Targeted insertion of reporter transgene into a gene safe harbor in human blood fluke, *Schistosoma mansoni*

Wannaporn Ittiprasert[1*], Max F. Möscheid[2], Cristian Chaparro[3], Victoria H. Mann[1], Thomas Quack[2], Rutchanee Rodpai[1,4], Andre´ Miller[5], Prapakorn Wistiphongpun[1,6], Watunyoo Buakaew[1,7], Margaret Mentink-Kane[5], Legendre´ Miller[5], Sarah Schmid[5], Anastas Popratiloff[8], Christoph G. Grevelding[2*], Christoph Grunau[3*], Paul J. Brindley[1*]

[1] Department of Microbiology, Immunology & Tropical Medicine, & Research Center for Neglected Diseases of Poverty, School of Medicine & Health Sciences, George Washington University, Washington, D.C. 20037, USA.

[2] Institute of Parasitology, Biomedical Research Center Seltersberg (BFS), Justus Liebig University Giessen, Giessen, Germany

[3] Interactions Hôtes Pathogènes Environments, University of Perpignan Via Domitia, 66860 Perpignan cedex 9 France

[4] Department of Parasitology and Excellence in Medical Innovation, and Technology Research Group, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand

[5] Schistosomiasis Resource Center, Biomedical Research Institute, Rockville, MD 20850

[6] Faculty of Medical Technology, Rangsit University, Pathum Thani province, Thailand 12000, Thailand

[7] Cellular and Molecular Immunology Research Unit (CMIRU), Faculty of Allied Health Sciences, Naresuan University, Phitsanulok province 65000, Thailand

[8] Nanofabrication and Imaging Center, Science and Engineering Hall, George Washington University, Washington, D.C. 20052, USA.

[*] Correspondence: , Wannaporn Ittiprasert, email, wannaporni@gwu.edu; Christoph G. Grevelding, email, christoph.grevelding@vetmed.uni-giessen.de; Christoph Grunau, email, christoph.grunau@univ-perp.fr; Paul J. Brindley, email, pbrindley@gwu.edu

Abstract

We identified genomic safe harbor sites (GSH) in the human blood fluke, *Schistosoma mansoni* and developed a CRISPR-focused protocol for insertion of a reporter transgene into a representative GSH. The protocol employed ribonuclear protein complexes of Cas9 nuclease, three overlapping guide RNAs, and phosphorothioate-modified, double stranded donor DNAs encoding green fluorescent protein driven by a strong endogenous promoter. Gene-editing efficiencies of >20% and reporter transgene fluorescence of >50% of gene-edited eggs were obtained by five days after CRISPR transfection. These methods and results advance functional genomics for multicellular parasites, and represent a tractable path towards transgenic schistosomes using homology directed repair catalyzed transgene insertion. Identification and characterization of GSH is expected to facilitate consistent transgene activity with neutral influence on the host cell genome and, concurrently, provide a privileged locus for transgene activity. This approach should be adaptable to platyhelminths generally.

Main

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) technology has revolutionized genome manipulation in biology, agriculture, and medicine[1-3]. Transgenesis technologies are integral in diverse applications including gene therapy, biotherapeutics, agricultural crop and breed enhancements, and deciphering host-pathogen interactions. With progress emanating from model species and cell lines, tools and techniques can frequently be adapted and transferred to non-model species. Among these are helminth parasites, which are responsible for substantial mortality and disease. . According to the WHO, many important helminth parasites are responsible for 'neglected tropical diseases'[4]. These mostly occur in the global south and are responsible for a global burden of disease that exceeds malaria and tuberculosis. In the post-genomic era of parasitic helminths, like schistosomes, access to CRISPR-based transgenesis protocols is a public health research priority. Progress for one helminth species will facilitate technology transfer to major phyla of invertebrates, e.g., Platyhelminthes, other Lophotrochozoans and Protostomia, for which CRISPR-based reverse and forward genetics have yet to be reported and/or are challenging.

CRISPR enables targeted site-specific mutation(s), obviating an impediment of earlier transgenesis approaches that rely on lentiviruses[5] and transposons such as *piggyBac*[6]. These latter approaches may lead to genetic instability, multi-copy insertion, unstable expression or even inactivation of the transgene and interference with the endogenous gene. In the process of genome editing, the DSBs are resolved by several different repair mechanisms, including the predominant error-prone non-homology end joining (NHEJ) and the templated homologous-directed repair (HDR). Sister chromatids provide a natural repair template, whereas the latter also can be donated by exogenous DNA such a plasmid, oligodeoxynucleotides, and PCR amplicons. When supplied with double-strand (ds) donor DNA with modifications HDR efficiency can be markedly improved [7]. The system of CRISPR/Cas-assisted HDR has been applied in *Schistosoma mansoni*[8,9] with promoter free-single strand-deoxynucleotide donor. Multiple overlapping CRISPR target sites improve precise HDR insertion of large cargoes in embryonic stem cells[10,11], while modification of 5'-termini of long dsDNA donors enhance HDR, favoring efficient, single-copy integration through the retention of a monomeric donor confirmation and thereby facilitating gene replacement and tagging[12].

We identified safe harbor sites in the genome of *S. mansoni*, and adapted CRISPR/Cas9-based approaches to insert a reporter transgene into a located site termed GSH1, situated in euchromatin of chromosome 3. GSH1 was free of repetitive sequences and neighboring long non-coding regions, a situation likely to minimize off-target effects of CRISPR/Cas activity. Multiple sites within this region were targeted with programmed RNAs to boost efficiency of nuclease cleavage and repair in the presence of 5' 5× phosphorothioate bond modified-donor template bearing GSH1-specific homology arms. The approach delivered knock-in efficiencies ~70% in independent replicates of the reporter GFP transgene when the CRISPR materials were delivered by electroporation to the egg stage of *S. mansoni*, with GFP expression controlled by the promoter and terminator of an abundantly transcribed *S. mansoni* ubiquitin gene. In brief, here we provide a new strategy to target-oriented transgene integration into the schistosome genome.

Results

Genome safe harbors predicted in the schistosome genome

To identify sites that could serve as potential GSHs, we conducted a genome-wide bioinformatic search based on established, widely accepted criteria[13], along with newly introduced criteria (below), that would satisfy benign and stable gene expression (Table 1). At the outset, we identified euchromatic regions in all developmental stages of *S. mansoni* to avoid silencing genes to be integrated upon CRISPR/Cas manipulation. With these criteria, we enriched for regions that were, firstly, close to peaks of H3K4me3, a histone modification that is associated with euchromatin and transcription start sites, secondly, regions that did not include H3K27me3, a histone modification that is associated with heterochromatin[14], thirdly, regions of open euchromatin accessible to Tn5 integration, in an Assay of Transposase Accessible Chromatin sequencing (ATAC-seq) providing a positive display of integration events, and fourth, |given that HIV-1 integrates preferentially into euchromatin in human cell lines[15], we used sites of HIV proviral integration known from *S. mansoni* [16] to likewise support predictions of euchromatic regions.

