

RESEARCH ARTICLE

# The Epigenome of *Schistosoma mansoni* Provides Insight about How Cercariae Poise Transcription until Infection

David Roquis<sup>1,2</sup>✉, Julie M. J. Lepesant<sup>1,2</sup>✉, Marion A. L. Picard<sup>1,2</sup>✉, Michael Freitag<sup>3</sup>, Hugues Parrinello<sup>4</sup>, Marco Groth<sup>5</sup>, Rémi Emans<sup>1,2</sup>, Céline Cosseau<sup>1,2</sup>, Christoph Grunau<sup>1,2\*</sup>

1 Université de Perpignan Via Domitia, Perpignan, France, 2 CNRS, UMR 5244, Interactions Hôtes-Pathogènes-Environnements (IHPE), Perpignan, France, 3 Oregon State University, Corvallis, Oregon, United States of America, 4 CNRS, GenomiX IBIISA, Montpellier, France, 5 Leibniz Institute for Age Research—Fritz Lipmann Institute, Jena, Germany

✉ These authors contributed equally to this work.

\* [christoph.grunau@univ-perp.fr](mailto:christoph.grunau@univ-perp.fr)



**OPEN ACCESS**

**Citation:** Roquis D, Lepesant MJM, Picard MAL, Freitag M, Parrinello H, Groth M, et al. (2015) The Epigenome of *Schistosoma mansoni* Provides Insight about How Cercariae Poise Transcription until Infection. PLoS Negl Trop Dis 9(8): e0003853. doi:10.1371/journal.pntd.0003853

**Editor:** Sergio Costa Oliveira, Federal University of Minas Gerais, BRAZIL

**Received:** November 26, 2015

**Accepted:** May 27, 2015

**Published:** August 25, 2015

**Copyright:** © 2015 Roquis et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** ChIP-Seq and RNA-Seq fastq files are available at the NCBI SRA as studies SRP034587 and SRP035609, respectively. We also provide a Gbrowse interface at <http://genome.univ-perp.fr>

**Funding:** The study was funded by the Agence Nationale de la Recherche, contract ANR 2010 BLAN 1720 01 (EPIGEVOL) ([http://www.agence-nationale-recherche.fr/programmes-de-recherche/recherches-exploratoires-et-emergentes/blanc-generalite-et-contacts/blanc-presentation-synthetique-du-projet/?tx\\_lmwsuivibilan\\_pi2\[CODE\]=ANR-10-BLAN-1720](http://www.agence-nationale-recherche.fr/programmes-de-recherche/recherches-exploratoires-et-emergentes/blanc-generalite-et-contacts/blanc-presentation-synthetique-du-projet/?tx_lmwsuivibilan_pi2[CODE]=ANR-10-BLAN-1720)).

## Abstract

### Background

Chromatin structure can control gene expression and can define specific transcription states. For example, bivalent methylation of histone H3K4 and H3K27 is linked to poised transcription in vertebrate embryonic stem cells (ESC). It allows them to rapidly engage specific developmental pathways. We reasoned that non-vertebrate metazoans that encounter a similar developmental constraint (*i.e.* to quickly start development into a new phenotype) might use a similar system. Schistosomes are parasitic platyhelminthes that are characterized by passage through two hosts: a mollusk as intermediate host and humans or rodents as definitive host. During its development, the parasite undergoes drastic changes, most notable immediately after infection of the definitive host, *i.e.* during the transition from the free-swimming cercariae into adult worms.

### Methodology/Principal Findings

We used Chromatin Immunoprecipitation followed by massive parallel sequencing (ChIP-Seq) to analyze genome-wide chromatin structure of *S. mansoni* on the level of histone modifications (H3K4me3, H3K27me3, H3K9me3, and H3K9ac) in cercariae, schistosomula and adults (available at <http://genome.univ-perp.fr>). We saw striking differences in chromatin structure between the developmental stages, but most importantly we found that cercariae possess a specific combination of marks at the transcription start sites (TSS) that has similarities to a structure found in ESC. We demonstrate that in cercariae no transcription occurs, and we provide evidences that cercariae do not possess large numbers of canonical stem cells.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Conclusions/Significance

We describe here a broad view on the epigenome of a metazoan parasite. Most notably, we find bivalent histone H3 methylation in cercariae. Methylation of H3K27 is removed during transformation into schistosomula (and stays absent in adults) and transcription is activated. In addition, shifts of H3K9 methylation and acetylation occur towards upstream and downstream of the transcriptional start site (TSS). We conclude that specific H3 modifications are a phylogenetically older and probably more general mechanism, *i.e.* not restricted to stem cells, to poise transcription. Since adult couples must form to cause the disease symptoms, changes in histone modifications appear to be crucial for pathogenesis and represent therefore a therapeutic target.

## Author Summary

The blood fluke *Schistosoma mansoni* causes intestinal bilharzia. The parasite has a complex life cycle in which a freshwater snail serves as intermediate host from which the human infecting larvae hatch. These larvae will actively seek skin contact, penetrate through the epithelium and start developing straight away into adult worms. Development from larvae into adults needs thorough adjustment of gene expression through repositioning or modification of proteins that are associated with DNA (the chromatin). We decided to compare the chromatin of human infective larvae (cercariae), the first developmental stage after infection of the vertebrate host (schistosomula) and adults of *S. mansoni*. We found that cercariae possess chromatin structures (modifications of histone H3) around the beginning of genes that are very different from schistosomula and adults. We conclude that this structure serves to keep gene transcription in a poised state, *i.e.* transcription is initiated and can start immediately when the blocking histone modification is removed. A similar type of histone modification was found in embryonic stem cells of vertebrates and our data indicate that it is either a more ancient and/or more general means to poise transcription than previously assumed. Since many parasites possess infective stages that develop rapidly within the host, this particular chromatin structure could be a therapeutic target for a new class of antiparasitic drugs.

## Introduction

Histones are closely associated with DNA. Amino acids in their side-chains carry several modifications, such as methylation, acetylation and others. Specific combinations of histone modifications are associated with transcriptionally permissive or repressive chromatin structures, thus controlling gene expression at the transcriptional level [1]. Acetylation of histone H3 at lysine 9 (H3K9ac) is a mark associated with euchromatin and facilitation of transcription [2–6]. In model species the main enrichment of this mark is observed between the beginning of promoters and the 5' end of genes and sometimes in enhancers [2,3,7]. In contrast, methylation at the same position (H3K9me3) is a landmark of constitutive heterochromatin. It is often highly enriched at telomeric, pericentromeric and repeat-rich regions, and believed to be responsible for stable large-scale repression [8–11]. In *Plasmodium falciparum* for instance, it is found in subtelomeric and some small intrachromosomal regions containing almost exclusively clusters of genes involved in antigenic variations [6]. In human, it was shown that this

modification is mutually exclusive with trimethylation of histone H3 at lysine 4 (H3K4me3) on a given nucleosome [12]. H3K4me3 is often associated with transcription start sites (TSS), while trimethylation of H3 at lysine 27 is generally found in transcriptionally repressive heterochromatin. Consequently, the discovery of methylation at both sites (bivalent methylation) in several hundred genes of embryonic stem cells has attracted wide attention [13]. Embryonic stem cells (ESCs) are generated from the inner cell mass of mammalian blastocysts. These cells are self-renewing and can give rise to all lineages of the developing organism. The current view of bivalent methylation is that histone H3 trimethylated at lysine 27 (H3K27me3) represses transcription of lineage control genes during pluripotency, while H3K4me3 maintains them poised, *i.e.* transcription machinery is in position at the TSS, ready for activation upon reception of a signal that triggers differentiation (reviewed in [14]). When ESCs initiate differentiation, genes marked with bivalent trimethylation have two possible outcomes: H3K27me3 is removed and transcription occurs, or H3K4me3 is removed and genes stay silent. Since histones and histone modifications are extremely conserved through all kingdoms, we wondered if H3K9 methylation/acetylation and bivalent H3K4/K27 trimethylation might be a general mechanism by which a transient developmental state is maintained until development and differentiation into the next developmental state is triggered. Parasites are one of the models of choice to study this question since many of them have life cycles with very strong phenotypic differences between the stages, they are phylogenetically only distantly related to vertebrates and the environmental triggers are well-known, even though the molecular bases of switching to a new developmental stage are mostly unknown. Here we used the platyhelminth trematode *Schistosoma mansoni*. *S. mansoni* is the causative agent of intestinal bilharzia, a disease affecting 67 million people [15]. In contrast to many other parasites, the stage of the parasite infecting humans is actively seeking skin contact and penetrates the epidermis by use of cytolytic enzymes secreted from the pre- and post-acetabular glands and mechanical force. During skin penetration, cercariae lose their tails and undergo a drastic morphological and physiological transformation. Within two hours, the free-living larvae become obligatory endoparasitic schistosomula [16] that develop into adult worms, which reproduce in the blood stream of the host. Cercariae have therefore a double function: they are a vehicle for the parasites heritable information, but also true free-living animals with highly organized structures such as a nervous system and sensory organs, a digestive tract, excretory bladder and ducts, and a gland cell network. In addition to this, they have the capacity to develop very rapidly into full-grown adult worms, which are phenotypically very distinct from the cercariae.

The core histone isoforms of *S. mansoni* are canonical with 100% identity to human H4.A, 79–99% identity to H3 forms and 62–92% to H2A and H2B [17]. The distribution of histone modifications in *S. mansoni* was completely unexplored, and we decided to study the parasite epigenome by chromatin immunoprecipitation followed by sequencing (ChIP-Seq) of DNA associated with the above-mentioned histone isoforms H3K4me3, H3K9ac, H3K9me3 and H3K27me3. The genome of *S. mansoni* has been recently assembled and most of the gene structure data comes from automatic annotation [18]. Since gene prediction algorithms often miss the first exon (especially if they are untranslated), we performed RNA-Seq to unambiguously identify transcriptional start sites (TSS) and the expression status of genes.

## Methods

### Ethics statement

Housing, feeding and animal care followed the national ethical standards established in the writ of February 1st, 2013 (NOR: AGRG1238753A) setting the conditions for approval, planning and operation of establishments, breeders and suppliers of animals used for scientific

purposes and controls. The French Ministère de l'Agriculture et de la Pêche and French Ministère de l'Éducation Nationale de la Recherche et de la Technologie provided permit A 66040 to our laboratory for experiments on animals and certificate for animal experimentation (authorization 007083, decree 87–848) for the experimenters.

## Animal breeding and preparation of parasites

The *S. mansoni* strain *SmBre*, originally sampled in Brazil was used in this study. The strain was maintained in its sympatric intermediate host strain *BgBRE* of the mollusk *Biomphalaria glabrata*, and *Mus musculus* as definitive vertebrate host. Cercariae were collected from infected snails 30 days after infection by pipetting from spring water and sedimentation on ice. For isolation of adults, infected mice were killed after 60 days of infection by a lethal intraperitoneal injection of sodium pentobarbital, and adult worms were recovered by retrograde perfusions of the hepatic portal system with citrate (7.5%) saline (8.5%) solution administrated through the left ventricle [19]. Worms trapped in the liver or mesenteric system were collected after excising these organs.

