perfusion, males and females were cultured *in vitro* for 2, 4 or 8 days as paired (P) couples or unpaired (U) worms. RT-qPCR results (solid-colored graphs) are normalized to the geometric mean of reference genes Smp_099690 and Smp_023150. Expression values from 4 different biological replicates are shown. Standard error of the mean (SEM) is shown in the error bars. (*) = $p < 0.05$; (**) = $p < 0.01$, Student t test. N.D.: Not detected. ns: p-value > 0.05. For comparison, RNA-Seq data from re-analysis of Lu *et* al., 2016 is shown (plaid-colored graphs) and the expression is measured in TPM (transcripts per million); RNA-Seq data is retrieved from males (M), females (F), testes (T) or ovaries (O) from either a mixed-sex (b) or a single-sex (s) infection; (#) = FDR<0.005. The fold-change differences between the compared groups are represented under the brackets.

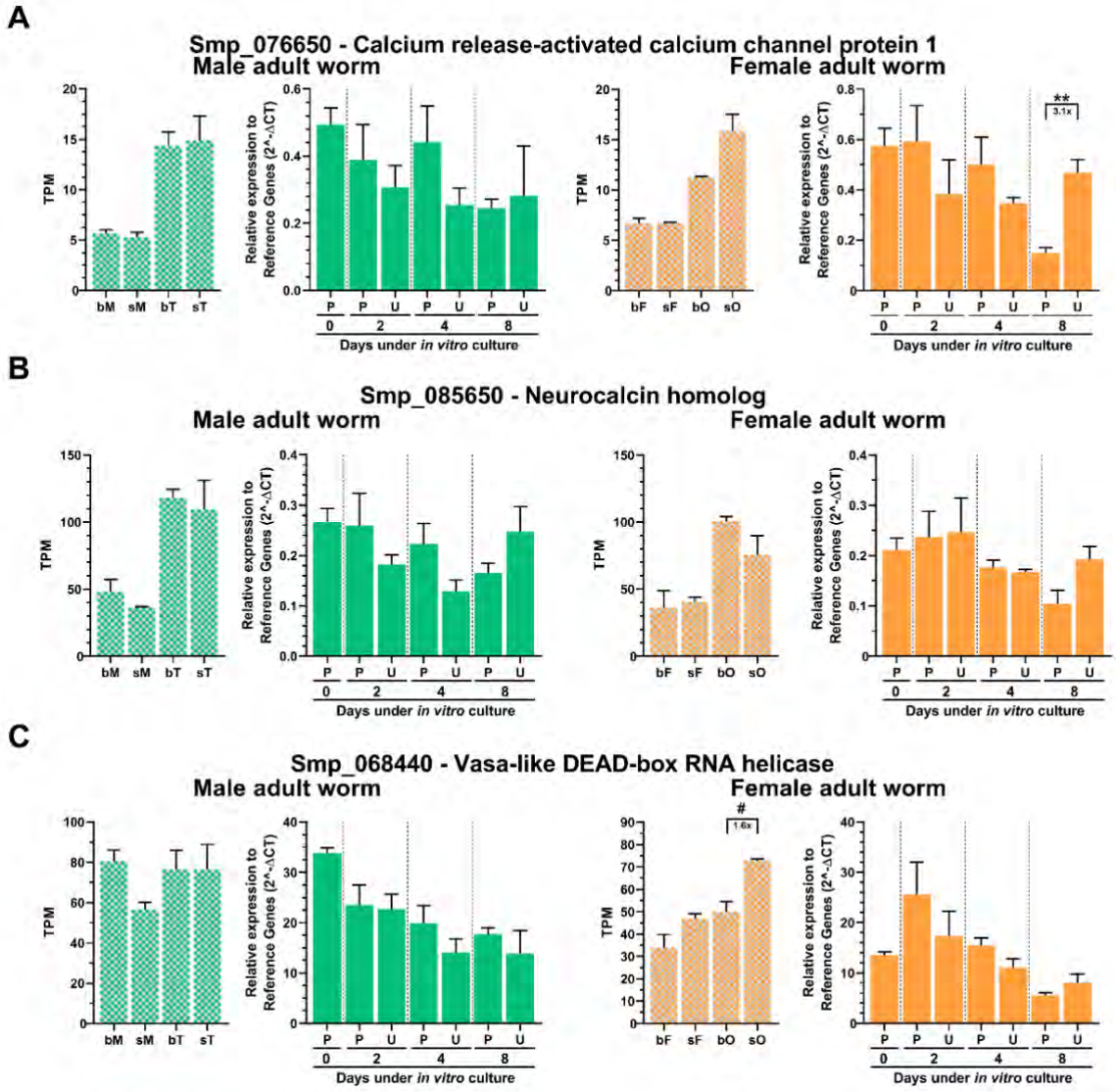


Figure S7: Expression validation of protein-coding genes known to be related to the reproductive system and measured by RT-qPCR in *S. mansoni* cultured *in* vitro for 2, 4 or 8 days as paired couples or unpaired males and females. Three protein-coding genes selected for RT-qPCR assays were validated as non-differentially expressed genes in the *in vitro* mimetic assays, namely (**A**) Smp_076650, Calcium release-activated calcium channel protein 1; (**B**) Smp_085650, Neurocalcin homolog; and (**C**) Smp_068440, Vasa-like DEAD-box RNA helicase. Male related results are shown on the left (green) and female results on the right (orange). Paired couples (P) or unpaired (U) parasites were obtained by perfusion of hamsters infected for 42 days with *S. mansoni* cercariae. After perfusion, males and females were cultured *in vitro* for 2, 4 or 8 days as paired (P) couples or unpaired (U) worms. RT-qPCR results (solid-colored graphs) are normalized to the geometric mean of reference genes Smp_099690 and Smp_023150. Expression values from 4 different biological replicates are shown. Standard error of the mean (SEM) is shown in the error bars. (**) = $p < 0.01$, Student t test. For comparison, RNA-Seq data from re-analysis of Lu *et* al., 2016 is shown (plaid-colored graphs) and the expression is measured in TPM (transcripts per million); RNA-

Seq data is retrieved from males (M), females (F), testes (T) or ovaries (O) from either a mixed-sex (b) or a single-sex (s) infection; (#) = FDR<0.005. The fold-change differences between the compared groups are represented under the brackets.