Examination of the draft genome of *S. mansoni* in Worm Base Parasite, version 7 (WormBase Parasite)[17-20] identified 6,884 regions with enrichment of H3K4me3 in the absence of H3K27me3 in available developmental stages (H3K4me3 not K3K27me3). In mature, adult schistosomes, we found consistently 10,533 ATAC positive regions. There were 4,027 ATAC regions that overlapped with H3K4me3 but not K3K27me3, and 2,915 genes overlapped with (ATAC and H3K4me3 not H3K27me3). Forty-two unambiguous HIV integration sites were identified, and eight genes were < 4 kb upstream or downstream from these integration sites. Annotated expression data were available for six of these genes: endoplasmic reticulum Golgi intermedia (Smp_040360), metal tolerance protein C3 (Smp_150230), aldo keto reductase family (Smp_053220), RUN domain containing protein 1 (Smp_067010), endophilin III (Smp_036990), and actin protein ARP2 (Smp_127830) (Table 1). Genes occurring in these putative GSH regions are expressed in all developmental stages, although not uniformly (Fig. 1a). The locations of these potential intragenic GSH sites, which satisfied the above criteria, were Smp_053220, Smp_150230, Smp_040360, Smp_127830, Smp_067010 and Smp_036990 (Table 1).

To identify intergenic GSH, we located 10,149 intergenic regions. There were 9,985 regions beyond 2 kb upstream and 8,837 regions outside long non-coding-RNA (lncRNA), which were intersected to 95,587 unique intergenic regions outside 2 kb and lncRNA of ≥ 100 bp. Two hundred regions were identified intersecting with merged ATAC H3K4me3 signal. Four of these were situated ≤ 11 kb distance from HIV integration sites. We termed these four potential gene free-GSH regions, which satisfied all our criteria, GSH1 (1,416 nt, chromosome 3:13380432-13381848), GSH2 (970 nt, chromosome 2: 15434976-15435954), GSH3 (752 nt, chromosome 2: 9689988-9690739), and GSH4 (138 nt, chromosome 3: 13381901-13382038), respectively (Table 1, Fig. 1b), with the names GSH1 to GSH4 based in rank order on their size from longest to shortest. Several protein-coding gene loci were located proximal to the GSH, although these were > 2kb distant from these intergenic GSH: Smp_052890, Smp_33810, Smp_071830, Smp_245610, Smp_016380, Smp_131070, Smp_052890 and Smp_150460. These may be non-essential genes, when based on orthology to essential genes in model eukaryotes[21].

Efficiency of programmed mutation at GSH1 enhanced by multiple guide RNAs

Based on bioinformatic screening for GSH, here we focused here on GSH1, located on chromosome 3: 13380432-13381848 (1,416 nt) because with 1,416 bp, it was the longest of the four putative intergenic GSH sites. Guide RNAs (gRNA) possessing high on-target specificity; three overlapping guide RNAs, sgRNA1, sgRNA2 and sgRNA3 that did not exhibit self-complementarity, and off-target matches against the reference *S. mansoni* genome (Fig. 2a) were selected from the list of CRISPR/Cas9 targets predicted by CHOPCHOP [22,23]. Ribonucleoprotein complexes (RNP) of Cas9 nuclease and each sgRNA were assembled, after which two mixtures of the three assembled RNPs were prepared. One mixture of CRISPR materials included two assembled RNPs, sgRNA1 RNP and the sgRNA3 RNP (dual gRNAs) and the second included three assembled RNPs, sgRNA1 RNP, sgRNA2 RNP and sgRNA3 RNP (triple, overlapping gRNAs). For each, the mixture of two or three RNP complexes was delivered to schistosome eggs (or other developmental stages) by electroporation (EP), after which the transfected eggs were maintained in culture for 15 days. At that point, genomic DNAs and total RNAs were extracted from the eggs.

Efficiency of genome editing was estimated by DECODR[24] analysis of chromatograms of Sanger sequencing tracings of PCR amplicons spanning the programmed DSBs, amplified from genomic DNAs using primers flanking DSBs, as indicated in Fig. 2a, among experimental and control (mock) treatments. The dual RNPs, sgRNA1+ sgRNA3, delivered mutation frequencies at GSH1 of 4.6-6.3% and ~3.7%, respectively, with short deletions (of one to several, nucleotide in length both on target region 1 and region 3 (Fig. 2b, 2c). Mutations were not detected in the wild type (control, no treatment) and mock treatment groups (not shown). The three overlapping sgRNAs (that shared at least six overlapping nucleotides) induced higher mutation efficiencies at GSH1, with mutation frequencies of 7-15.8%, 9.1-20.2%, and 9.9-19.3% indels at target regions 1, 2 and 3, respectively. In addition, with the triple, overlapping RNPs, larger deletion sizes, up to 115 nt in length, were recorded among the biological replicates (Fig. 2d-f). CRISPR efficiency for each sgRNA was analyzed from six independent biological replicates: both sgRNAs provided similar CRISPR efficiency (Fig. 2g). Notably, however, the combination of the three overlapping gRNAs delivered higher mutation efficiency compared to the two-overlapping guide RNAs ($P = 0.002$) (Fig. 2h).

Knock-in efficiency increased with overlapping guide RNAs

Multiple sgRNAs with overlapping sequences can enhance CRISPR/Cas9-mediated HDR efficiency[10]. Here, three overlapping sgRNAs performed better than dual gRNAs from programmed mutation at GSH1. Subsequently, we investigated transgene knock-in (KI) at GSH1 with three overlapping sgRNAs. As the donor template for programmed homology directed repair, we used the enhanced green fluorescent protein (EGFP) gene with expression driven by an optimized *S. mansoni* endogenous ubiquitin promoter and the cognate ubiquitin terminator region. The donor template also included symmetrical homology arms specific for GSH1, located on the 5' flank of target site 1 and the 3' flank of target site 3 (Fig. 3a, 3b). The donor template was delivered to the schistosomes as linearized long double-strand DNA (lsDNA). Aiming to enhance and favor precise and efficient single-copy integration of the donor transgene into GSH1 by HDR, we biotinylated the 5' termini of the DNA donor amplicons[12] to shield the template from multimerization and from integration at the DSB via the non-homologous end-

joining (NHEJ) repair pathway (Fig. 3a). First, we investigated the impact of length of the homology arms (HA), by comparison donor templates (as above) with homology arms of 200 bp, 400 bp and 600 bp in length after DNA restrictions by either dual or triple sgRNAs. We did not observe EGFP from 200 bp and 400 bp HA lsDNA donor in either donor transfected control and KI parasite eggs at days 5-6 after transfection in both dual and triple sgRNAs treatment conditions (data now shown). There was inconsistent EGFP expression (<1% or absent) in the live miracidium with the schistosome egg i with 600 bp HA along with dual sgRNAs (data not shown). Subsequently, we focused this investigation the donor transgene flanked by homology arms of 600 bp each in length, with EGFP expression in the eggs examined every second day for 15 days. On examination using spectrally resolved, confocal laser scanning microscopy, EGFP signals were not observed in the negative control groups, although the "autofluorescence" characteristic of eggs was apparent[25]. The EGFP signal was also detected in the lsDNA donor control (without RNPs) for several days. The EGFP signal was detected in the CRISPR materials group that included the lsDNA donor with 600 bp HA; the signal was detected up to15 days (at which point the experiment terminated).

Next, we investigated the 3' and 5' KI at GSH1 by a PCR approach on genomic DNA of integrated ubiquitin promoter in frame with EGFP and its terminator. For the analysis of 5' PCR KI,|, we used a forward primer specific for several nucleotides upstream of the 5' end HA with a reverse primer specific for the ubiquitin promoter (Fig. 3b). For analysis of the 3' KI integration junction, , the reverse primer was specific for a site downstream of the 3'end of the HA and was paired with a forward primer specific for the ubiquitin terminator. Fragments re[resenting the 3' KI and 5' KI integration regions of 983 bp and 728 bp, respectively, were observed in the KI treatment groups but not in the other (control) groups (Fig. 3c). Expression of the EGFP transgene was monitored using DNase-treated RNA (to eliminate the possibility of lsDNA donor contamination and genomic DNA): EGFP transcripts were observed in the KI experimental groups, among which we observed slight variability in transcript abundance among biological replicates (Fig. 3d).