*In-vitro* transformation of cercariae to schistosomula was done as described before [20]. The shells of infected snails were cleaned to remove rotifers, transferred to de-chlorinated water (about 75 mL for 10 snails), and exposed to a light source (60 W) for two hours in order to promote cercariae emission. Cercariae were gently collected with a 1 mL micro-pipet previously moistened, avoiding snail feces and mucus. They were transferred to a sterile 50 mL conical tube and placed on ice in the dark. After 30 minutes, cercariae were concentrated by centrifugation (400×g/10 minutes/4°C) and the supernatant was removed, leaving 5 mL of liquid. Five milliliters of phosphate buffered saline, supplemented with 200 U/mL of penicillin and 200 µg/mL of streptomycin (PBS-PS) were added and put at 37°C. The tails of cercariae were sheared by vortexing at maximum speed during one minute and subsequently centrifuged (400×g/1 minute/4°C). All further handling was done under sterile conditions. The supernatant, containing only tails, was completely removed and discarded. The pellet was resuspended in 2 mL PBS-PS. Schistosomula were separated from tails by centrifugation along a gradient of polyvinylpyrrolidone-coated colloidal silica particles (Percoll) [21] prepared in 15 mL conical tubes (6 mL of the Percoll [Sigma, P4937], 1 mL 10 fold PBS, 0.25 mL 100 mM HEPES, 0.25 mL penicillin/streptomycin, 2.5 mL sterile water). One milliliter of the schistosomula suspension was applied to each column without disturbing the surface of the Percoll gradient. The preparation was centrifuged at 450×g for 10 minutes, with weakest possible brake. The top layer containing the tails was removed, and the pellet containing schistosomula was resuspended and washed twice in 10 mL of Dulbecco's modified Eagle's medium (DMEM) with 200 U/mL of penicillin, 200 µg/mL of streptomycin and 10 mM HEPES at 37°C (centrifugation 400×g/1 minute/4°C). Finally, schistosomula were resuspended in modified Basch's medium without antimycotics [20,22] and maintained at 37°C under 5% CO<sub>2</sub> for 21 hours.

## RNA extraction and reverse transcription, quantitative PCR, and native chromatin immunoprecipitation

As the genome annotation of *S. mansoni* has mostly been done through prediction software, we performed RNA-Seq on cercariae to unambiguously identify TSS and exon-intron structures. Since published RNA extraction protocols for cercariae did not yield RNA suitable for Illumina sequencing, we developed the following method: 5,000 cercariae were ground with glass beads in liquid nitrogen for 5 minutes, resuspended in 0.5 mL TRIzol (Invitrogen), and 0.1 mL of chloroform was added. The tube was shaken vigorously for 1 minute and left at room temperature for 3 minutes, and was then centrifuged at ≥12,000×g for 15 minutes at 4°C. The

upper phase was transferred to a new RNase-free tube, an equal volume of 70% ethanol was added and the tube was vortexed. PureLink RNA Mini kit (Ambion, 12183020) was used for further purification following the manufacturer's protocol. RNA was recovered in 30  $\mu$ L RNasecure (Ambion, AM7006), left for 1 minute at room temperature, collected by centrifugation for 1 minutes at 8,000xg at room temperature and incubated at 65°C for 10 minutes. To remove remaining DNA, 3  $\mu$ L of 10X TURBO DNase Buffer (TURBO DNA-free, Ambion, M1907) were added to 30  $\mu$ L of total RNA in RNasecure, followed by 3  $\mu$ L of TURBO DNase (2 Units/ $\mu$ L). The preparation was gently mixed and incubated at 37°C for 30 minutes. DNase was removed with an RNA Clean-up system (RNeasy mini kit, QIAGEN, 74104) and eluted in 30  $\mu$ L RNase-free water. For schistosomula, total RNA extraction, DNase treatment and purification methods were adapted from the protocol described above. 500 ng of this purified total RNA were reverse transcribed using identical concentrations (250 nmol/L) of random hexamers and oligo(dT)<sub>18</sub> primers of Maxima H Minus Reverse Transcriptase kit (Thermo-SCIENTIFIC). Quantitative PCR was performed using a LightCycler 480 System (Roche Diagnostics). We used 2.5  $\mu$ L of cDNA (diluted 1/10 in water) in a total volume of 10  $\mu$ L containing 1X LightCycler 480 SYBR Green I Master Mix (Roche Diagnostics) and 500 nmol/L of each primer. To ensure that a single product was generated, we separated amplification products on a Labchip GX DNA assay system (PerkinElmer). For each studied stage, the level of mRNA was normalized using the mean geometric transcription rate of three reference sequences (Smp\_093230, Smp\_197220 and Smp\_089880) previously described by Wang and colleagues [23]. mRNA of Smp\_034000 was measured using primers Smp034000.2F: GCGTGCTGTGG AGGAAAGCGA and Smp034000.2R: ACGCAGACAGGGCATCGTCA. The normalized relative quantities (NRQs) were calculated as described by Hellemans and colleagues [24], using equation 1 where  $E_{\text{target}}$  is the amplification efficiency of the gene of interest;  $E_{\text{ref}}$  is the amplification efficiency of the reference gene;  $\Delta C_{p,\text{ref}} = C_{p,\text{ref}}(\text{calibrator}) - C_{p,\text{ref}}(\text{sample})$ ; and  $\Delta C_{p,\text{target}} = C_{p,\text{target}}(\text{calibrator}) - C_{p,\text{target}}(\text{sample})$ . For each reaction, the crossing point (Cp) was determined using the second derivative maximum method applied using Light Cyclor Software version 3.3 (Roche Diagnostics). Male cercariae were arbitrarily chosen as calibrator. We obtained male cercariae by infecting host mollusk *B. glabrata* strain BgBre with single miracidium from SmBre strain. Sex determination was performed as described in [25]. Experiments were done in six biological replicates.

Native chromatin immunoprecipitation was done according to an earlier published protocol [26] and at least 1,500 cercariae, 2000 schistosomula or 1 adult couple using the following antibodies: H3K4me3 (Millipore, cat# 04-745 lot# NG1680351, 4  $\mu$ L per reaction), H3K9Ac (Millipore, cat# 07-352 lot# DAM16924924, 8  $\mu$ L per reaction), H3K9me3 (Abcam, cat# ab8898 lot# 733951, 4  $\mu$ L per reaction), and H3K27me3 (Diagenode, cat# pAb-069-050 lot# A29900242, and cat# C15410069 lot# A1821D, 8  $\mu$ L per reaction). For each sample, we used a control without antibody, to assess unspecific background (bound fraction) and input (unbound fraction). Inputs were used for normalization in all subsequent bioinformatics analyses. Experiments were done in duplicate. Antibodies were carefully tested for specificity as described in [27] and saturating quantities (see [27]) were used. Further details are available at <http://methdb.univ-perp.fr/epievo/>

### Illumina library generation and sequencing

For ChIPseq library construction, the ChIP sample preparation kit (Illumina Inc., USA) was used according to the manufacturer's protocol except for the indexed adapters. Those adapters came from the Truseq DNA sample preparation kit (Illumina Inc., USA). Briefly, 30 ng of DNA per condition was blunt ended and adenylated on its 3' ends. Illumina's indexed adapters

were ligated to both ends of the adenylated DNA. Ligated DNA was size separated by electrophoresis through a 2% agarose gel and a size selection was performed at 400 bp. Once extracted from the agarose, size selected DNA was enriched by 18 cycles of PCR for 10 seconds at 98°C, 30 seconds at 60°C and 30 seconds at 72°C using Illumina's proprietary primers and Phusion DNA polymerase (New England Biolabs Inc., USA). Each library was verified using a DNA 1000 Chip on a Bioanalyzer 2100, quantified by real time PCR with the KAPA Library Quantification Kit for Illumina Sequencing Platforms (Kapa Biosystems Ltd, SA) and then diluted to 10 nM.

RNAseq library construction was done with the TruSeq RNA Sample Seq kit (Illumina Inc., USA) according to the manufacturer's protocol. Briefly, poly-A containing mRNA molecules were purified from 70 to 450 ng of total RNA using oligo-dT magnetic beads. The purified mRNA was fragmented by addition of the fragmentation buffer and was heated at 94°C in a thermocycler for 8 minutes. First strand cDNA was synthesized using random primers. Second strand cDNA synthesis, end repair, A-tailing, and adapter ligation was done in accordance with the manufacturer's instructions. Purified cDNA templates were enriched by 15 cycles of PCR for 10 seconds at 98°C, 30 seconds at 65°C, and 30 seconds at 72°C using Illumina's proprietary primers and Phusion DNA polymerase. Each indexed cDNA library was verified using a DNA 1000 Chip on a Bioanalyzer 2100 and quantified by real time PCR with the KAPA Library Quantification Kit for Illumina Sequencing Platforms (Kapa Biosystems Ltd, SA) then diluted to 10 nM using Illumina's resuspension buffer.

For each lane of sequencing, four libraries (ChIPseq) or 6 libraries (RNAseq) were pooled in equimolar concentrations, denatured using 0.1N NaOH and diluted to a final concentration of 7 pM (ChIPseq) or 8 pM (RNAseq) using Illumina's HT1 buffer. 120 µL of the dilution was then transferred into a 200 µL strip tube and placed on ice before loading onto the cBot. The flow cell was clustered using Single Read Cluster Generation Kit (Illumina Inc., USA), according to the Illumina SR\_amplification\_Linearization\_Blocking\_PrimerHyb recipe. The flow cell was loaded onto the Illumina HiSeq 2000 instrument following the manufacturer's instructions and sequencing was performed with the 50 cycles, single read, indexed protocol. Image analyses and base calling were performed using the HiSeq Control Software (HCS) and Real-Time Analysis component (RTA). Demultiplexing was performed using CASAVA (Illumina). The quality of the data was assessed using fastqc from the Babraham Institute and the Illumina software SAV (Sequence Analysis Viewer).

## Illumina sequencing quality control, alignment and ChIP-Seq peak calling

All data treatment was carried out under a local galaxy instance [28] (<http://galaxy.univ-perp.fr>). Fastq Groomer was used for verification of the fastqsanger format, and the FASTX-Toolkit (Compute quality statistics, Draw quality score boxplot, Draw nucleotides distribution chart) was used for initial quality control. Read quality was judged sufficiently good (the majority of reads showed a quality score  $\geq 24$  for all positions) and no further quality filter was applied.

For ChIP-Seq, reads were aligned to the *S. mansoni* reference genome, assembly version 5.0 (downloaded from <ftp://ftp.sanger.ac.uk/pub/pathogens/Schistosoma/mansoni/genome/Assembly-v5/>) using Bowtie2 [29] evoking parameters—sensitive-k 2. Mapping quality (mapq) in Bowtie2 is related to “uniqueness” of the read. The higher the quality score is, the lower is the probability that a read can map elsewhere in the genome. By filtering mapped reads with a score of 255 (the highest possible) in samtools [30] (samtools view-hS-q 255), we kept reads with a single match on the genome. In Bowtie 2.1 score 255 was abandoned and filtering of uniquely mapped reads was done with “samtools view-Sh-q *quality value* 40–42-F 0x0004 - |

grep -v XS:i. “. SAM alignment files were converted into the bed format with pyicos [31] and sorted with sortBed-i of the bedtools suite [32]. For peak identification an equal number of random lines in the bed-file was chosen for each biological replicate (Table 1). Although comparative analysis software usually include a normalization step to handle the different numbers of mapped reads in conditions or replicates when comparing samples, we found that the random sampling step drastically reduces the amount of false positives without lowering the sensitivity. Mostly, this removes background peaks. Identification of peaks was done with ranger of Peak-ranger v1.16 [33] with P value cut off 0.0001, FDR cut off 0.01, Read extension length 200, Smoothing bandwidth 99 and Delta 1. We used the input samples as negative controls for the peakcalling (-c). For data visualization, wiggle files were generated and uploaded to a local GBrowse instance [34] (<http://genome.univ-perp.fr>).