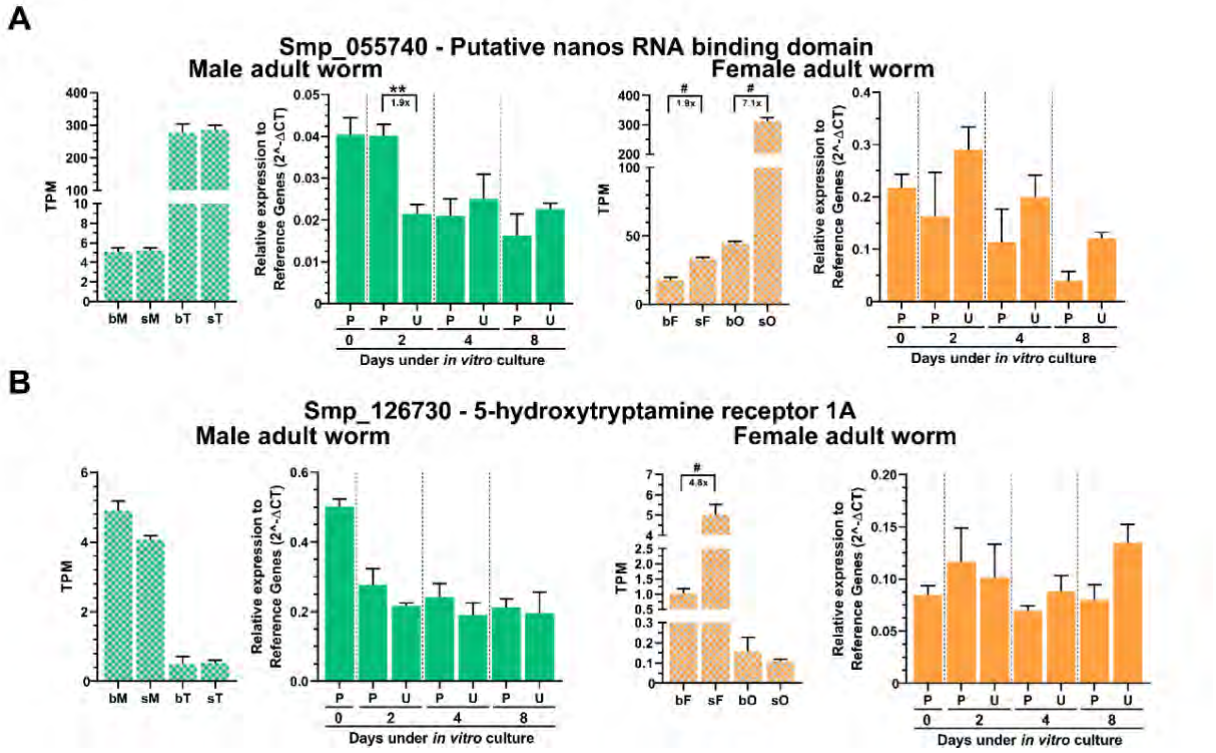


Figure S8: Expression validation of protein-coding genes known to be related to the reproductive system and measured by RT-qPCR in *S. mansoni* cultured *in* vitro for 2, 4 or 8 days as paired couples or unpaired males and females. Two selected protein-coding genes were not validated by RT-qPCR in the *in vitro* mimetic assays, namely (**A**) Smp_055740, Putative nanos RNA binding domain; and (**B**) Smp_126730, 5-hydroxytryptamine receptor 1A. Male related results are shown on the left (green) and female results on the right (orange). Paired couples (P) or unpaired (U) parasites were obtained by perfusion of hamsters infected for 42 days with *S. mansoni* cercariae. After perfusion, males and females were cultured *in vitro* for 2, 4 or 8 days as paired (P) couples or unpaired (U) worms. RT-qPCR results (solid-colored graphs) are normalized to the geometric mean of reference genes Smp_099690 and Smp_023150. Expression values from 4 different biological replicates are shown. Standard error of the mean (SEM) is shown in the error bars. (**) = $p < 0.01$, Student t test. For comparison, RNA-Seq data from re-analysis of Lu *et* al., 2016 is shown (plaid-colored graphs) and the expression is measured in TPM (transcripts per million); RNA-Seq data is retrieved from males (M), females (F), testes (T) or ovaries (O) from either a mixed-sex (b) or a single-sex (s) infection; (#) = FDR<0.005. The fold-change differences between the compared groups are represented under the brackets.

Supplementary Methods

Selection of lincRNAs to be tested for their involvement in adult worm pairing

Because our RNA-Seq re-analysis used a different reference transcriptome (Maciel et al., 2019) comprised of lncRNAs besides protein-coding RNAs, and 16,583 lncRNAs were detected as expressed in addition to 14,520 protein-coding genes, we first performed an overall inspection of the patterns of expression, using a principal components analysis (PCA) of the entire set of expressed lncRNAs and protein-coding RNAs. A pattern similar to that of Lu et al., 2016 was obtained **(Figure A)**; of note, the expression pattern of immature single-sex whole females was clustered closer to that of single-sex and bisex whole males, whereas the bisex female samples clustered quite apart **(Figure A)**, similar to what has been shown by Lu et al., 2016 for the pattern of protein-coding genes alone. This suggests a pattern of lncRNAs expression in the tissues and organs of immature and mature males and females that closely follows the patterns of expression of protein-coding genes.