Genome safe harbor accessible in the adult developmental stage of the schistosome

To investigate the impact of CRISPR manipulation on the adult developmental stage of *Schistosoma*, we electroporated the triple RNPs (sgRNAs1+2+3) and lsDNA into 20 worms (10 females, 10 males). *In vitro* maintenance for 10 days yielded similar movement and egg-laying between the experimental and control cultures. The EGFP expression was observed in some non-dividing cells of adult worms (Fig. S1), and these EGFP-positive worms survived and were actively mobile at the time of cessation of the culture (day 11). Efficacy of programmed mutation was investigated at the genome level using genomic DNAs extracted from EGFP-positive schistosomes (six females, two males), which involved deep sequencing of amplicons spanning the predicted DSB by CRISPREsso2[26,27]. Alleles with deletions ranging up to 150 nt, likely resulting from NHEJ, were characterized. The analysis estimated mutation efficiencies of 13.3% and 16.8% indels for the female and males, respectively (Fig. S1b, S1d). Based on both the *ex vivo* findings with both eggs and adult stages of the schistosome, and the *in silico* predictions for the presence and criterion conformity of predicted GSH within the genome of *S. mansoni*, we considered that the intergenic (gene-free region) GSH1 represented a suitable candidate locus for CRISPR/Cas-catalyzed insertion of a large sized, exogenous, and over expressed reporter transgene.

Transgene expression in the miracidium following programmed insertion

EGFP positivity and intensity in the treated eggs were assessed quantified by using spectral laser scanning confocal fluorescence microscopy[25]. Active transgene expression was confirmed within miracidia developing inside transfected eggs (Fig. 4a, b). Firstly, EGFP appeared to be expressed by cells ubiquitously throughout many tissues and cells of the schistosome larva. Morphological malformation was not observed in transgenic eggs and their enclosed larvae (eggs, $n = 402$ aggregated from four independent, biological replicates). More intense GFP fluorescence was consistently seen and quantified in eggs from the experimental treatment group than the mock control eggs and in eggs transfected solely with donor template (Fig. 4a1, a2) at the 509 nm. Subsequently, we used the average EGFP background from negative eggs to normalize intensity values for specific EGFP fluorescence in the donor lsDNA control and the CRISPR with lsDNA treated eggs.

In schistosome miracidia within the eggshell at 15 days post electroporation, fluorescence intensity of transgenic parasites markedly differed from the wild type eggs. Seventy five percent of 402 eggs examined displayed EGFP fluorescence in miracidia. About 25% of eggs containing a miracidium transfected with the donor transgene exhibited EGFP (Fig. 1c). However, after reading fluorescence intensity of EGFP (established by subtracting the signal from autofluorescence at 509 nm, the emission wavelength for EGFP [25]), the EGFP-specific signal in the control donor transgene group, 856-1,713 arbitrary units [28] (average, 1,290 au) was significantly lower than the experimental group transfected with overlapping guide RNPs and the donor transgene bearing 600 bp HA, 4972.5-8,963.1 au (average, 6,905 au) ($P < 0.001$). EGFP expression within developing miracidia among the eggs was not apparently localized; diverse cells and parasite organs expressed the fluorescence reporter gene.

Discussion

Schistosomes are water-borne pathogens and pose a constant threat to human health in the global south and beyond. Only a single antiparasitic drug, praziquantel, is available for treating schistosomiasis. In light of the possibility of resistance development, reinfection after treatment, the absence of immunity to reinfection following curative treatment, and reemergent spread into southern Europe[29], likely precipitated by the occurrence of the intermediate hosts, globalization, and increasing global temperatures, are causes for increasing concern. To advance functional genomics for schistosomes in the post-genomic era, here (to our knowledge for the first time) we localized genome safe harbor sites in *S. mansoni*, optimized conditions for delivery and structure of transgene cargo, inserted the reporter transgene into a predicted intergenic genome safe harbor (GSH) by programmed CRISPR/Cas9 homology-directed repair by targeted mutation using three overlapping guide RNAs, and quantified transgene activity using confocal imaging of emission spectra specific for EGFP green fluorescence protein. More specifically, delivery to the schistosome egg by electroporation of multiple overlapping guide RNAs delivered with Cas9 nuclease as ribonuclear complexes lead to efficient programmed cleavage of the GSH1. Double-stranded DNA flanked by chemically modified termini, encoding enhanced EGFP driven by the endogenous schistosome ubiquitin gene promoter and terminator, served as the model repair template. Our studies are consequential for two principal reasons. First, the results advance functional genomics and forward genetics for a hitherto unmet challenge to manipulate a pathogen of global public health significance. Second, transgenes can be targeted to safe integration sites to endow individual stages or populations of these pathogens with novel functions, which will have broad potential for basic and translational studies[30-32]. Third, the editing methods developed can be adapted for knock-out approaches of genes of interest, in schistosomes and probably other platyhelminths, for which genome-project data are available.

Targeting transgenes using homology-directed repair (HDR) at intergenic GSH sites catalyzed by RNA-programmed Cas9 can be expected to enable, in effect, a mutation-independent genome modification to support forward genetics investigation. In the human genome, GSHs, which are situated either in intergenic or intragenic regions, promote stable expression of integrated transgenes without negatively affecting the host cell[13]. Access to schistosome intergenic GSH will also provide a step-change advance for functional genomics of these pathogens. For *S. mansoni*, our prediction criteria for GSH included location in euchromatin to avoid silencing of the transgene, unique genome-target sequence to minimize off-target events, avoidance of lncRNA genes, presence of epigenetic marks for open chromatin structure, and the absence of epigenetic marks indicating heterochromatin. We named the intergenic sites GSH1, -2, -3, and -4, which were located on chromosomes 2 and 3. (*S. mansoni* has seven pairs of autosomes and Z and W sex chromosomes.). In addition, we assessed one intergenic GSH1 locus for CRISPR/Cas9 gene editing and over expression EGFP integration. We predicted potential GSH in non-essential coding regions of aldo keto reductase, metal tolerance protein C3, endoplasmic reticulum Golgi intermediate, actin subunit, RUN domain and endophilin III. In similar fashion to findings that have been reported in human and mouse genomes[33], we posited that schistosome GSH will tolerate the integration by CRISPR-catalyzed HDR donor templates and enable stable expression of the integrated transgenes without negatively impacting the genome of the transfected helminth and progeny.

We edited GSH1 using two and three ribonucleoprotein complexes of Cas9 endonuclease with the overlapping guide RNAs, sgRNAs numbers 1 and 3 (dual guides approach) and with sgRNAs numbers 1, 2 and 3 (triple guides approach). Overlapping CRISPR target sites were selected from lists of target site predicted and ranked by the CHOPCHOP algorithm. The triple overlapping guides approach delivered higher CRISPR/Cas9 efficiency, and larger sized deletion mutations than dual guides (Fig. 2), and outcome that would enhance the efficiency of HDR in the presence of the long stranded DNA (lsDNA) donor template. Accordingly, we deployed the three overlapping sgRNAs in the CRISPR/Cas9 system to deliver the 3551 bp lsDNA encoding the reporter transgene - here a *S. mansoni* ubiquitin gene promoter and terminator flanking the EGFP reporter - to schistosome eggs.