ChIP-Seq results in adults and cercariae were verified by qPCR (primers in S1 Table) in 20 randomly chosen regions using the following conditions: denaturation 95°C for 10 seconds, annealing 60°C for 10 seconds, extension 72°C for 20 seconds, over 40 cycles. The kit used for the amplification is the LightCycler 480 SYBR Green 1 Master (Roche, 04887352001) and the instrument is a LightCycler 480 (Roche). To compare ChIP-Seq and qPCR results, coverage was calculated with coverageBed of the bedtools suite for the regions whose boundaries were defined by the corresponding qPCR primer positions and divided by the coverage in the alpha-tubulin reference locus (Smp\_scaff000423:21,107–21,289). Spearman rank correlation gave values > 0.8, R2 was 0.64 for H3K4me3 and 0.59 for H3K27me3 indicating that both methods gave comparable results (S1 Fig).

For RNA-Seq, reads were aligned to the genome with TopHat v1.4.1 with default parameters and exon-intron structure was reconstructed with cufflinks v1.3.0 evoking -q -I 40000 -F 0.100000 -j 0.150000 -p 8 on each library and data were merged and linked to the gene prediction of the Sanger Centre (<ftp://ftp.sanger.ac.uk/pub/pathogens/Schistosoma/mansoni/genome/Assembly-v5/>) with Cuffmerge v1.0.0 [35].

## Comparative ChIP-Seq analysis

To identify peaks that were different between cercariae and adults, we used the DiffBind library 1.6.1 [36] available on Bioconductor [37] for R version 2.15.1. As input, we used the randomly sampled aligned reads that also served as inputs for PeakRanger and the regions file generated with PeakRanger, all in bed format. Diffbind allows for replicates and “input” negative control to be taken into account to search for differential peak distribution between conditions. For H3K9ac, PCR duplicates were removed with pyicos, which improved sample autocorrelation. For H3K4me3 all differences were inspected on GBrowse. For the other histone isoforms, one hundred regions identified by DiffBind were randomly sampled with a custom Perl script and verified by visual inspection. Average histone modification profiles around TSS were generated by doing a “window analysis” around the TSS of genes, using EpiChIP v0.9.7-e [38]. EpiChIP allows for generating genome-wide, average profiles for histone modifications at a given genomic feature (in our case, around the TSS for genes). For this purpose, we used the output file of Cuffmerge and selected only those 12,871 genes located on linkage groups that were assembled at the chromosome level. Chromosome names were converted in the input bed files into Epi-ChIP canonical names (“Chrnumber”) with a custom script. EpiChIP was also used to correlate transcription and histone modification, and to quantify the proportion of peaks that contribute significantly to the average profiles. Common peaks were detected from Peakranger region bed-files with multiIntersectBed of the BedTools package.

For the analysis of chromatin structure around repetitive sequences, reads were aligned to a fasta file containing all 3,145 known repeats of *S. mansoni* as consensus sequences (available

**Table 1. Summary of next-generation sequencing data and analysis results.**

| Condition  | H3K4Me3    |           |               | H3K9Ac   |            | H3K9Me3    |            | H3K27Me3   |            |               |
|--|------------|-----------|---------------|--|------------|------------|------------|------------|------------|---------------|
|  | cercariae  | adults    | schistosomula | cercariae  | adults     | cercariae  | adults     | cercariae  | adults     | schistosomula |
| <b>Number of samples</b>   | 2          | 2         | 2             | 2  | 2          | 2          | 2          | 3          | 3          | 2             |
| <b>Input reads (average per sample)</b>                              | 34,519,715 | 6,421,317 | 49,109,097    | 54,833,141   | 43,884,060 | 53,702,142 | 43,884,060 | 40,895,894 | 33,865,527 | 46,158,137    |
| <b>Alignment to unique sequences</b>                                 |            |           |               |  |            |            |            |            |            |               |
| Aligned exactly 1 time (average per sample)                          | 21,455,086 | 3,097,077 | 28,275,831    | 28,626,785   | 28,035,327 | 23,326,061 | 21,945,541 | 22,882,234 | 27,123,469 | 30,651,658    |
| Aligned reads used for peak-calling (per sample)                     | 2,900,000  | 2,900,000 | 2,900,000     | 23,000,000   | 23,000,000 | 21,000,000 | 21,000,000 | 17,900,000 | 17,900,000 | 17,900,000    |
| Regional peaks identified by peakranger 1.16 (average for condition) | 7,746      | 8,868     | 7,376         | 17,116   | 421        | 128,518    | 100,884    | 145,322    | 27,544     | 65,629        |
| Significantly different peaks identified by DiffBind                 | 224        |           |               | 3,844  |            | 9,505      |            | 8,480      |            |               |
| DiffBind result confirmed by visual inspection                       | 124        |           |               | random sample of 100 peaks for each condition(33–50% confirmed by 2 independent observers) |            |            |            |            |            |               |
| EpiChIP peaks with signal at FDR 0.01 (per condition)                | 3,598      | 3,566     | 3,954         |  |            |            |            | 4,801      | 4,370      | 5,024         |
| <b>Alignment to repetitive sequences</b>                             |            |           |               |  |            |            |            |            |            |               |
| Aligned exactly 1 time (average per sample)                          | 1,326,620  | 242,000   |               | 3,805,162  | 3,091,945  | 3,881,193  | 3,612,100  | 3,999,829  | 4,554,097  |               |
| Random sample used for DESeq   | 220,000    | 220,000   |               | 2,900,000  | 2,900,000  | 3,400,000  | 3,400,000  | 3,900,000  | 3,900,000  |               |

doi:10.1371/journal.pntd.0003853.t001

<http://methdb.univ-perp.fr/downloads/>) [39] using Bowtie2 [29] evoking parameter -k 2 and parameter preset “sensitive”. Unique matches were filtered using samtools [30] (mapping quality mapq = 255). In order to obtain identical effective library sizes for each sample, the same number of aligned reads was randomly sampled for each of the ChIP experiments (Table 1). After the counts had been normalized, they were compared between samples using the DESeq



package of the R software [40]. This package uses the negative binomial distribution to calculate the P-value for each element between two compared samples and find elements that present significant differences in their read enrichment levels.

### *In-situ* detection of replication and transcription

For the detection of mRNA synthesis, cercariae were incubated in spring water at 28°C with a concentration of 1mM of 5-ethynyl-uridine (EU, Invitrogen Click-iT RNA Imaging Kit, MP10329) for 2, 18 h and 21 h. Freshly transformed schistosomula were grown in Basch medium at the same EU concentration in an incubator at 37°C and 5% CO<sub>2</sub>, also for 2, 18, and 21 h. Labeling of replicating cells was performed in the exact same conditions, only substituting EU by a concentration of 10 mM of 5-ethynyl-2'-deoxyuridine (EDU, Invitrogen Click-iT DNA Imaging Kit, MP10337). Negative controls without the addition of 5-ethynyl-uridine (EU) and 5-ethynyl-2'-deoxyuridine (EDU) were also performed for every condition. Fixation, development and Hoechst 33342 coloration after incubation were performed following the protocols of the supplier. Imaging was done on a confocal microscope Zeiss LSM 700 at 488 nm (EU and EDU) and 405 nm (Hoechst 33342). For every condition, 100 animals were observed for transcription and replication.

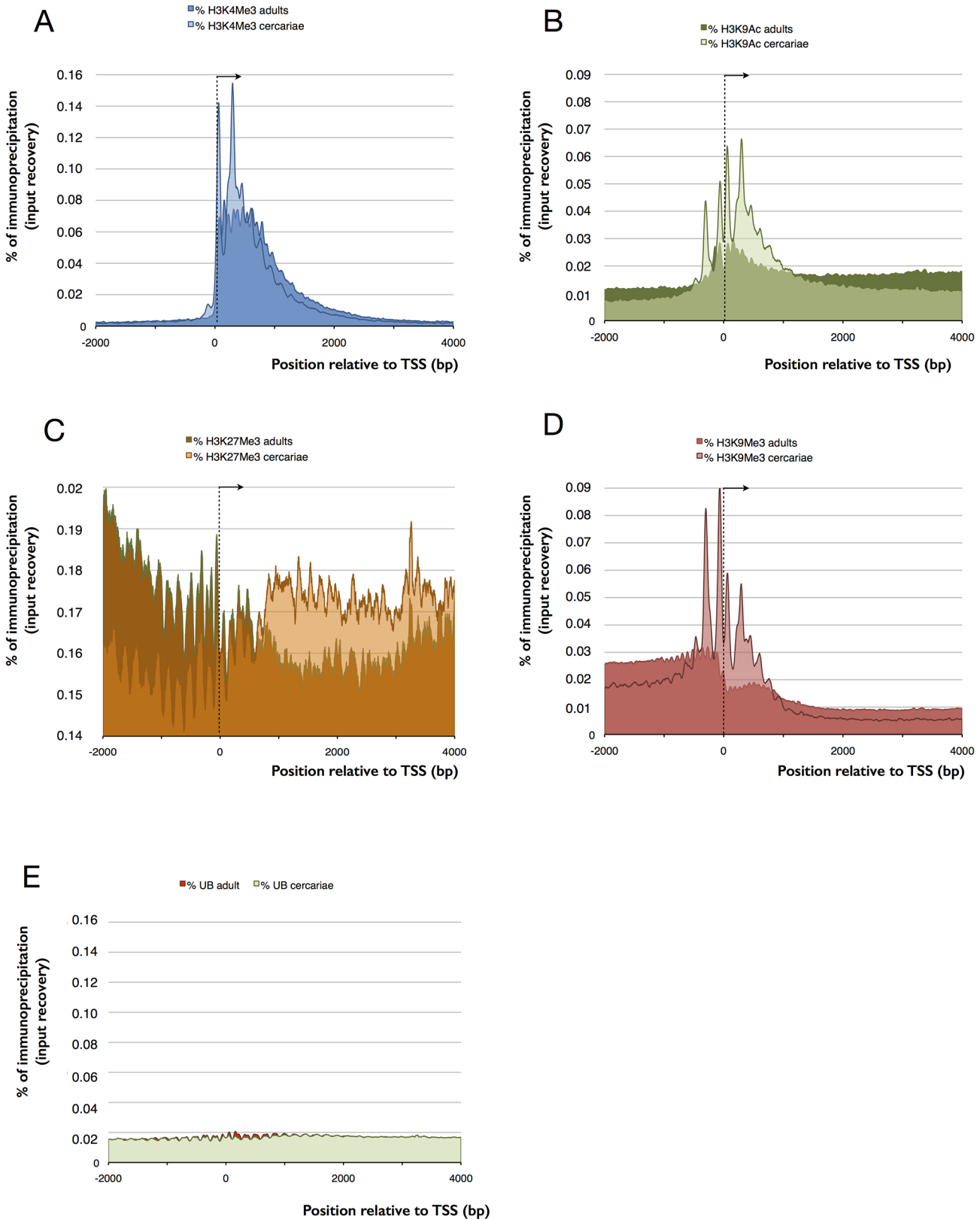
### Accession numbers

RNA-Seq and ChIP-Seq reads are available at the NCBI-SRA under study accession numbers SRP034587 (ChIP-Seq) and SRP035609 (RNA-Seq).