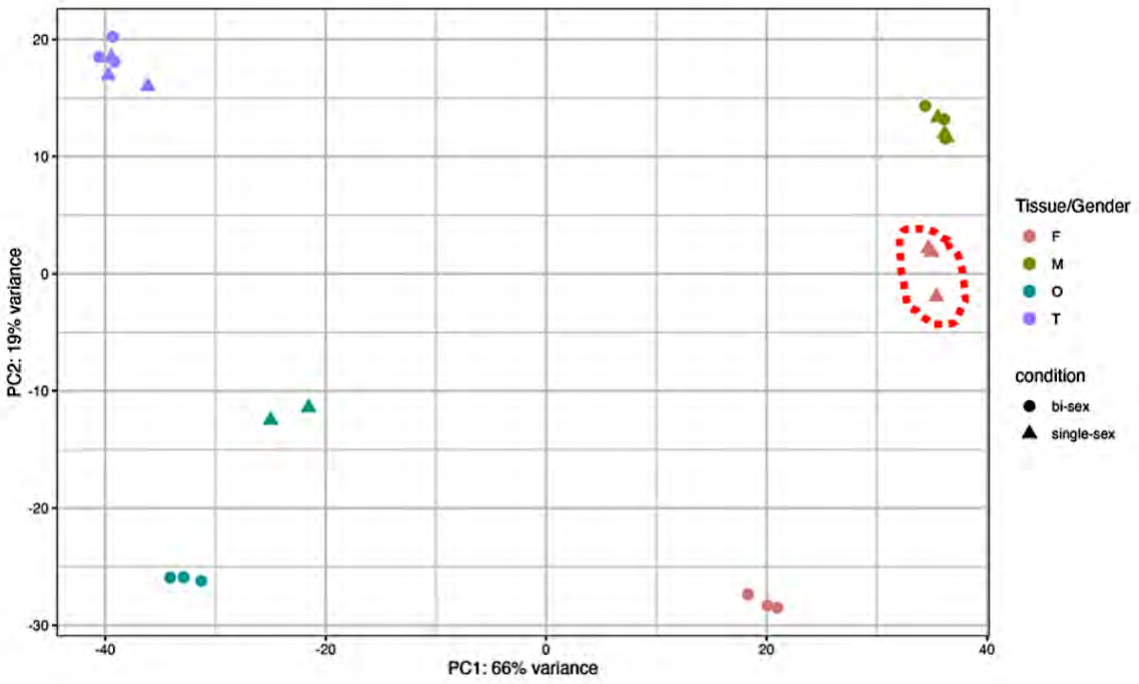


Figure A: Principal Component Analysis of expression profiles of *S. mansoni* parasites and their gonads. Principal component analysis (PCA) of expressed genes (dots and triangles) in three biological replicates (different dots or triangles) of *S. mansoni* parasites retrieved from single-sex (triangles) and bisex infections (dots). Females from bisex infections are represented by the orange dots, and from single-sex infections, by the orange triangles. Males from bisex infections are represented by the green dots, and from single-sex infections by the green triangles. Ovaries from bisex infection females are represented by the blue dots, and from single-sex infection females by the blue triangles. Testes from bisex infection males are represented by the purple dots, and from single-sex infection males by the purple triangles. Variance stabilizing transformation (VST) from counts data was used to generate the PCA plot. Single-sex female samples are highlighted by the red dashed line.

Identification of lincRNAs enriched in bisex females, males, and their reproductive organs

To identify a subset of lincRNA candidates to be tested for their possible involvement in adult worm pairing, we devised the filtering pipeline described below to single-out the lincRNAs enriched in bisex females, males, and their reproductive organs.

First, we took each of the bisex female, male, ovary, and testes samples and performed the most relevant pairwise comparisons between samples (**Table A)**, computing the number of significant DE transcripts detected in each comparison. The lists of all DE transcripts found in each comparison are shown in **Supplementary Tables S14 to S23**.

The subgroup of DE long intergenic ncRNAs (DE lincRNAs), which is contained within the group of DE lncRNAs, was singled out in a separate column, as we concentrated for further analyses on those lincRNAs that were upregulated in each of the indicated comparisons (**Table A)**.

Table A: Number of differentially expressed (DE) transcripts at each of the *Schistosoma mansoni* pairing status comparisons analyzed. The results presented here are based on a re-analysis of Lu *et al.,* 2016, which now includes long non-coding RNAs besides protein-coding genes. Statistically significant DE genes were determined with DESeq2 (FDR < 0.05). Smps: *S. mansoni* protein-coding transcripts. lncRNAs: long non-coding RNAs. lincRNAs: long intergenic non-coding RNAs, a subgroup of the lncRNAs. LincRNAs upregulated in the comparison: number of lincRNAs that were more expressed in the first analyzed condition compared to the second one. Conversely, LincRNAs downregulated in the comparison: number of lincRNAs that were more expressed in the second analyzed condition compared to the first.