This cohort of eggs, termed "liver eggs", LE, was expected to include > 50% eggs that included the mature miracidium with the remainder dead eggs or immature/developing eggs[34-36]. By 10 days following transfection, higher expression of EGFP at GSH1 was apparent based on examination of the normalized fluorescence intensity of the miracidium inside the egg carrying the transgene. Precise knock-in (KI) was confirmed using target site-specific amplicons (Figs. 3c). Moreover, RT-PCR of the EGFP expression confirmed KI of donor transgene (Fig. 3d). Using a similar approach, we also investigated programmed gene editing at GSH1 and impact on fitness in the adult stage schistosome as reflected in motility, morphology, mortality of the transfected adult schistosomes, and release of eggs *in vitro*. The worms remained active, did not exhibit apparent morphological changes for at least 11 days after transfection, and the females released eggs *in vitro*, all of which were similar to the phenotype of the control group schistosomes. (Fig. S1).

These findings indicated that GSH1 represented a promising safe harbor site for forward genetics-focused forward genomics with this schistosome. With the longer-term goal of deriving lines of transgenic parasites carrying gain- or loss-of-function mutations, we also undertook preliminary studies with the newly laid egg of *S. mansoni*, a stage that at its origin includes a single zygote (surrounded by vitelline yolk cells) and which, therefore, is a window to the germline[37,38]. As noted, highly efficient HDR resulted from the combination of multiple target site-overlapping RNPs programmed to cleave GSH1 in the presence of the chemically modified a repair template protected by chemical modifications. Notably, ~75% of mature eggs (n = 402 eggs from four independent replicates) exhibited reporter transgene fluorescence with the miracidium developing within the eggshell (Fig. 4c), and significantly more fluorescence than seen in the control eggs transfected with donor template but not with the RNPs, ~25% (*n* = 397 eggs, from four independent biological replicates). EGFP signals were not present in the control, untreated wild type egg, which by 9 days following transfection exhibits minimal background fluorescence[39].

We also transfected *in vitro* newly laid eggs, termed IVLE, and of the adult stage of the parasite using the same multiple RNPs targeting GSH1 and donor template. The IVLE, deposited *in vitro* up to 12 hours after recovery of the adult schistosomes from the euthanized mouse, contains the zygote and developing blastula including the germ tissue, and were maintained thereafter in culture for 10 days following transfection with CRISPR materials. These eggs expressed EGFP by about 7 days after transfection as they developed (Fig. S2). During this 7 days interval, the eggs grew to contain the fully developed miracidium[40,41]. In contrast to the findings with LE, the miracidium failed to develop in > 50% transfected IVLE and, of those that did develop, <1%

were EGFP-positive. Given the fragility of the IVLE[38,41], alternative delivery methods to electroporation, such as lipid nanoparticle containing the RNP and the donor template within a single lipid enclosed sphere, may improve efficiency[42-46] of delivery to the schistosome nucleus while minimizing loss of fitness of the manipulated schistosome larva.

Nonetheless, this study demonstrated that the GSH1 locus is a prospective safe harbor locus site for germline transgene integration in *S. mansoni* although further validation is needed. Comparison of the utility of GSH1 with the other gene-free GSH, GSH2, -3, and -4 (Fig. 1), might also uncover profitable modifications. Likewise, the intragenic sites may exhibit expedient attributes functional genomics. The safe harbors in human gene therapy, *CCR5*, *AAVSI*, and *Rosa26*, all reside within intragenic, gene-rich loci and whereas they have been targeted with therapeutic gene cargo, including for example the insertion of *FANCA* at AAVS1 in $CD34^+$ hematopoietic progenitors from Fanconi anemia patients[47], ideally intergenic sites might be inherently safer[33]. Overall, this report and the methods presented here enabled novel insight into efficient transgenesis and forward genetics for *S. mansoni* and will promote forward genetics approaches in functional genomics for schistosome and related helminth parasites.

Online content

Methods

Computational search for gene safe harbors in *Schistosoma mansoni*

We undertook a genome analysis of intergenic (gene-free) and intragenic (gene-linked) regions to identify prospective a genome safe harbor, in like fashion approaches focused on the human genome[13]. In essence, we aimed to locate a GSH to enable stable expression of the integrated transgene free of interference from the host genome, and which in parallel integrates and transcribes transgenes without negative consequences for the host genome or cell[33]. *Gene-linked GSH*, deployed four criteria. First, adjacent to peaks of H3K4me3, a histone modification associated with euchromatin and transcription start sites; second, not near or containing H3K27me3, which is associated with heterochromatin, in any of the life-cycle stages; third, open, euchromatic chromatin was accessible to Tn5 integration and ATAC-sequence provides a positive display of such integration events. Consequently, safe harbor candidate regions should deliver an ATAC-sequence signal; and fourth, near known HIV integration sites. Given that HIV integrates preferentially into euchromatin in human cells, HIV integration into the schistosome genome may likewise indicate a euchromatic region.

To find loci conforming to the four criteria, pooled ChIP-seq data for H3K4me3 and K3K27me2 from previous studies[48] was aligned against on *S. mansoni* genome data (version 7 on the date of analysis). ATAC-seq was performed as previously described with slightly modification[49]. Peakcalls of ChIP-seq and ATAC-Seq were done with ChromstaR[48,50] and stored as Bed files. Bed files were used to identify the presence of H3K4me3 and absence of H3K27me3 in adults, miracidia, *in vitro* sporocysts, cercariae and *in vitro* schistosomule with hbedtools intersect. Thereafter, ATAC-seq data from males and females (two replicates each) were intersected to find common ATAC positive regions. H3K4me3-only (H3K27me3 absent) common to all stages and ATAC signals were intersected to find common regions. H3K4me3 common to all parasite stages and ATAC signals were intersected to find common regions. Next, the HIV integration sites were identified by using data from ERR338338[51]. Reads were mapped to the lentivirus genome (HIV-1 vector pNL-3, accession AF324493.2) using Bowtie2 with default parameters. Those paired reads were extracted where one end mapped to HIV and the other end mapped to schistosome genome at a unique location. Genes from the BED files above that located ≤ 11 kb HIV-1 integration sites were identified with bedtools closestbed. Gene expression data of these genes were obtained from https://meta.schisto.xyz/analysis/. Information remains unavailable for Smp_343520.

Intergenic GSH. Given that transgene integration into and existing gene could disrupt key functions and endow selective (dis)advantage to the genetically modified cell and its progeny[52,53], we scanned constitutively euchromatic regions for a gene-free region. We defined genes as protein coding sequences and sequences coding for long non-coding RNA (lncRNA). In view of our goal to use CRISPR/Cas mediated-HDR to insert the transgene, ideally, we searched preferentially for unique sequences, to obviate off-target gene modification, and excluded gene free-regions composed of repetitive sequences. Those unique sequences were also annotated outside lncRNA, regions beyond putative promotors that we deemed as 2 kb upstream of the transcription termination site (TTS), and the regions close to peaks of H3K4me2 in all

parasite stages which never contained H3K27me3. The regions also overlapped ATAC-seq positive sites with ≤11 kb distance from HIV integration sites were included (~10 kb is the size of the HIV genome).