## Results

### Chromatin landscape in *S. mansoni* cercariae and adults

To link transcription and chromatin structure, we considered the correct identification of the TSS as crucial. Anecdotal evidences had suggested that the available automatic annotation had missed a certain proportion of first exons and we decided to perform RNA-Seq with RNA extracted from cercariae. Indeed, our data showed that in about 20% of the genes, TSS had to be corrected. We then performed ChIP-Seq. For H3K4me<sub>3</sub>, we found relatively sharp peaks with a mean peak maximum located 250 bp (or 1–2 nucleosomes) downstream of the transcription start site (TSS) of genes (Fig 1A). We detected on average 7,746 peaks in cercariae and 8,868 peaks in adults (Table 1). For about ~24% of genes, EpiChIP identified transcripts but no significant H3K4me<sub>3</sub> peaks (FDR = 0.01). Examples for genes that are free of H3K4me<sub>3</sub> peaks are all known micro exon genes (MEGs) supplementary table S5.4 in [41] (Fig 2A), with the exception of Smp\_163710 that shows small peaks in adults. MEGs contain unusually small exons of 3–36 bp that are repeated in tandem [42]. Their function remains elusive but it is thought that hypervariable proteins can be generated by MEGs through a 'pick and mix' splicing strategy. It is conceivable that such types of genes have a specific control mechanism. Other H3K4me<sub>3</sub>-peak free genes are polycistronic genes that are located downstream of the trans-splicing site. Protasio *et al.* estimated trans-splicing events to occur in 11% of *S. mansoni* genes [18] and had compiled a list of putative polycistronic transcripts. We used this list and show that in 143 (89.9%) genes, the H3K4me<sub>3</sub> peak is located exclusively in the gene upstream of the trans-splicing acceptor site (Fig 2B), in 2 (1.2%) cases peaks are present in both genes (upstream and downstream), and in 13 (8.2%) genes we did not find any peak, although in 3 cases we suspect that an upstream gene with a peak may be the first of the polycistronic unit (S2 Table). Our data are in line with the view that these genes are indeed transcribed as one unit from a single TSS (that has the H3K4me<sub>3</sub> signal) and fragmented by trans-splicing. For

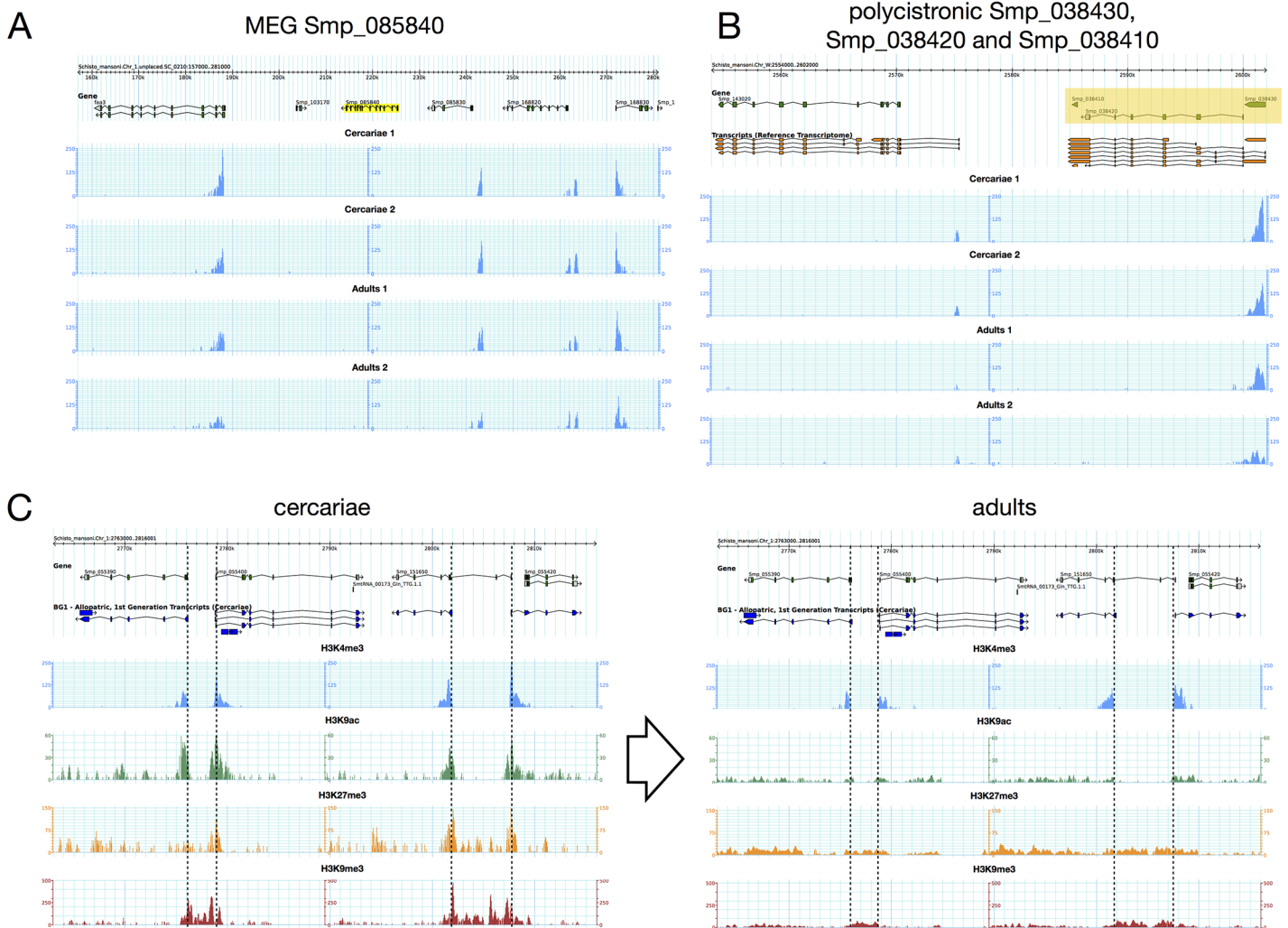


**Fig 1. Comparison of EpiChIP profiles 2kb upstream and 4kb downstream of the TSS (average for 12,871 genes located on linkage groups that are assembled on chromosome level) for cercariae (light color) and adults (dark color). TSS and direction of transcription indicated by an arrowhead. X-**

axis in bp, y-axis in relative enrichment (% of all mapped reads). (A) H3K9me3, (B) H3K9ac, (C) H3K27me3, (D) H3K9me3, and (E) unbound fractions UB of mock ChIP without antibody that corresponds to input that shows no difference between cercariae and adults and no enrichment at the TSS.

doi:10.1371/journal.pntd.0003853.g001

H3K27me3 we found with Diagenode antibody cat# pAb-069-050 lot# A29900242 enrichment at the TSS (-500 to +1,000 bp), while it was reduced to background level in adults. Since there was anecdotic evidence for cross-reactivity of this antibody with H3K4me3, we repeated the experiment with cat# C15410069 (that replaces pAb-069-050) lot# A1821D. With this antibody we also observe enrichment of H3K27me3 in cercariae compared to adults, but shifted towards the body of the gene (Fig 1C and Fig 2C). Methylation and acetylation in H3K9 are

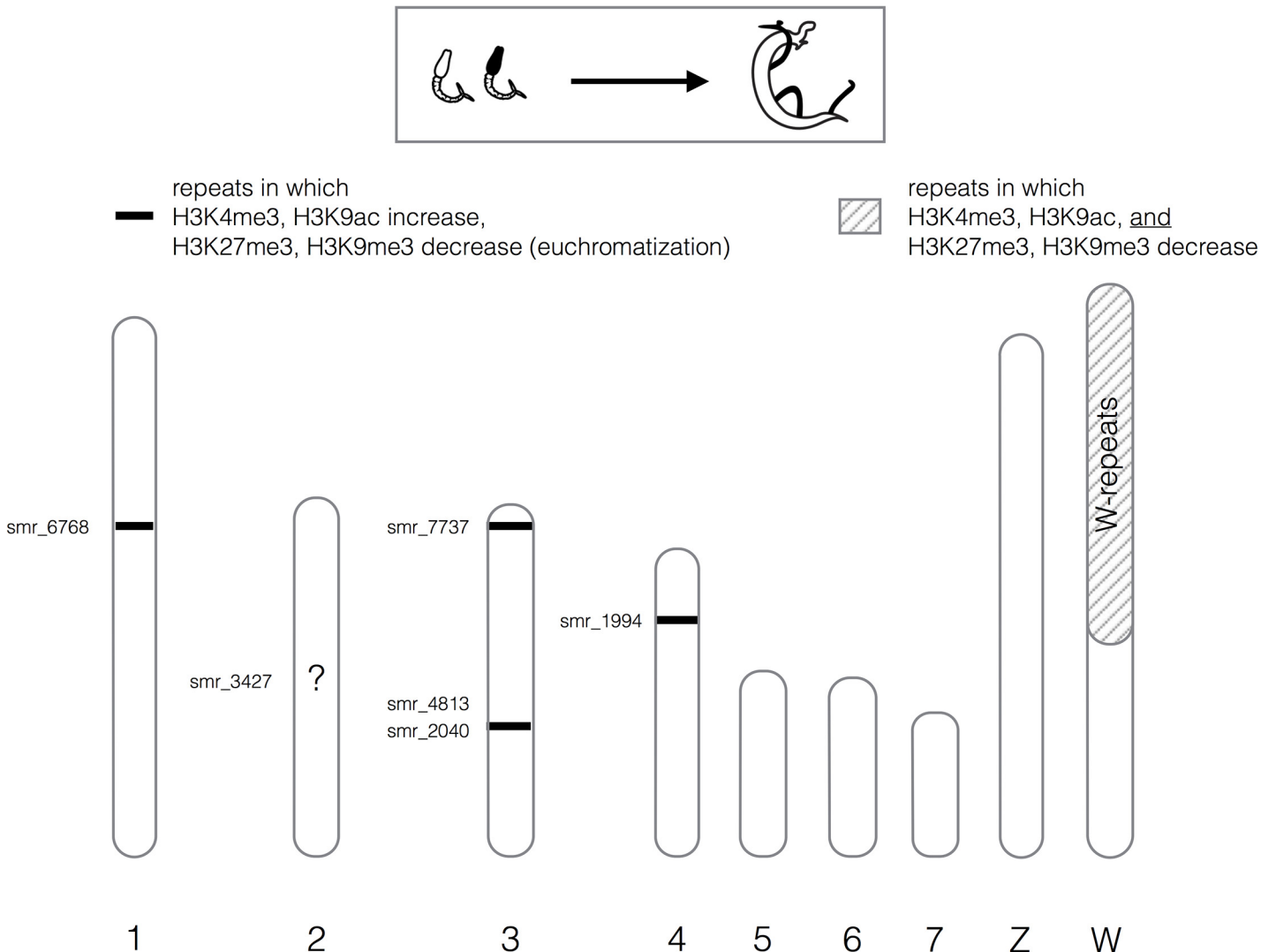


**Fig 2. Typical examples of chromatin profiles.** (A) H3K4me3 peaks are absent in the TSS of MEG Smp\_085840 (highlighted in yellow) while surrounding genes show characteristic peaks in all 4 samples. (B) Polycistronic genes Smp\_038430, Smp\_038420 and Smp\_038410 (from right to left, highlighted in yellow) show an H3K4me3 peak only in the first gene. (C) Changes in chromatin structure in cercariae (left) and adults (right). On top predicted genes, below RNA-Seq based exon/intron structure, followed by H3K4Me3 (blue), H3K9Ac (green), H3K27Me3 (yellow) and H3K9Me3 (red), identical scale for cercariae and adults. Positions of TSS that were detected by RNA-Seq are indicated by a dotted line.

doi:10.1371/journal.pntd.0003853.g002

mutually exclusive at the same histone and H3K9me3 is considered a repressive mark. H3K9me3 shows a peak about 200 bp upstream of the TSS in cercariae, which is reduced to an upstream background in adults (Fig 1D and Fig 2C). H3K9ac is a hallmark of active transcription. In cercariae we detected this modification primarily around the TSS (-500 to +1,000 bp) while in adults acetylation spans the body of the genes (Fig 1B and Fig 2C). Taken together, our results indicate that all four studied histone modifications show an unusual enrichment around the TSS in cercariae, while in adults their distribution is similar to that found in other metazoans. We saw no differences between the input fractions (negative controls used for normalization) of cercariae and adults (Fig 1E).