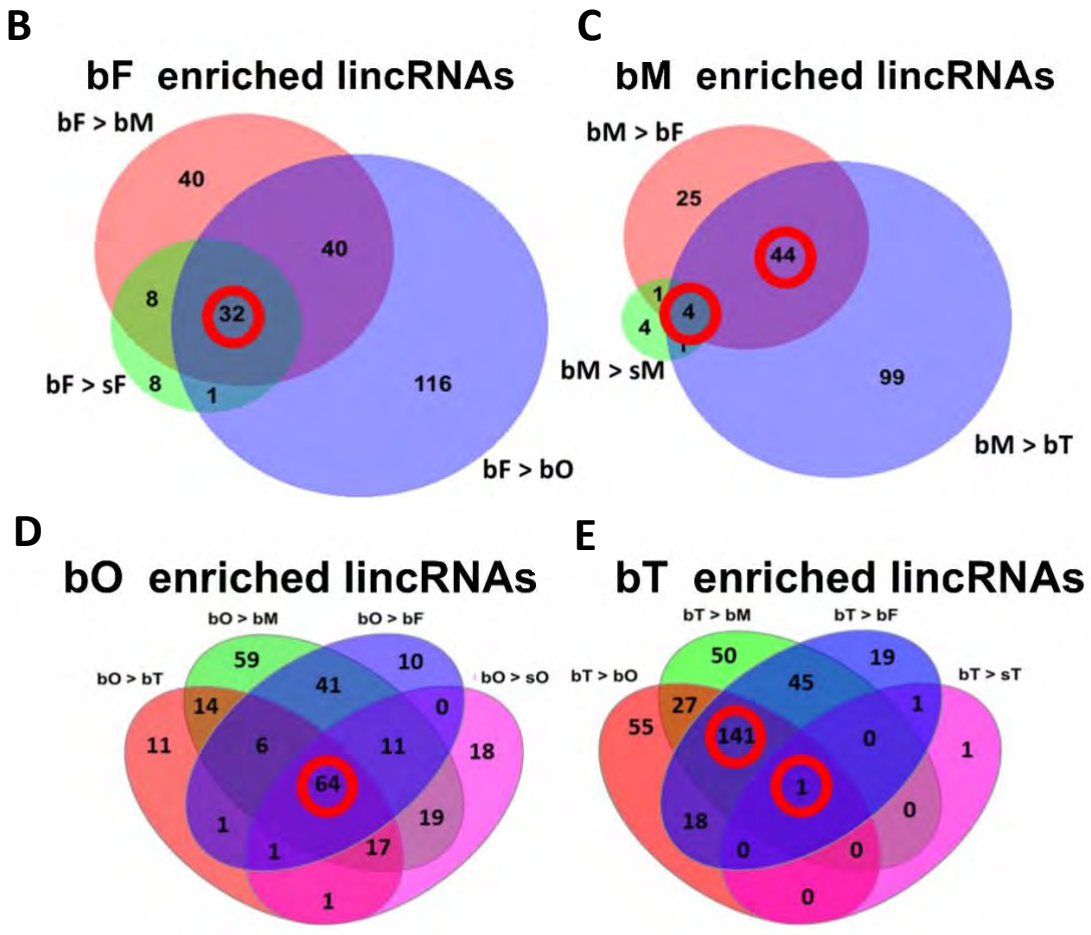
Analyzed Conditions	Total number of DE transcripts	Number of DE Smps	Number of DE lncRNAs	Number of DE lincRNAs	LincRNAs upregulated in the comparison	LincRNAs downregulated in the comparison
bF x sF	5070	4482	588	275	175	100
bF x bM	5264	4638	626	263	172	91
bF x bO	6797	5860	937	430	256	174
bM x sM	718	667	51	25	12	13
bM x bT	6912	5814	1098	473	196	277
bO x bT	6532	5563	969	409	155	254
bO x sO	5990	5232	758	347	183	164
bO x bM	9102	7426	1676	775	364	411
bT x sT	77	63	14	5	3	2
bT x bF	5922	5068	854	379	233	146

A subset of lncRNAs show higher expression in bisex worms or gonads compared with single-sex worms

Next, to further focus on a reduced list of differentially expressed lincRNAs that could play roles in *S. mansoni* worm sexual development and pairing, we cross-referenced the lists of lincRNAs found as more expressed in bisex females (bF) compared with single-sex females (sF), with bisex ovaries (bO) and with bisex males (bM), and found 32 bF enriched lincRNAs in common among these three pairwise comparisons **(Figure B, red circle).**

Similarly, when looking at lincRNAs more expressed in bisex males (bM), we found four bM lincRNAs in common in comparison with sM, bT, and bF, with an additional 44 lincRNAs in common in comparison with bT and bF alone **(Figure C, red circles)**.

When the lincRNAs more expressed in bisex ovaries (bO) were analyzed, we found 64 lincRNAs in common in the pairwise comparisons with bT, bM, bF, and sO **(Figure D, red circle)** while looking at lincRNAs more expressed in bT we found 141 in common in the pairwise comparisons with bO, bM, and bF and not with sT **(Figure E, red circle)**, with only one additional lincRNA being more expressed in bT compared with sT and present in the other comparisons **(Figure E, red circle)**.