Integrating a transgene into an existing gene or its putative promotor region could disrupt important function and provide a selective disadvantage to the genetically modified cells. Therefore, we deliberately searched for gene-free constitutive euchromatic regions. Here we define gene as protein coding genes and genes coding for long non-coding RNA (lncRNA). Integration *via* Crispr/Cas relies on guide RNA with specific, ideally unique sequences. To exclude gene-free regions that are composed of repetitive sequences, we also searched for repeats. A total of 10,129 protein coding gene locations and 27 pseudogenes were extracted from the schistosome genome annotation. BEDtools were used to delimit 2 kb upstream regions (FlankBed). Annotations of 16,583 lncRNA were pooed from http://verjolab.usp.br/public/schMan/schMan3/macielEtAl2019/files/macielEtAt2019.bed12[54]. Repeats were masked with RepeatMasker V4.1.0 using a specific repeat library produced with RepeatModeler2 V2.0.1 and stored as a GFF file. BED files with coordinates outside these annotations were generated by BedTools complementBed. Finally, BedTools Multiple Intersect was used to identify regions that are common to unique regions (complement of repeatmasker), intergenic regions, > 2 kb upstream and outside of lncRNA. Only regions ≥ 100 bp were retained. We reasoned that otherwise it would be too difficult to design guide RNAs. These regions were intersected with merged H3K4me3-only common to all developmental stages and ATAC signals (euchromatic signal). BedTools ClosestBed was used to determine distance to the nearest HIV provirus integration.

Gene-linked GSH. Here we also used the above criteria. Overlapping genes were identified using published *S. mansoni* version 7 annotation. Then, the HIV integration sites were identified by ERR338338[51]. Reads were mapped to the virus genome (HIV-1 vector pNL-3, accession number AF324493.2) using Bowtie2 with default parameters. Those paired reads were extracted where one end mapped to HIV and the other end mapped to schistosome genome on a unique location. Genes from the BED files above that located ≤ 4 kb HIV integration sites were identified with bedtools closestbed. Gene expression data of these genes were obtained from https://meta.schisto.xyz/analysis/.

Table 1 summarizes the criteria used to predict schistosome genome safe harbor sites.

Developmental stages of the schistosome

Mice (female, Swiss Webster) infected with *S. mansoni* were obtained from the Schistosomiasis Resource Center (Biomedical Research Institute, Rockville, MD) within seven days of infection by cercariae (180 cercariae/mouse/ percutaneous route of infection). The mice were housed at the Animal Research Facility of George Washington University, which is accredited by the American Association for Accreditation of Laboratory Animal Care (AAALAC no. 000347) and has the Animal Welfare Assurance on file with the National Institutes of Health, Office of Laboratory Animal Welfare, OLAW assurance number A3205. All procedures employed were consistent with the Guide for the Care and Use of Laboratory Animals. The Institutional Animal Care and Use Committee (IACUC) of the George Washington University approved the protocol used for maintenance of mice and recovery of schistosomes.

Mice were euthanized at about 46 days after infection, after which schistosomes were recovered by portal vein perfusion with 150mM NaCl, 15mM sodium citrate, pH 7.0. The worms were washed with 1×PBS, 2% antibiotic/antimycotic and maintained thereafter in DMEM, 10% heat inactivated bovine serum, and 2% antibiotic/antimycotic at 5% CO_2, 37°C [55] In addition, at necropsy, the liver were resected, rinsed in 70% ethanol, washed twice with 1×PBS, before blending with a tissue homogenizer. Liver tissue homogenate was incubated with collagenase at 37°C for 18 h after which schistosome eggs were recovered by Percoll gradient centrifugation, as described[56]. Eggs isolated from livers, termed LE [41], were cultured overnight before transfection. From cultures of the perfused adult worm population, concurrently, eggs laid in culture by adult female schistosomes from 0 to 12 hours after necropsy, were collected and maintained in high nutrient medium (modified Basch's medium)[55]. We termed these eggs, *in vitro* laid eggs (IVLE)[41]. At its release from the female, the IVLE contains the zygote surrounded by yolk cells. The larva grows and by day 7 has developed into the mature miracidium[40].

Guide RNAs, ribonucleoprotein complexes

For transfection, we focused on GSH1, located on *S. mansoni* chromosome 3; 13380432-13381848 (Table 1), an intergenic safe harbor site with the longest region (1,416 nt) among the predicted intergenic GSHs. Single gRNAs (sgRNA) for GSH1 were designed with assistance of the CHOPCHOP[22,23,57] tools, using the version 7 annotation of the *S. mansoni* genome[17], to predict target sites, off-targets, and efficiency of CRISPR/Cas9 programmed cleavage. Three overlapping (expected DSB sites, 6-12 nt apart) CRISPR target sites; sgRNA1, sgRNA2, and sgRNA3 with predicted absence of both off-target effects and self-complementarity, and each with similar CRISPR efficiency ~50% were selected. Although these sgRNAs were not among the top five predicted by CHOPCHOP, they did exhibit off-target identity to the genome. Their CRISPR efficiency was 55.7% (rank 7), 47.0% (rank 16), and 36.0% (rank 23), and they were located on the forward strand of GSH1 at nucleotide positions 605-624, 617-636, and 623-642, respectively (Fig. 2a). Synthetic guide RNAs, Alt-R CRISPR-Cas9 sgRNA chemically modified to enhance functional stability, and recombinant *Streptococcus pyogenes* Cas9 nuclease, Alt-R HiFi Cas9 which includes nuclear localization sequences (NLS), were purchased from Integrated DNA Technologies, Inc. (IDT) (Coralville, IA). Each ribonucleoprotein complex (RNP) was prepared in the separate tube, with Cas9 and a single sgRNA at 1:1 ratio, in 25 µl Opti-MEM. The sgRNA was mixed with the nuclease by gentle pipetting and incubated for 10 min at room temperature to allow assembly of the RNP.

Donor plasmid construct and preparation of long double strand DNA donor

The donor plasmid vector (pUC-Ubi-EGFP-ubi) was synthesized and ligated into pUC by Azenta Life Sciences (Chelmsford, MA). The construct included homology arms of 600 bp length corresponding to GSH1 at 22-621 nt (5'-homology arm) and 640-1239 nt (3'-homology arm), respectively, flanking the in frame expression cassette composed of the *S. mansoni* ubiquitin promoter (2,056 bp), EGFP (717 bp), and the ubiquitin terminator (578 bp). Plasmid DNA was amplified by PCR using Phusion High-Fidelity DNA Polymerase (New England Bio-Labs, Ipswich, MA, cat no. M0530) with primers specific for the 5' and 3' termini of the homology arms. These primers were 5'end biotinylated and 5×phosphorothioate-modified to enhance stability; 5'-modified long dsDNA donor (lsDNA) enhances HDR and favors efficient single-copy integration by it retention of monomeric conformation[12] (Fig. 3a).

PCRs were carried out in 50 µl reaction volume containing 200 µM dNTPs, 0.5µM of each primer, 100 ng pUC-Ubi-EGFP-ubi, 3% DMSO and 1 unit of Phusion DNA polymerase, with thermocycling at 98°C, 30 sec, 30 cycles of 98°C, 10 sec, 55°C, 30 sec, 72°C, 3 min, and final extension at 72°C, 10 min. Amplification products were isolated using the NucleoSpin Gel and PCR Cleanup and gel extraction kit (Takara, San Jose, CA, cat no. 740609), eluted in 30 µl nuclease-free water, and the long stranded (ls) DNA donor stored at -20°C until used.

Transfection of schistosomes

Ten thousand eggs (LE) of *S. mansoni*, 20 adult stage schistosomes, or ~300 *in vitro* laid eggs (IVLE) were washed three times with ice-cold 1×PBS before transfer into 4 mm pathway cuvettes (BTX, Holliston, MA) with ~100 µl Opti-MEM as electroporated buffer. Each 25 µl of RNP and lsDNA donor was immediately added into the cuvette, to a total cuvette volume of ~300 µl. Transfection of schistosome eggs and adults with CRISPR materials was accomplished using square wave electroporation (Electro SquarePorator ECM 830, BTX), with a single pulse of 125 volts for 20 ms, transfection conditions as optimized previously[8,9,58]. Thereafter, the transfected schistosome stages were transferred to culture medium (as above).