In cercariae and adults, both H3K4me3 and H3K9ac showed globally low association with repeats ( $4.10 \pm 0.43\%$  and  $6.45 \pm 0.65\%$ , respectively of the total reads aligned) (details in Table 1). For H3K27me3 and H3K9me3 the percentage was higher with  $7.60 \pm 1.00\%$  and  $7.75 \pm 0.60\%$ . DESeq [40] identified repeats that change chromatin structure during transformation from cercariae to adults. DESeq uses a negative binomial distribution to model read counts that can be assigned to a locus. To estimate variance within a sample it pools the data from genes with similar hit counts, thus allowing variance estimation and identification of significant differences based on a small number of biological replicates [40]. We used DESeq because we had earlier compared it to edgeR [43], and DESeq2 and in our hands it gave the most consistent results. Using a p-value of  $< 10^{-5}$  as the selection criterion, we identified 17 (0.54%) repeats that change their chromatin status for H3K4me3, 182 (5.79%) for H3K9ac, 263 (8.36%) for H3K9me3, and 40 (1.27%) for H3K27me3. Results are summarized in S3 Table. Interestingly, not all repeats were concerned equally: in roughly two thirds of these repeats ( $n = 363$ , 11.5%) more than one histone isoform had changed meaning that about 10% of all repeat families undergo important chromatin structure changes during the transformation of cercariae into adults. Two results were particularly interesting (Fig 3). The first is a chromatin change in five regions that are composed of tandemly repeated stretches of class I LTR retrotransposons and class II DNA transposons. These regions are located on the large *S. mansoni* chromosomes (1 to 4), but there might be similar regions on the three smaller chromosomes that were overlooked. Each of the regions is composed of a different repeat, but they all belong to the same repeat class (*i.e.* transposon). In these five regions, enrichment of histone forms H3K9me3 and H3K27me3 decreases during cercariae to adult metamorphosis, while enrichment of H3K9ac and H3K4me3 increases. Considering five states (H3K4me3 up, H3K9ac up, H3K9me3 down, H3K27me3 down, transposon or other), the individual term binomial distribution probability to obtain such a combination by chance is  $3.7^{-36}$ . Our data suggest therefore that the process is not randomly distributed along the genome. Stretches of (retro)transposons could form heterochromatic knobs in the cercariae and their change in chromatin structure could be an indication of large scale chromatin rearrangements or chromosome displacements in the nucleus during transformation into adults. Another peculiar observation during transformation from cercariae to adults is the nearly complete loss of H3 modifications in 22 of the 36 W-specific repeats (61%). We had shown earlier that transcription of these repeats can only be observed during larval stages and is undetectable in adults [44]. From qPCR data for a small subset we had already postulated that these repeats acquire an “adult-specific” chromatin structure. This view is confirmed by the new NGS data: H3K9/K4/K27 methylation and H3K9 acetylation are lost. Cytogenetic images suggest that the repeat containing large pericentromeric regions becomes heterochromatic when cercariae develop in the snail, but no clear data is available for adult worms [45]. It is tempting to speculate that H3 is replaced by another H3 isoform similar to what is observed in centromeres and neo-centromeres of other species for instance [46]. A homologue of the centromere specific form of H3, CENP-A, was detected in *S. mansoni*



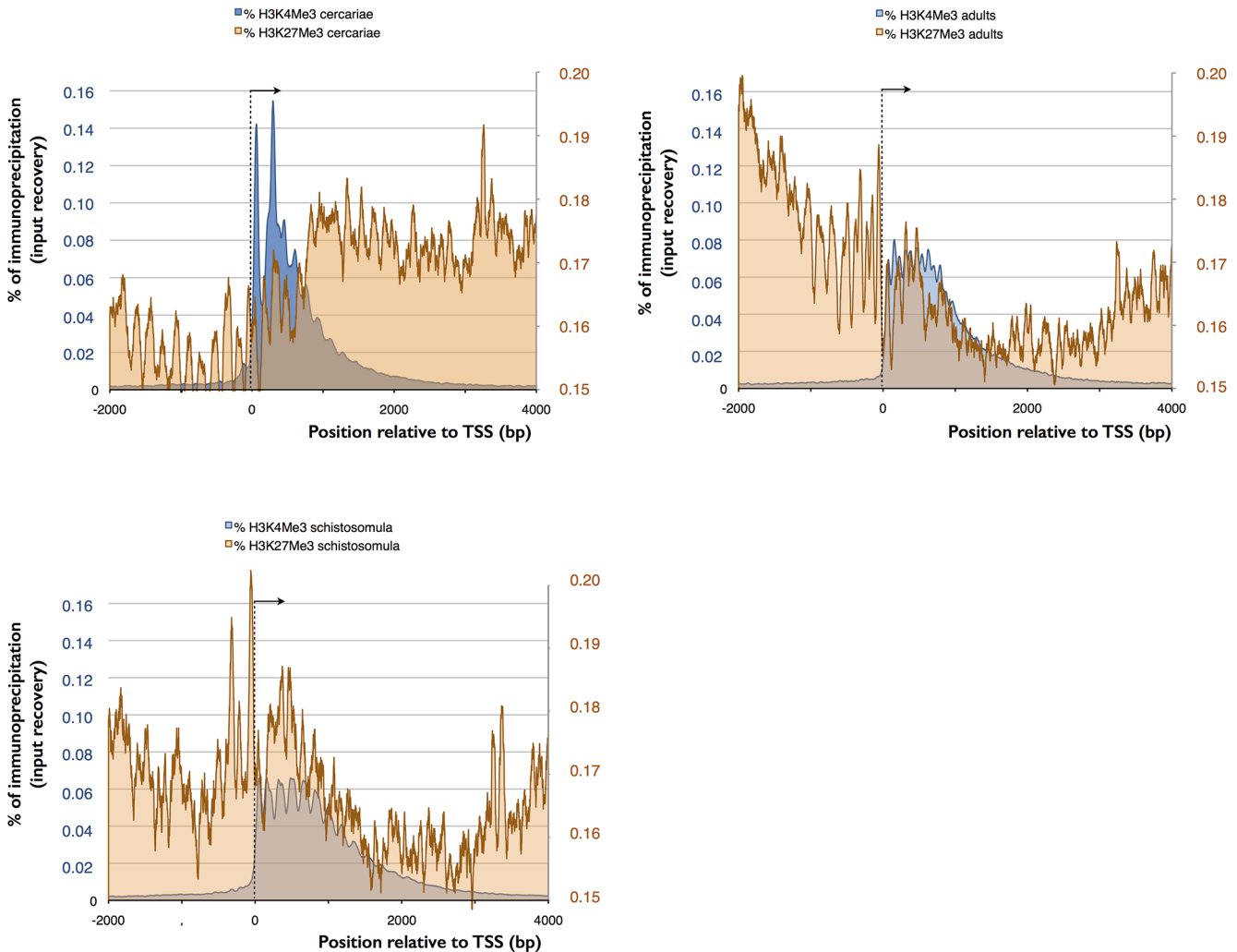
**Fig 3. Schematic representation of chromatin structure changes in tandem-repeat regions during metamorphosis of cercariae into adults.** Autosomes 1–7 are labeled with numbers, sex chromosomes with letters. Repeat regions are indicated on each chromosome as inferred by alignment to the reference genome, or previously by FISH on the W-chromosome. Repeat class smr\_3427 is presumably on chromosome 2 but the exact position is unknown (unpositioned contig).

doi:10.1371/journal.pntd.0003853.g003

(Smp\_040020) [47] but homology with CENP-A of other species is vague and basically restricted to the core H3/H4 domain.

### Bivalent trimethylation of H3K4 and H3K27 is specific for a subset of genes in cercariae

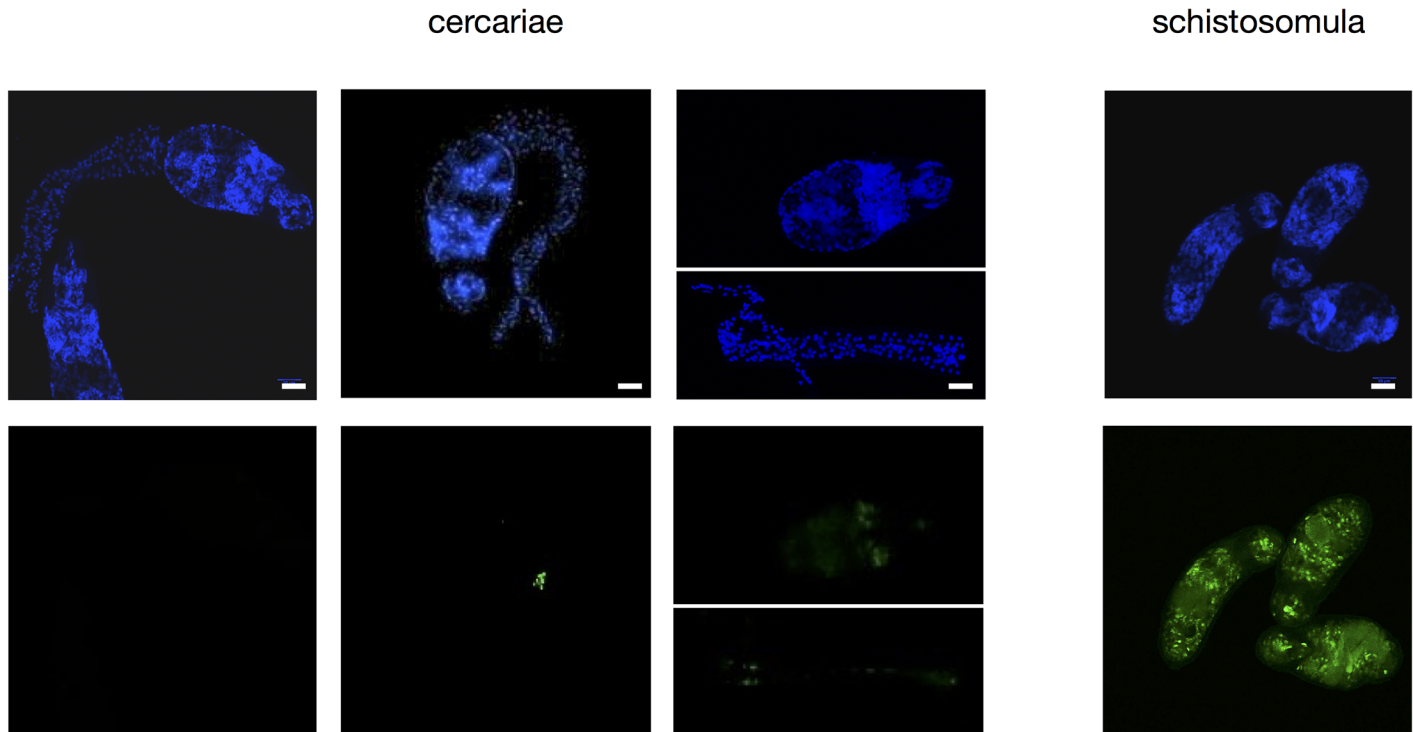
While PeakRanger identify 1,122 more peaks in adults than cercariae for H3K4me3, Diffbind considers only 224 differences to be statistically significant. After visual inspection, we confirmed differential enrichment in only 124 regions along the genome with 95 (77%) hypermethylated in cercariae and 29 (23%) in adults. This means that less than 2% of H3K4me3 peaks are different, while gene transcription patterns of cercariae and adults are clearly distinct [48]. A peculiar feature of the *S. mansoni* epigenome was revealed when we analyzed the



**Fig 4. Comparison of EpiChIP profiles 2 kb upstream and 4 kb downstream of the TSS (average for 12,871 genes located on linkage groups that are assembled on chromosome level) between different histone isoforms in cercariae (left up), schistosomula (left down) and adults (right). TSS and direction of transcription indicated by an arrowhead. Bivalent methylation of H3K4 and H3K27 occurs in cercariae but not in adults, left y-axis for H3K4me3, right y-axis for H3K27me3.**

doi:10.1371/journal.pntd.0003853.g004

distribution of the heterochromatic H3K27me3. In cercariae, H3K27 trimethylation is high in the gene body (Fig 4). H3K27me3 and H3K4me3 are also statistically significantly enriched around the TSS in 150 genes. This number was reduced to 121 after visual inspection (S4 Table). Thirty-five days of development separate cercariae and full-grown adults worms. To determine more precisely the time window during which chromatin is remodeled we performed ChIP-Seq on schistosomula. Our results show that the adult chromatin structure is already established in schistosomula by 18 h after infection (Fig 4). Most importantly, in schistosomula, bivalent methylation is absent in the 121 genes that were significantly enriched in cercariae.