Figures B to E: Analysis of gender and gonad enriched lincRNAs. The Venn diagrams show the number of lincRNAs that were detected as more expressed in: (**B**) bisex females (bF) in pairwise comparisons with bM, bO, and sF, and the circle highlights the number of lincRNAs detected in common in the three pairwise comparisons; (**C**) bisex males (bM) in pairwise comparisons with bF, bT and sM,

and the circles highlight the number of lincRNAs detected in common in the three pairwise comparisons and in one of the two pairwise comparisons, due to the fact that bM and sM were highly similar in terms of expression; (**D**) bisex ovaries (bO) in pairwise comparisons with bT, bM, bF and sO, and the circle highlights the number of lincRNAs detected in common in the four pairwise comparisons; (**E**) bisex testes (bT) in pairwise comparisons with bO, bM, bF and sT, and the circles highlight the number of lincRNAs detected in common in the first three pairwise comparisons, as well as in the four comparisons.

We further narrowed the selected set by excluding lincRNAs with TPM values lower than 2 (lowly expressed) and those with more than one isoform in their genomic locus, resulting in a total of 40 lincRNA candidates containing a chromatin epigenetic mark at their transcription start site (TSS), an indicative of lincRNA regulation (**Table B, upper panel**) and 31 lincRNA candidates without the epigenetic mark at their TSS (**Table B, lower panel**).

Table B: Number of lincRNAs selected for RT-qPCR validation and phenotypic assays validation. LincRNAs identified as being more expressed in common in the pairwise comparisons described in Figure 3 were further selected for RT-qPCR validation based on the level of expression above 2 transcripts per million (TPM) or above 5 TPM, and on the presence of only one isoform in their genomic locus. They were further divided into the ones having epigenetic marks at their Transcriptional Start Site (TSS) (upper panel) or not having the marks (lower panel), as annotated by Maciel *et al*. 2019. Finally, lincRNAs were selected for further investigation based on our ability to design specific sets of primer pairs that passed the 80% primer efficiency threshold.

Enriched Condition	Number of lincRNAs with epigenetic marks around TSS	Number of lincRNAs selected for further investigation	2>TPM>5	TPM>5
bF	13	5	1	4
bM	5	1		1
bO	9	1		1
bT	13	2	1	1
Total	40	9	2	7

Enriched Condition	Number of lincRNAs without epigenetic marks around TSS	Number of lincRNAs selected for further investigation	2>TPM>5	TPM>5
bF	11	3		3
bM	-	-		
bO	6	0		
bT	14	0		
Total	31	3		3

Finally, we excluded from the list of candidates those for which no specific pairs of primers for PCR could be designed, or those for which the primer pairs did not pass the 80% primer efficiency threshold. The final set of lincRNAs selected for further investigation **(Table B)** comprised 12 transcripts, being 8 bF enriched lincRNAs**,** 1 bM enriched lincRNA, 1 bO enriched lincRNA, and 2 bT enriched lincRNAs.

LISTA DE ANEXOS:

A seguir estão anexados a Súmula Curricular do aluno e 4 artigos para os quais o aluno de doutorado contribuiu como colaborador durante o período de sua Tese. Em cada artigo, a contribuição que o aluno deu para o referido trabalho está descrita no tópico "Author Contributions", ao final do artigo.

A. **Súmula Curricular de Gilbert de Oliveira Silveira**

B. Maciel LF, Morales-Vicente DA, Silveira GO, Ribeiro RO, Olberg GGO, Pires DS, Amaral MS and Verjovski- Almeida S (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

C. Pereira ASA, Silveira GO, Amaral MS, Almeida SMV, Oliveira JF, Lima MCA, et al. (2019) ***In vitro* activity of aryl-thiazole derivatives against *Schistosoma mansoni* schistosomula and adult worms.** *PLoS ONE* **14**(11): e0225425. doi: 10.1371/journal.pone.0225425

D. Coutinho Carneiro V, de Abreu da Silva IC, Amaral MS, Pereira ASA, Silveira GO, Pires DdS, et al. (2020) **Pharmacological inhibition of lysine specific demethylase 1 (LSD1) induces global transcriptional deregulation and ultrastructural alterations that impair viability in *Schistosoma mansoni*.** *PLoS Negl Trop Dis* **14**(7): e0008332. doi: 10.1371/journal.pntd.0008332

E. Lopes-Junior, E.H., Bertevello, C.R., de Oliveira Silveira, G. et al. **Human tumor necrosis factor alpha affects the egg-laying dynamics and glucose metabolism of *Schistosoma mansoni* adult worms in vitro.** *Parasites Vectors* **15**, 176 (2022). doi: 10.1186/s13071-022-05278-8