Nucleic acids

To recover genomic DNA and total RNA, eggs from each replicate were triturated in ~100 µl DNA/RNA Shield solution (Zymo Research, cat no. R1100, Irvine, CA) using a motor-driven homogenizer fitted with a disposable pestle and collection tube (BioMasher II, Bunkyo-ku, Tokyo, Japan). DNA was isolated from 50% of the homogenate, and RNA from the remainder. 250 µl DNAzol® ES (Molecular Research Center, Inc., Cincinnati, OH; cat no. DS128) was dispensed into the homogenate, and DNA recovered according to the manufacturer's protocol. Total RNA was extracted from the homogenate by adding 250 µl RNAzol RT (Molecular Research Center, Inc., cat no. RN190). Yields and purity were assessed quantified by spectrophotometry (NanoDrop One Spectrophotometer, ThermoFisher Scientific), using ratios of absorbance at 260/280 and 260/230 nm[59].

Analysis of CRISPR on-target efficiency

Amplicons of GSH1 spanning the programed DSBs were obtained using population genomic DNA (above) and primers termed 'control-F and control-R primers' that cover the region flanking expected double strand break of all the CRISPR target sites. Amplification products were purified (NucleoSpin Gel and PCR Cleanup and gel extraction kit, cat no. 740609, Takara) and the nucleotide sequences determined by Sanger cycle sequencing (Azenta Life Sciences, South Plainfield, NJ). Chromatograms of the sequence traces of experimental and control group(s) was compared using DECODR[24] at default parameters. NGS deep sequencing was untaken on y genomic DNAs of eggs, IVLE and adult schistosomes, using the Amplicon EZ sequencing with 2 x 300 bp configuration (Azenta Life Sciences). Subsequently, >100,000 sequence reads per sample were analyzed by CRISPResso2[9,26,27] with window analysis 200 bp parameter, multiple sgRNA targets. Deeply sequenced reads (>100,00 reads) were analyzed using CRISPRsso2, with resulting merged images used to plot the indel size distributions of the experimental compared to the wild type reference.

Detection of transgene integration into the schistosome genome

Integration of donor transgene at GSH1 was analyzed by PCR with GoTaq G2 DNA polymerase (cat no. M7841, Promega, Madison, WI) using two pairs of primers; one locates on the GSH1 using specific primers upstream or downstream of the homology arms paired with primers specific for the transgene (Fig. 2b), as described previously[9]. PCR conditions: 95°C, 2 min, 40 cycles 94°C, 15 sec, 58°C 30 sec, 72°C, 60 sec. Amplification products were size separated by electrophoresis and stained with ethidium bromide. The expected product sizes for the 5' and 3' integration site specific amplicons were 728 bp and 983 bp, respectively, and an amplification control was included, expected product size 764 bp (Fig. 2b).

Quantification of EGFP transgene expression by RT-PCR

To examine the mRNA expression of EGFP, total RNAs were extracted from the LE by RNAzol® RT (Molecular Research Center, Inc., cat no. RN190) as manufacturer's manual. The total RNA was transcribed into cDNA after treated with DNase enzyme to get rid of genomic DNA contamination or unuse lsDNA donor using Maxima First Strand cDNA synthesis kit with DNase (Thermo Fisher Scientific). The qPCR was performed using the GoTaq® G2 DNA polymerase (cat no. M7841, Promega, Madison, WI) with the specific primers; EGFP-F 5'-atggtgagcaagggcgagg-3' and EGFP-R 5'-cttgtacagctcgtccatgcc-3' (Fig. 3b) with expected amplicon at 717 bp. *S. mansoni* GAPDH (Smp_056970) was used as the reference gene. The specific primer for GAPDH-specific oligos: GAPDH-F; 5'-atgggacatttccaggcgag-3', GAPDH-R; 5'-ccaacaacgaacatgggtgc-3', expected amplicon of 213 bp in length. PCR cycling conditions: 95°C, 2 min, 25 cycles 94°C, 15 sec, 58°C, 30 sec, 72°C, 30 sec, after which amplification products were separated by electrophoresis through 1% agarose and stained with ethidium bromide.

Quantification of EGFP fluorescent by spectral fluorescent unmixing in schistosome parasite

Spectral and spatial distribution of EGFP fluorescence were assessed using confocal laser scanning microscopy, using a Zeiss LSM710 Meta detector fitted Axiovert 200 (Carl Zeiss, Jena, Germany). Images were collected with the C-Apochromat 20×, 1.2 NA water immersion objective. Spectroscopic measurements were performed in response to excitation by 458 nm (16.5 μW) Arion laser line and 633 nm He/Ne laser line (Lasos Lasertechnik, Jena, Germany), which were used for focus and transmission mode imaging. Emission was detected with spectral META detector at 16 channels 477-638 nm simultaneously. A hurdle when viewing of EGFP via fluorescence microscope autofluorescence known to originate from the egg shell and adult female *S. mansoni*[60,61,62], with vitelline cells determined to be the source of the emission signals[63]. Accordingly, all spectra of EGFP expressed in a miracidium inside each eggshell or cell in the adult stage worm were obtained by selecting the interest area (a whole miracidium inside egg or spots of cells inside the worm) in multispectral images using LSM Image Examiner and were collected for solvent background by subtracting autofluorescence regions from the entire auto fluorescent egg. Total EGFP intensity was calculated by the software at 509 nm[63] from a total of ~400 eggs containing a miracidium in each of both the control and experimental groups, all of which contained the miracidium (~100 eggs from each of four biological replicates). Images from adult worms were collected at day 15 following transfection.

Acknowledgements

Schistosome-infected mice were provided by the NIAID Schistosomiasis Resource Center of Biomedical Research Institute, Rockville, MD through NIH-NIAID contract HHSN272201700014I for distribution through BEI Resources. This work was supported by Wellcome Trust award 107475/Z/15/Z (PI, Karl F Hoffmann).

Author contributions

P.B., C.G, and C.G.G. conceived the study. W.I. planed and performed the gene editing, transgene knock in and data analysis. M.M. and T.Q. contributed promoter and terminator cloning and analysis for EGFP expression. R.R. performed optimization of the donor to use. V.H., A.M, L.M, S.S, and M.M. maintained the parasite life cycle, parasite collection and purifications. P.W. and W.B. investigated indel and transgene knock in, M.M and A.P. contributed to analysis of confocal micrographs, C.C., C.G. and K.H. contributed the gene safe harbor analysis. All authors contributed to the writing, and all approved the final version of the manuscript.

Competitive interests

The authors declare no competing interests.

Figure legends

Figure 1. Normalized gene expression of predicted intragenic GSH sites and the locations of gene-free stretches bearing GSHs on chromosomes 2 and 3 of *Schistosome mansoni*. Panel **a.** Normalized expression (X-axis) for the intragenic GSH at each developmental stage of the schistosome. The protein coding sequences for Smp_036990, Smp_053220, Smp_150230, Smp_040360, Smp_127830, and Smp_067010 are shown in green, blue, yellow, gray, orange, and dark blue colored bars, respectively. **b.** Four extragenic GSH sites (blue rectangles), specifically GSH2 and GSH3 on chromosome 2 and GSH1 and GSH4 on chromosome 3. The red boxes and bars indicate the endogenous genes proximal to the predicted GSHs. The accession number, Smp_xxxxxx, of each gene is indicated. Black and white bars indicate GSH position coordinates on the chromosome.