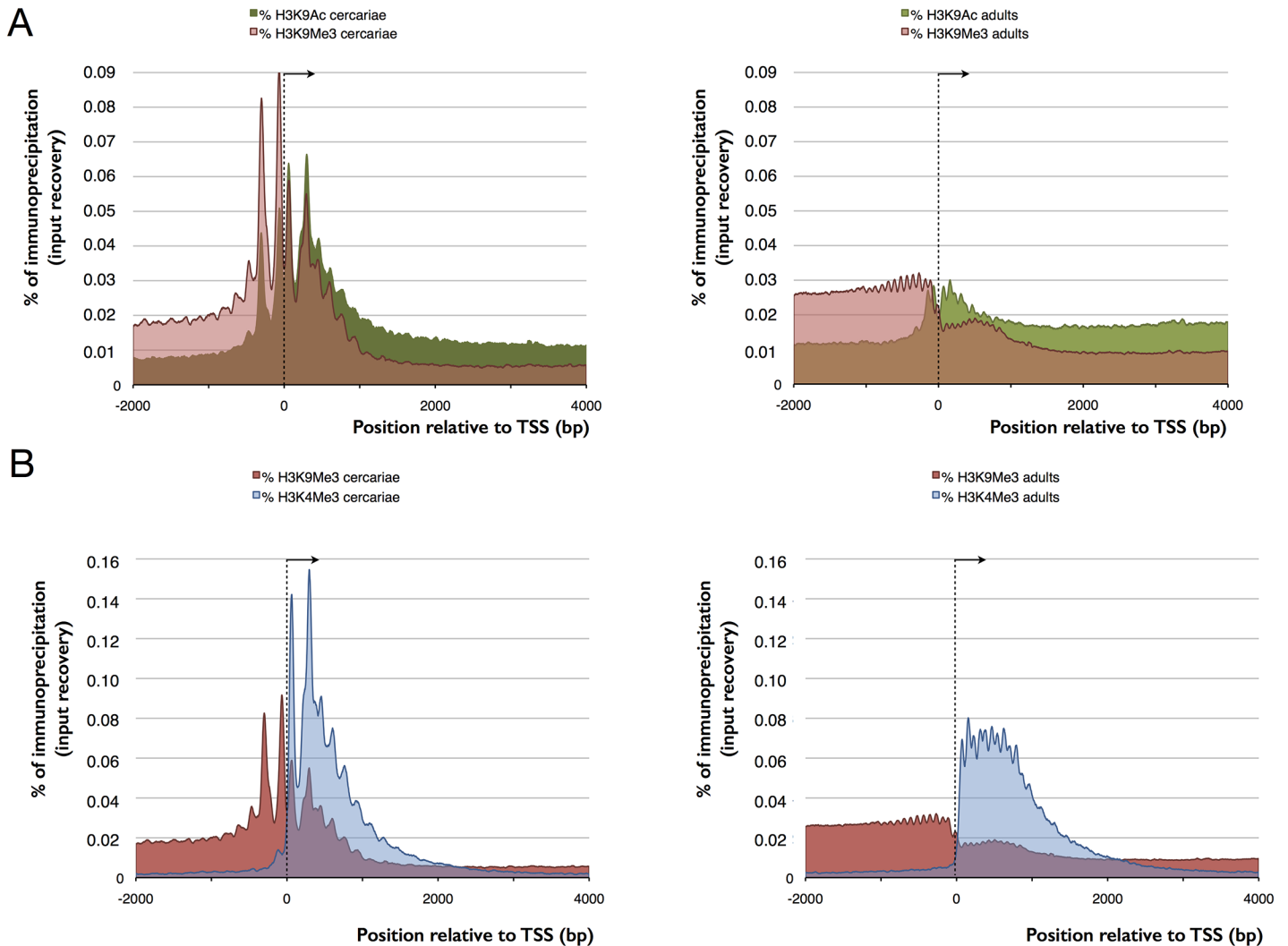
**Fig 5. Detection of RNA synthesis by incorporation of EU by fluorescence microscopy.** Top panel (blue) Hoechst 33342 staining of DNA, bottom panel (green) EU incorporation and detection with Alexa Fluor 488. White bar 20 μm.

doi:10.1371/journal.pntd.0003853.g005

### Cercariae are transcriptionally silent

While RNA is clearly present in cercariae, the co-localization of all four histone modifications at the TSS made us wonder if transcription is indeed ongoing or if it is poised. Cercariae are short lived and they do not grow outside the host. It is therefore conceivable that their RNA was produced during their development in the snail. We used incorporation of the cytosine analog 5-ethynyl-uridine (EU) and fluorescence labeling to detect active transcription *in vivo*. At 28°C, the maximum life span of a cercaria is 32 h (90% of mortality occurs past 26 h [49]) but infection capacity decreases rapidly and probably does not exceed 6 h. We decided therefore to use a relatively long EU pulse of 18 h. Despite this, we detected fluorescence signals in less than 10% of the cercariae. With the exception of a single individual (out of one hundred) signals were detected in only a few of the roughly 1,000 cells of an individual larva, and were always restricted to the nucleus (examples shown in Fig 5). In contrast, schistosomula, the first developmental stage after infection of the vertebrate host, showed consistent staining in the 100 arbitrarily chosen individuals and staining was more dispersed. We conclude that there is very little transcription in cercariae and that transcription is rapidly activated upon transformation into schistosomula. Our observation that in cercariae the repressive mark H3K9me3 is slightly upstream and the permissive mark H3K9ac is slightly downstream but close to the TSS is in accordance with stalled transcription. In contrast, in adults H3K9 acetylation and methylation are spread out with H3K9ac occurring along the gene body and H3K9me3 in the upstream regions separating genes (Fig 6A).

Bivalent H3K4/H3K27 trimethylation and absence of transcription are reminiscent of what occurs in ESC of vertebrates, *e.g.* mice [13], human [50] and zebrafish [51]. There, bivalent



**Fig 6. Comparison of EpiChIP profiles 2 kb upstream and 4 kb downstream of the TSS (average for 12,871 genes located on linkage groups that are assembled on chromosome level) between different histone isoforms in cercariae (left) and adults (right).** TSS and direction of transcription indicated by an arrowhead. (A) methylation of H3K9 upstream of TSS and acetylation downstream, concentrated around TSS in cercariae and spanning the gene body (H3K9ac) and the intergenic region (H3K9me3) in adults, (B) methylation of H3K9 and H3K4 are mutually exclusive in adults but overlap in cercariae.

doi:10.1371/journal.pntd.0003853.g006

H3K4/H3K27 trimethylation occurs at the transcription start site of lineage-specific genes and it was proposed that it keeps them transcriptionally “poised”, with the transcription machinery in place, ready to start transcription upon removal of the H3K27me3 mark (reviewed in [51]). ESC do not exist in schistosomes, but totipotent somatic stem cells (neoblasts) were found in platyhelminthes from the free-living turbellarians (for review [52]) to the endo- or ecto-parasitic flatworms (monogenea, cestoda, trematoda) [53]. Recently, somatic stem cells were detected in adult schistosomes and called proliferating somatic cells (PSCs) [23,54]. Another type of stem cells, totipotent germinal cells, is present in the intramolluscan stages of *S. mansoni* but they are restricted to a small germinal cell cluster in cercariae [23]. Chromatin structure profiles for these cells have not been studied but it could be that cercariae possess simply a



large number of ESC-like cells. The ultrastructure of cercariae was described in great detail in the past [55,56] and apart from the above-mentioned small population of germinal cells, no stem cells were detected. Stem cells could be identified thanks to the expression of characteristic marker genes such as vasa, piwi, tudor and nanos. However, piwi and tudor-domain containing genes are absent from flukes and tapeworms and no true vasa orthologue has been identified in schistosomes [41]. Thus, we sought to exploit rapid replication as another characteristic feature of these cells to identify them in schistosomes.

### No DNA replication in cercariae

We used 5-ethynyl-2'-deoxyuridine (EDU) incorporation to detect replication in cercariae and schistosomula. While the latter showed at least one replicating cell per individual in 27% (36 out of 135) after 2 h EDU pulse and 86% (156 out of 181) after 18 h, in none of the cercariae did we detect replication. Consequently, cercariae do not contain large amounts of rapidly replicating ES-like cells or neoblasts that could explain the ChIP-Seq profiles.

### Discussion

We describe here the epigenome of *S. mansoni* using four major histone modifications. We analyzed unique sequences, in particular those containing genes, but also the repetitive compartment of the genome.

The enrichment of histone forms around TSS shares similarities with other metazoans, but also shows particularities. Similar to what is found in model organisms, H3K4me3 is enriched at transcription start sites (TSS) and promoter regions of genes with actual or potential transcription [2,5,57–59]. In mammalian genomes, it marks active or poised (when combined with a repressive mark) high CpG-content promoters, while is only present at low CpG-content promoters when the latter are active. Further analysis will show if this is also the case in schistosomes and for what reason some TSS are devoid of H3K4me3. H3K9ac is another mark associated with euchromatin and facilitation of transcription [2–6]. The main enrichment of this mark is often observed between the beginning of promoters and the 5' end of genes, and sometimes in enhancers [2,3,7]. In *S. mansoni* there is a clear difference between cercariae and adults with H3K9ac spanning the gene body only in the latter. H3K9me3 is a landmark of constitutive heterochromatin. In many organisms it is highly enriched at telomeric, pericentromeric and repeat-rich regions, and is responsible for stable large-scale repression [8–11]. In humans, it was shown that this modification is mutually exclusive with H3K4me3 [12]. This is also the case in adult schistosomes (Fig 6B). Nevertheless, H3K9me3 can be present at a very high level over transcribed regions of active mammalian genes, especially 1.5 kb after the TSS, and at a low levels at the promoter [60]. H3K9me3 increases during activation of transcription and is quickly removed afterwards [60]. In our study, association of H3K9me3 with the TSS is only observed in cercariae, where this mark is predominately located upstream of the gene. In *S. mansoni*, a combination of H3K9me3 and H3K9ac allows the prediction of the direction of transcription (always from the H3K9me3 peak to the H3K9ac peak) and it seems that blocks of H3K9me3 separate promoters of genes that are transcribed back-to-back (example shown in Fig 2C). This might be a signature of protein-coding genes. The fourth studied histone isoform, H3K27me3 is known to maintain facultative heterochromatin and is involved in the regulation and sequential expression/repression of developmental genes [11,61]. It is present at inactive or poised (when combined with H3K4me3) high CpG-content promoters and is anti-correlated with the level of transcription [9,10]. In some unicellular organisms, such as *Plasmodium falciparum* this mark can even be absent [6]. In *S. mansoni*, the mark is weak in genes of schistosomula and adults, but present in cercariae. Similar to ESC, bivalent trimethylation of H3K4

and H3K27 is present in at least 121 genes. Bivalent methylation is often considered as an ESC specific mark that leads to “poised” transcription that can rapidly be activated when the cells enter into differentiation. Our data indicate that there are no rapidly proliferating stem cells in cercariae, and we show that there is no transcriptional activity. Consequently, neoblasts of cercariae are either in a dormant state, or somatic cells of cercariae use the same mechanism as ESC to poise transcription that is rapidly activated after infection, upon transformation into schistosomula (at least during the first 18 h after penetration in the host). Bivalent H3K4/H3K27 trimethylation could therefore be a phylogenetically very old and/or more general mechanism to poise transcription than currently thought. However, as only 121 genes were detected to have these bivalent marks, another mechanisms could be involved in the transcription silencing observed in cercariae. At the moment, we can only hypothesize that the major shift of H3K27me3 enrichment (from gene body in cercariae to upstream of TSS in schistosomula and adults, [Fig 4](#)) and/or the strong enrichment of H3K9me3 and H3K9ac at the TSS of cercariae play a role in this phenomenon.