Figure 2. Programmed mutation of genome safe harbor enhanced by three overlapping guide RNAs. Panel **a.** Schematic diagram to indicate sites of the overlapping guide RNAs within GSH1. **b-c, g.** Representative examples of indel percentages at GSH1, as a measurement of CRISPR catalyzed gene editing efficiency, as estimated using analysis of nucleotide sequences by the Deconvolution of Complex DNA Repair (DECODR) algorithm using distance from two overlapping guide RNAs, gRNA 1 and gRNA 3. Small deletions, 1-3 nt in length, of 1.7-13.8% indel mutations were estimated from each target site (panel g, left box). **d-f, g.** Larger mutations of ≤ 115 nt with higher CRISPR/Cas9 efficiency (2.5-71.9%) were observed at each target site following KI using guide RNAs numbers 1, 2 and 3 (panel g, right side). **h.** The approach deploying the three-overlapping guide RNAs was significantly more efficient than that using two overlapping guide RNAs, as assessed from six, independent biological replicates ($P = 0.0021$, unpaired *t*-test) with the 95% confidence interval (CI) for the difference between the means, 6.17 to 20.74; 20.18 ± 13.45% (X ± SEM) observed using three overlapping guide RNAs and 6.73 ± 3.27 % (X ± SEM) with two gRNAs.

Figure 3. Targeted insertion and transgene expression at GSH1 in the egg stage of *Schistosoma mansoni.* Programmed CRISPR/Cas9 insertion (knock-in, KI) at GSH1 on chromosome 3 of *S. mansoni* of a donor repair template of 4.451 kb in length, encoding an EGFP transgene driven by the endogenous *S. mansoni* ubiquitin promoter and terminator. Panel **a.** Topology of double-stranded DNA donor prepared from a primer pair with 5' 5x-phosphorothioate modification. The donor template encoded the *S. mansoni* ubiquitin promoter (pink bar) driving expression of the EGFP reporter gene (green) and ubiquitin terminator (pink) and was flanked at its termini with symmetrical 600 bp homology arms (black bars). The homology arm on the left (HAL) was situated 600 nt of upstream of the position of sgRNA1 and the homology arm on the right (HAR) is 600 nt of downstream of that of sgRNA 3. **b.** Schematic illustration of the WT and knock-in alleles after multiple double stranded breaks programmed by sgRNAs 1, 2 and 3 (scissors). The PCR primers used are shown as purple arrows. **c.** Targeted knock-in of the EGFP cassette detected by genomic PCR using 5'KI (728 bp) or 3'KI (983 bp) primer pairs. Negative controls for KI included wild type (WT), mock, and donor treatment groups not exposed to RNPs/Cas9 nuclease. **d.** EGFP transcript expression (717 bp) by RT-qPCR following the integration into the egg of the parasite into the GSH1as well as schistosome GAPDH (213 bp). The three biological replicates of knock-in and its terminator are shown in lanes KI-1, KI-2 and KI-3 represent three independent biological replicates of programmed insertion of the ubiquitin promoter-driven EGFP, and lanes 1-3 show the RT-qPCR outcomes

from schistosome RNA with donor DNA electroporation (without CRISPR materials - nuclease or guide RNAs). Double-stranded DNA donor was used as the positive PCR template. Transcription of GAPDH was seen in all treatment and control groups (lanes 1-3 and KI-1 to KI-3 in bottom panel) but not in the donor group. Primer-dimer and/or non-specific PCR band(s) from DNA donor transfected-eggs were ≤ 100 bp in size.

Figure 4. Markedly higher numbers of eggs emitting green fluorescence following programmed knock-in of the reporter transgene at genome safe harbor as assessed by spectral image analysis. Confocal laser scanning micrographs: Panel **a**, eggs exhibiting background signal (autofluorescence) from the control group, *i.e*. eggs transfected with donor repair template only; a1 and a2, representative images from biological replicates. Panel **b**, eggs expressing the EGFP encoding transgene from the experimental group transfected with RNPs and the donor repair template; b1, b2, representative images from two biological replicates. Many eggs expressed EGFP with the broad range in intensity of fluorescence ranging from higher intensity (green arrow) and lower levels (yellow arrow) following programmed homology directed repair; micrographs taken at day 5 after transfection. Eggs expressed EGFP until day 15 (experiment terminated). Panel **c**, micrograph showing representative images to demonstrate the EGFP and autofluorescence of individual eggs. Panel **d**, mean emission spectral intensity for eggs, scanned from 477-638 nm, with curves for each of the four biological replicates presented. Spectral signal, and the signal at 509 nm (peak wavelength for EGFP) for each positive egg was normalized with the average autofluorescence signal from the same biological experiment. and with the points showing mean values. Panel **e**, Percentage of egg population positive for EGFP fluorescence. Control group (gray), experimental group (green); findings from four independent, biological replicates (~100 eggs per group); eggs expressing EGFP in the control group, 23.7% (range,19 to-32%), eggs expressing EGFP in the experimental group, 74% (range, 68-79%) ; $P < 0.001$, two-tailed $t = 69.87$, $df = 142$; difference between means (EGFP-KI – only donor) ± SEM, 49.7 ± 0.7, 95% CI, 48.3 to 51.1. Panel **f**, normalized fluorescence spectral intensity from control eggs (transfected with donor repair template) exhibiting higher intensity than autofluorescence; these eggs were also scored as EGFP-positive, and with a normalized EGFP intensity mean, 1290 au (range, 856 - 1712.8); experimental group, normalized-EGFP intensity, mean 6905 au (range 4971.5 – 8963.1); $P < 0.001$, unpaired *t*-test, n = 402; difference between means of experimental and control group eggs ± SEM, 5651 ± 57.40, 95% CI, 5502 to 5728).

Supporting information

Figure S1. **GSH1 deletions resulting from CRISPR/Cas9-derived NHEJ and random expression of GFP transgene resulting from HDR in 100% survival *S. mansoni* mature worms.** Ten males or 10 female *S. mansoni* were transfected with multiple RNPs and lsDNA donor encoding EGFP driven by the ubiquitin promoter to investigate fitness of the schistosomes following CRISPR-associated manipulation. EGFP expression was evident in six females and two males (green arrow) (panels a, c). Blue arrows indicate autofluorescence that was also apparent in these worms (Zeiss LSM710 confocal microscope, 20× magnification). At this magnification, it was not possible to capture micrographs of the entire worm (panel e). Genomic DNA from EGFP-positive worms was pooled and analyzed for programmed mutations (indels) (y-axis panels b, d). Large-sized gene deletions were apparent, up to 150 nt in female and 120 nt in the males CRISPResso2 is limited in its analysis of efficiency of HDR in this study given the donor transgene is 4.4 kb in length. All female and male worms survived until day 11 when the experiment was terminated

Figure S2. EGFP expression in *in vitro* laid eggs. *In vitro*-laid eggs (IVLE) released overnight from adult schistosomes (~200 worms) were transfected by electroporation with RNPs (three overlapping guide RNAs) and donor repair template. At transfection (day 0), the ILVEs contained a few parasite cells and germ cells. Transfected IVLEs were maintained in high nutrition medium for 10 days. At this point, some of the eggs (<10%) contained the fully developed miracidium (panel a). EGFP expression in the miracidium (green arrow) was apparent in a few of these eggs of the population (<1%) in culture (Zeiss LSM 710, 20X magnification). In similar fashion to the outcome with the adult schistosomes (Fig. S1), programmed deletions were seen in the genome of these eggs, following CRISPResso2 analysis of the sequence reads (b).