While the importance of repetitive sequences for genome structure and evolution is generally accepted, they are often considered too difficult to analyze by genome-wide approaches. Almost half of the *S. mansoni* genome is composed of repeats [39] and we deemed it to be important to include these in our analysis. Here we used a straightforward approach to evaluate chromatin structure in all known repeat classes of *S. mansoni*. We show that in five mini-satellite regions composed of retrotransposons H3K4me3 and H3K9ac increase while H3K9me3 and H3K27me3 decrease during the transition from cercariae into adults. It is difficult to evaluate the functional importance of this chromatin structure change. It was shown in cultured human cells that another type of satellite DNA (gamma satellites) is enriched for H3K4me3, and that it shields adjacent genes from heterochromatization [62]. Acetylation of H3K9 was not investigated in this latter study. A similar effect might occur in our system but further work will be necessary to elucidate the function of these repeat regions.

In summary, we present here the epigenomes of *S. mansoni* for three life cycle stages. We show that in cercariae, bivalent H3K4/H3K27 trimethylation exists. To the best of our knowledge, this bivalent trimethylation has, in the animal kingdom, so far only been found in vertebrate ESCs. We also show that TSS occupancy by H3K4me3 is stable between cercariae and adults in the large majority of genes. It changes drastically for H3K9 trimethylation and acetylation. In addition, we identified at least five repeat locations that gain euchromatic status during development into adults. Our findings support a model in which transcription is poised in cercariae due to the presence of both activating and repressive marks. It is reactivated, presumably through the removal of the repressive methylation at the TSS and in the gene body, and in conjunction with large scale chromatin remodeling (indicated by changes in repeat regions). In this model, histone-modifying enzymes, in particular H3K27 and H3K9 demethylases would act within 18 h after skin penetration to release the poised state and allow for rapid activation of the transcription machinery that is already in place around the TSS. We had shown before that histone-modifying enzymes are indeed essential for transformation and infection success of another life cycle stage, the miracidia [63]. Others have shown that inhibition of HDAC8 leads to parasite death [64]. In zebrafish, a H3K27 demethylase of the KDM6 type (UTX) is needed for fin regeneration [65]. An orthologue of UTX is encoded in the *S. mansoni* genome (Smp\_034000) and by q-RT-PCR we found that mRNA for this gene is present in cercariae, all four stages of schistosomula *in vivo*, and in male and female adults. Further studies will show if blocking this candidate will result in developmental arrest. Besides being of academic interest, such enzymes would also be promising new drug candidates, especially in the light of the fact that the only drug currently available to cure schistosomiasis, praziquantel, is ineffective against the larval forms [66]. In addition, many parasites possess infective life cycle stages

that develop immediately after infection. Our findings should motivate search for characteristic histone modifications in these stages.

## Supporting Information

**S1 Table. (a) Primers targeting regions with bivalent H3K4me3 and H3K27me3 marks in cercariae, and monovalent H3K4me3 marks in adults. (b) Primers targeting regions with H3K4me3 differences between cercariae and adults.**

(DOCX)

**S2 Table. H3K4me3 peaks in putative polycistrons with a maximum intergenic distance of 200bp.** Upstream indicates the transcript that is located upstream of the trans-splicing acceptor site. Down-stream indicates the putative trans-spliced transcript. Polycistron list from <http://www.plosntds.org/article/info%3Adoi%2F10.1371%2Fjournal.pntd.0001455#s5> (S2

Table).

(XLS)

**S3 Table. DESeq results for ChIP-Seq on repetitive sequences.**

(XLSX)

**S4 Table. List of transcripts for which TSS has H3K4me3 and H3K27me in cercariae, evidence from GeneDB (for Smp\_#####) and from AmiGO v.1.8 (for TCONS\_#####).**

(PDF)

**S1 Fig. Relative enrichment compared to the alpha-tubulin locus: ChIP-Seq vs qPCR, p-value is given for the t-test of the null hypothesis that the corresponding slope is equal to zero against the two-sided alternative.**

(TIFF)

## Acknowledgments

The authors are grateful for technical support by Anne Rognon, Jean-François Allienne, Nathalie Aranciabia, Rémi Emans, Bernard Dejean and Claire Picard. We thank Ivonne Heinze for her diligent and skillful technical assistance in Illumina sequencing. We also thank the French National Research Agency for providing financial support to this project (ANR INVIMORY).

## Author Contributions

Conceived and designed the experiments: CG. Performed the experiments: DR JM JL MALP MF HP MG. Analyzed the data: DR CC CG. Contributed reagents/materials/analysis tools: RE. Wrote the paper: DR JM JL MALP CC CG.

## References

1. Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature* 403: 41–45. PMID: [10638745](#)
2. Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM et al. (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 122: 517–527. PMID: [16122420](#)
3. Nishida H, Suzuki T, Kondo S, Miura H, Fujimura Y et al. (2006) Histone H3 acetylated at lysine 9 in promoter is associated with low nucleosome density in the vicinity of transcription start site in human cell. *Chromosome Res* 14: 203–211. PMID: [16544193](#)
4. Koch CM, Andrews RM, Flicek P, Dillon SC, Karaoz U et al. (2007) The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Res* 17: 691–707. PMID: [17567990](#)

5. Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC et al. (2011) Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* 471: 480–485. doi: [10.1038/nature09725](https://doi.org/10.1038/nature09725) PMID: [21179089](https://pubmed.ncbi.nlm.nih.gov/21179089/)
6. Hoeijmakers WA, Stunnenberg HG, Bartfai R (2012) Placing the *Plasmodium falciparum* epigenome on the map. *Trends Parasitol* 28: 486–495. doi: [10.1016/j.pt.2012.08.006](https://doi.org/10.1016/j.pt.2012.08.006) PMID: [22999479](https://pubmed.ncbi.nlm.nih.gov/22999479/)
7. Hon GC, Hawkins RD, Ren B (2009) Predictive chromatin signatures in the mammalian genome. *Hum Mol Genet* 18: R195–201. doi: [10.1093/hmg/ddp409](https://doi.org/10.1093/hmg/ddp409) PMID: [19808796](https://pubmed.ncbi.nlm.nih.gov/19808796/)
8. Martens JH, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M et al. (2005) The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J* 24: 800–812. PMID: [15678104](https://pubmed.ncbi.nlm.nih.gov/15678104/)
9. Liu T, Rechtsteiner A, Egelhofer TA, Vielle A, Latorre I et al. (2011) Broad chromosomal domains of histone modification patterns in *C. elegans*. *Genome Res* 21: 227–236. doi: [10.1101/gr.115519.110](https://doi.org/10.1101/gr.115519.110) PMID: [21177964](https://pubmed.ncbi.nlm.nih.gov/21177964/)
10. Zhou VW, Goren A, Bernstein BE (2011) Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* 12: 7–18. doi: [10.1038/nrg2905](https://doi.org/10.1038/nrg2905) PMID: [21116306](https://pubmed.ncbi.nlm.nih.gov/21116306/)
11. Bannister AJ, Kouzarides T (2011) Regulation of chromatin by histone modifications. *Cell Res* 21: 381–395. doi: [10.1038/cr.2011.22](https://doi.org/10.1038/cr.2011.22) PMID: [21321607](https://pubmed.ncbi.nlm.nih.gov/21321607/)
12. Binda O, LeRoy G, Bua DJ, Garcia BA, Gozani O et al. (2010) Trimethylation of histone H3 lysine 4 impairs methylation of histone H3 lysine 9: regulation of lysine methyltransferases by physical interaction with their substrates. *Epigenetics* 5: 767–775. doi: [10.4161/epi.5.8.13278](https://doi.org/10.4161/epi.5.8.13278) PMID: [21124070](https://pubmed.ncbi.nlm.nih.gov/21124070/)
13. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ et al. (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125: 315–326. PMID: [16630819](https://pubmed.ncbi.nlm.nih.gov/16630819/)
14. Vastenhouw NL, Schier AF (2012) Bivalent histone modifications in early embryogenesis. *Curr Opin Cell Biol* 24: 374–386. doi: [10.1016/j.ceb.2012.03.009](https://doi.org/10.1016/j.ceb.2012.03.009) PMID: [22513113](https://pubmed.ncbi.nlm.nih.gov/22513113/)
15. Hotez PJ, Brindley PJ, Bethony JM, King CH, Pearce EJ et al. (2008) Helminth infections: the great neglected tropical diseases. *J Clin Invest* 118: 1311–1321. doi: [10.1172/JCI34261](https://doi.org/10.1172/JCI34261) PMID: [18382743](https://pubmed.ncbi.nlm.nih.gov/18382743/)
16. Brink LH, McLaren DJ, Smithers SR (1977) *Schistosoma mansoni*: a comparative study of artificially transformed schistosomula and schistosomula recovered after cercarial penetration of isolated skin. *Parasitology* 74: 73–86. PMID: [320543](https://pubmed.ncbi.nlm.nih.gov/320543/)
17. Anderson L, Pierce RJ, Verjovski-Almeida S (2012) *Schistosoma mansoni* histones: from transcription to chromatin regulation; an in silico analysis. *Mol Biochem Parasitol* 3: 105–114.
18. Protasio AV, Tsai IJ, Babbage A, Nichol S, Hunt M et al. (2012) A Systematically Improved High Quality Genome and Transcriptome of the Human Blood Fluke *Schistosoma mansoni*. *PLoS Negl Trop Dis* 6: e1455. doi: [10.1371/journal.pntd.0001455](https://doi.org/10.1371/journal.pntd.0001455) PMID: [22253936](https://pubmed.ncbi.nlm.nih.gov/22253936/)
19. Duvall RH, DeWitt WB (1967) An improved perfusion technique for recovering adult schistosomes from laboratory animals. *Am J Trop Med Hyg* 16: 483–486. PMID: [4952149](https://pubmed.ncbi.nlm.nih.gov/4952149/)
20. Mann VH, Morales ME, Rinaldi G, Brindley PJ (2010) Culture for genetic manipulation of developmental stages of *Schistosoma mansoni*. *Parasitology* 137: 451–462. doi: [10.1017/S0031182009991211](https://doi.org/10.1017/S0031182009991211) PMID: [19765348](https://pubmed.ncbi.nlm.nih.gov/19765348/)
21. Lazdins JK, Stein MJ, David JR, Sher A (1982) *Schistosoma mansoni*: rapid isolation and purification of schistosomula of different developmental stages by centrifugation on discontinuous density gradients of Percoll. *Exp Parasitol* 53: 39–44. PMID: [6276213](https://pubmed.ncbi.nlm.nih.gov/6276213/)
22. Basch PF (1981) Cultivation of *Schistosoma mansoni* in vitro. I. Establishment of cultures from cercariae and development until pairing. *J Parasitol* 67: 179–185. PMID: [7241277](https://pubmed.ncbi.nlm.nih.gov/7241277/)
23. Wang B, Collins JJ, Newmark PA (2013) Functional genomic characterization of neoblast-like stem cells in larval *Schistosoma mansoni*. *Elife* 2: e00768. doi: [10.7554/eLife.00768](https://doi.org/10.7554/eLife.00768) PMID: [23908765](https://pubmed.ncbi.nlm.nih.gov/23908765/)
24. Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* 8: R19. PMID: [17291332](https://pubmed.ncbi.nlm.nih.gov/17291332/)
25. Portela J, Grunau C, Cosseau C, Beltran S, Dantec C et al. (2010) Whole-genome in-silico subtractive hybridization (WISH)—using massive sequencing for the identification of unique and repetitive sex-specific sequences: the example of *Schistosoma mansoni*. *BMC Genomics* 11: 387. doi: [10.1186/1471-2164-11-387](https://doi.org/10.1186/1471-2164-11-387) PMID: [20565937](https://pubmed.ncbi.nlm.nih.gov/20565937/)
26. Cosseau C, Azzi A, Smith K, Freitag M, Mitta G et al. (2009) Native chromatin immunoprecipitation (N-ChIP) and ChIP-Seq of *Schistosoma mansoni*: Critical experimental parameters. *Molecular & Biochemical Parasitology* 166: 70–76.
27. Cosseau C, Grunau C (2011) Native chromatin immunoprecipitation. *Methods Mol Biol* 791: 195–212. doi: [10.1007/978-1-61779-316-5\\_15](https://doi.org/10.1007/978-1-61779-316-5_15) PMID: [21913081](https://pubmed.ncbi.nlm.nih.gov/21913081/)

28. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L et al. (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Res* 15: 1451–1455. PMID: [16169926](#)
29. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9: 357–359. doi: [10.1038/nmeth.1923](#) PMID: [22388286](#)
30. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079. doi: [10.1093/bioinformatics/btp352](#) PMID: [19505943](#)
31. Althammer S, Gonzalez-Vallinas J, Ballare C, Beato M, Eyraes E (2011) Pyicos: a versatile toolkit for the analysis of high-throughput sequencing data. *Bioinformatics* 27: 3333–3340. doi: [10.1093/bioinformatics/btr570](#) PMID: [21994224](#)
32. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842. doi: [10.1093/bioinformatics/btq033](#) PMID: [20110278](#)
33. Feng X, Grossman R, Stein L (2011) PeakRanger: a cloud-enabled peak caller for ChIP-seq data. *BMC Bioinformatics* 12: 139. doi: [10.1186/1471-2105-12-139](#) PMID: [21554709](#)
34. Stein LD, Mungall C, Shu S, Caudy M, Mangone M et al. (2002) The generic genome browser: a building block for a model organism system database. *Genome Res* 12: 1599–1610. PMID: [12368253](#)
35. Trapnell C, Roberts A, Goff L, Pertea G, Kim D et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7: 562–578. doi: [10.1038/nprot.2012.016](#) PMID: [22383036](#)
36. Stark R, Brown G (2013) DiffBind: differential binding analysis of ChIP-Seq peak data. *bioconductor/jp*
37. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5: R80. PMID: [15461798](#)
38. Hebenstreit D, Gu M, Haider S, Turner DJ, Lio P et al. (2011) EpiChIP: gene-by-gene quantification of epigenetic modification levels. *Nucleic Acids Res* 39: e27. doi: [10.1093/nar/gkq1226](#) PMID: [21131282](#)
39. Lepesant JM, Roquis D, Emans R, Cosseau C, Arancibia N et al. (2012) Combination of de novo assembly of massive sequencing reads with classical repeat prediction improves identification of repetitive sequences in *Schistosoma mansoni*. *Exp Parasitol* 130: 470–474. doi: [10.1016/j.exppara.2012.02.010](#) PMID: [22381218](#)
40. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11: R106. doi: [10.1186/gb-2010-11-10-r106](#) PMID: [20979621](#)
41. Tsai IJ, Zarowiecki M, Holroyd N, Garciarrubio A, Sanchez-Flores A et al. (2013) The genomes of four tapeworm species reveal adaptations to parasitism. *Nature* 496: 57–63. doi: [10.1038/nature12031](#) PMID: [23485966](#)
42. Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP et al. (2009) The genome of the blood fluke *Schistosoma mansoni*. *Nature* 460: 352–358. doi: [10.1038/nature08160](#) PMID: [19606141](#)
43. Gao D, Kim J, Kim H, Phang TL, Selby H et al. (2010) A survey of statistical software for analysing RNA-seq data. *Hum Genomics* 5: 56–60. PMID: [21106489](#)
44. Lepesant JM, Cosseau C, Boissier J, Freitag M, Portela J et al. (2012) Chromatin structure changes around satellite repeats on the *Schistosoma mansoni* female sex chromosome suggest a possible mechanism for sex chromosome emergence. *Genome Biol* 13: R14. doi: [10.1186/gb-2012-13-2-r14](#) PMID: [22377319](#)
45. Grossman AI, Short RB, Cain GD (1981) Karyotype evolution and sex chromosome differentiation in Schistosomes (Trematoda, Schistosomatidae). *Chromosoma* 84: 413–430. PMID: [7327052](#)
46. Chueh AC, Wong LH, Wong N, Choo KH (2005) Variable and hierarchical size distribution of L1-retroelement-enriched CENP-A clusters within a functional human neocentromere. *Hum Mol Genet* 14: 85–93. PMID: [15537667](#)
47. Mourao MM, Grunau C, Loverde PT, Jones MK, Oliveira G (2012) Recent advances in *Schistosoma* genomics. *Parasite Immunol* 34: 151–162. doi: [10.1111/j.1365-3024.2011.01349.x](#) PMID: [22145587](#)
48. Fitzpatrick JM, Peak E, Perally S, Chalmers IW, Barrett J et al. (2009) Anti-schistosomal intervention targets identified by lifecycle transcriptomic analyses. *PLoS Negl Trop Dis* 3: e543. doi: [10.1371/journal.pntd.0000543](#) PMID: [19885392](#)
49. Theron A, Pointier J-P (1985) Recherche des facteurs susceptibles d'empêcher la réalisation du cycle de *Schistosoma mansoni* dans les mares de la grande terre de Guadeloupe (Antilles françaises). *Annales de parasitologie humaine et comparée* 60:
50. Pan G, Tian S, Nie J, Yang C, Ruotti V et al. (2007) Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* 1: 299–312. doi: [10.1016/j.stem.2007.08.003](#) PMID: [18371364](#)

51. Vastenhouw NL, Zhang Y, Woods IG, Imam F, Regev A et al. (2010) Chromatin signature of embryonic pluripotency is established during genome activation. *Nature* 464: 922–926. doi: [10.1038/nature08866](https://doi.org/10.1038/nature08866) PMID: [20336069](https://pubmed.ncbi.nlm.nih.gov/20336069/)
52. Baguna J (2012) The planarian neoblast: the rambling history of its origin and some current black boxes. *Int J Dev Biol* 56: 19–37. doi: [10.1387/ijdb.113463jb](https://doi.org/10.1387/ijdb.113463jb) PMID: [22252540](https://pubmed.ncbi.nlm.nih.gov/22252540/)
53. Brehm K (2010) *Echinococcus multilocularis* as an experimental model in stem cell research and molecular host-parasite interaction. *Parasitology* 137: 537–555. doi: [10.1017/S0031182009991727](https://doi.org/10.1017/S0031182009991727) PMID: [19961652](https://pubmed.ncbi.nlm.nih.gov/19961652/)
54. Collins JJr, Newmark PA (2013) It's no fluke: the planarian as a model for understanding schistosomes. *PLoS Pathog* 9: e1003396. doi: [10.1371/journal.ppat.1003396](https://doi.org/10.1371/journal.ppat.1003396) PMID: [23874195](https://pubmed.ncbi.nlm.nih.gov/23874195/)
55. Collins JJr, Wang B, Lambrus BG, Tharp ME, Iyer H et al. (2013) Adult somatic stem cells in the human parasite *Schistosoma mansoni*. *Nature* 494: 476–479. doi: [10.1038/nature11924](https://doi.org/10.1038/nature11924) PMID: [23426263](https://pubmed.ncbi.nlm.nih.gov/23426263/)
56. Dorsey CH, Cousin CE, Lewis FA, Stirewalt MA (2002) Ultrastructure of the *Schistosoma mansoni* cercaria. *Micron* 33: 279–323. PMID: [11742750](https://pubmed.ncbi.nlm.nih.gov/11742750/)
57. Berger SL (2007) The complex language of chromatin regulation during transcription. *Nature* 447: 407–412. PMID: [17522673](https://pubmed.ncbi.nlm.nih.gov/17522673/)
58. Gu SG, Fire A (2010) Partitioning the *C. elegans* genome by nucleosome modification, occupancy, and positioning. *Chromosoma* 119: 73–87. doi: [10.1007/s00412-009-0235-3](https://doi.org/10.1007/s00412-009-0235-3) PMID: [19705140](https://pubmed.ncbi.nlm.nih.gov/19705140/)
59. Zentner GE, Henikoff S (2013) Regulation of nucleosome dynamics by histone modifications. *Nat Struct Mol Biol* 20: 259–266. doi: [10.1038/nsmb.2470](https://doi.org/10.1038/nsmb.2470) PMID: [23463310](https://pubmed.ncbi.nlm.nih.gov/23463310/)
60. Vakoc CR, Mandat SA, Olenchok BA, Blobel GA (2005) Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. *Mol Cell* 19: 381–391. PMID: [16061184](https://pubmed.ncbi.nlm.nih.gov/16061184/)
61. Trojer P, Reinberg D (2007) Facultative heterochromatin: is there a distinctive molecular signature? *Mol Cell*
62. Kim JH, Ebersole T, Kouprina N, Noskov VN, Ohzeki J et al. (2009) Human gamma-satellite DNA maintains open chromatin structure and protects a transgene from epigenetic silencing. *Genome Res* 19: 533–544. doi: [10.1101/gr.086496.108](https://doi.org/10.1101/gr.086496.108) PMID: [19141594](https://pubmed.ncbi.nlm.nih.gov/19141594/)
63. Azzi A, Cosseau C, Grunau C (2009) *Schistosoma mansoni*: developmental arrest of miracidia treated with histone deacetylase inhibitors. *Exp Parasitol* 121: 288–291. doi: [10.1016/j.exppara.2008.11.010](https://doi.org/10.1016/j.exppara.2008.11.010) PMID: [19094990](https://pubmed.ncbi.nlm.nih.gov/19094990/)
64. Marek M, Kannan S, Hauser AT, Moraes Mourao M, Caby S et al. (2013) Structural basis for the inhibition of histone deacetylase 8 (HDAC8), a key epigenetic player in the blood fluke *Schistosoma mansoni*. *PLoS Pathog* 9: e1003645. doi: [10.1371/journal.ppat.1003645](https://doi.org/10.1371/journal.ppat.1003645) PMID: [24086136](https://pubmed.ncbi.nlm.nih.gov/24086136/)
65. Stewart S, Tsun ZY, Izpisua Belmonte JC (2009) A histone demethylase is necessary for regeneration in zebrafish. *Proc Natl Acad Sci U S A* 106: 19889–19894. doi: [10.1073/pnas.0904132106](https://doi.org/10.1073/pnas.0904132106) PMID: [19897725](https://pubmed.ncbi.nlm.nih.gov/19897725/)
66. Doenhoff MJ, Cioli D, Utzinger J (2008) Praziquantel: mechanisms of action, resistance and new derivatives for schistosomiasis. *Curr Opin Infect Dis* 21: 659–667. doi: [10.1097/QCO.0b013e328318978f](https://doi.org/10.1097/QCO.0b013e328318978f) PMID: [18978535](https://pubmed.ncbi.nlm.nih.gov/18978535/)