References

1 Zhu, H., Li, C. & Gao, C. Applications of CRISPR-Cas in agriculture and plant biotechnology. *Nat Rev Mol Cell Biol* **21**, 661-677 (2020). https://doi.org:10.1038/s41580-020-00288-9

2 Zaib, S., Saleem, M. A. & Khan, I. CRISPR-Cas9 Genome Engineering: Trends in Medicine and Health. *Mini Rev Med Chem* **22**, 410-421 (2022). https://doi.org:10.2174/1389557521666210913112030

3 Wang, F. & Qi, L. S. Applications of CRISPR Genome Engineering in Cell Biology. *Trends Cell Biol* **26**, 875-888 (2016). https://doi.org:10.1016/j.tcb.2016.08.004

4 Hotez, P. J. *et al.* Helminth infections: the great neglected tropical diseases. *J Clin Invest* **118**, 1311-1321 (2008). https://doi.org:10.1172/JCI34261

5 Naldini, L. *et al.* In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272**, 263-267 (1996). https://doi.org:10.1126/science.272.5259.263

6 Li, X. *et al.* piggyBac transposase tools for genome engineering. *Proc Natl Acad Sci U S A* **110**, E2279-2287 (2013). https://doi.org:10.1073/pnas.1305987110

7 Guo, J. C. *et al.* Highly Efficient CRISPR/Cas9-Mediated Homologous Recombination Promotes the Rapid Generation of Bacterial Artificial Chromosomes of Pseudorabies Virus. *Front Microbiol* **7**, 2110 (2016). https://doi.org:10.3389/fmicb.2016.02110

8 Ittiprasert, W. *et al.* Programmed genome editing of the omega-1 ribonuclease of the blood fluke, Schistosoma mansoni. *Elife* **8:e41337** (2019). https://doi.org:10.7554/eLife.41337

9 Ittiprasert, W. *et al.* RNA-Guided AsCas12a- and SpCas9-Catalyzed Knockout and Homology Directed Repair of the Omega-1 Locus of the Human Blood Fluke, Schistosoma mansoni. *Int J Mol Sci* **23** (2022). https://doi.org:10.3390/ijms23020631

10 Jang, D. E. *et al.* Multiple sgRNAs with overlapping sequences enhance CRISPR/Cas9-mediated knock-in efficiency. *Exp Mol Med* **50**, 1-9 (2018). https://doi.org:10.1038/s12276-018-0037-x

11 Acosta, S., Fiore, L., Carota, I. A. & Oliver, G. Use of two gRNAs for CRISPR/Cas9 improves bi-allelic homologous recombination efficiency in mouse embryonic stem cells. *Genesis* **56**, e23212 (2018). https://doi.org:10.1002/dvg.23212

12 Gutierrez-Triana, J. A. *et al.* Efficient single-copy HDR by 5' modified long dsDNA donors. *Elife* **7** (2018). https://doi.org:10.7554/eLife.39468

13 Sadelain, M., Papapetrou, E. P. & Bushman, F. D. Safe harbours for the integration of new DNA in the human genome. *Nat Rev Cancer* **12**, 51-58 (2011). https://doi.org:10.1038/nrc3179

14 Zhang, T., Cooper, S. & Brockdorff, N. The interplay of histone modifications - writers that read. *EMBO Rep* **16**, 1467-1481 (2015). https://doi.org:10.15252/embr.201540945

15 Schroder, A. R. *et al.* HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110**, 521-529 (2002). https://doi.org:10.1016/s0092-8674(02)00864-4

16 Suttiprapa, S. *et al.* HIV-1 Integrates Widely throughout the Genome of the Human Blood Fluke Schistosoma mansoni. *PLoS Pathog* **12**, e1005931 (2016). https://doi.org:10.1371/journal.ppat.1005931

17 Berriman, M. *et al.* The genome of the blood fluke Schistosoma mansoni. *Nature* **460**, 352-358 (2009). https://doi.org:10.1038/nature08160

18 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). https://doi.org:10.1371/journal.pntd.0001455

19 Howe, K. L. *et al.* WormBase 2016: expanding to enable helminth genomic research. *Nucleic Acids Res* **44**, D774-780 (2016). https://doi.org:10.1093/nar/gkv1217

20 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). https://doi.org:10.1016/j.molbiopara.2016.11.005

21 Luo, H. *et al.* DEG 15, an update of the Database of Essential Genes that includes built-in analysis tools. *Nucleic Acids Res* **49**, D677-D686 (2021). https://doi.org:10.1093/nar/gkaa917

22 Labun, K. *et al.* CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res* **47**, W171-W174 (2019). https://doi.org:10.1093/nar/gkz365

23 Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M. & Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res* **42**, W401-407 (2014). https://doi.org:10.1093/nar/gku410

24 Bloh, K. *et al.* Deconvolution of Complex DNA Repair (DECODR): Establishing a Novel Deconvolution Algorithm for Comprehensive Analysis of CRISPR-Edited Sanger Sequencing Data. *CRISPR J* **4**, 120-131 (2021). https://doi.org:10.1089/crispr.2020.0022

25 Marcek Chorvatova, A., Kirchnerova, J., Cagalinec, M., Mateasik, A. & Chorvat, D., Jr. Spectrally and spatially resolved laser-induced photobleaching of endogenous flavin fluorescence in cardiac myocytes. *Cytometry A* **95**, 13-23 (2019). https://doi.org:10.1002/cyto.a.23591

26 Clement, K. *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol* **37**, 224-226 (2019). https://doi.org:10.1038/s41587-019-0032-3

27 Pinello, L. *et al.* Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat Biotechnol* **34**, 695-697 (2016). https://doi.org:10.1038/nbt.3583

28 Abudayyak, M., Jannuzzi, A. T., Ozhan, G. & Alpertunga, B. Investigation on the toxic potential of Tribulus terrestris in vitro. *Pharm Biol* **53**, 469-476 (2015). https://doi.org:10.3109/13880209.2014.924019

29 Rothe, C. *et al.* Developing Endemicity of Schistosomiasis, Corsica, France. *Emerg Infect Dis* **27** (2021). https://doi.org:10.3201/eid2701.204391

30 Hoffmann, K. F., Brindley, P. J. & Berriman, M. Medicine. Halting harmful helminths. *Science* **346**, 168-169 (2014). https://doi.org:10.1126/science.1261139

31 Cox, D. B., Platt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nat Med* **21**, 121-131 (2015). https://doi.org:10.1038/nm.3793

32 Douglas, B. *et al.* Immune System Investigation Using Parasitic Helminths. *Annu Rev Immunol* **39**, 639-665 (2021). https://doi.org:10.1146/annurev-immunol-093019-122827

33 Pavani, G. & Amendola, M. Targeted Gene Delivery: Where to Land. *Front Genome Ed* **2**, 609650 (2021). https://doi.org:10.3389/fgeed.2020.609650

34 Xu, Y. Z. & Dresden, M. H. Schistosoma mansoni: egg morphology and hatchability. *J Parasitol* **75**, 481-483 (1989).

35 Katsumata, T., Shimada, M., Sato, K. & Aoki, Y. Possible involvement of calcium ions in the hatching of Schistosoma mansoni eggs in water. *J Parasitol* **74**, 1040-1041 (1988).

36 Kassim, O. & Gibertson, D. E. Hatching of Schistosoma mansoni eggs and observations on motility of miracidia. *J Parasitol* **62**, 715-720 (1976).

37 Gabriel, R., Schmidt, M. & von Kalle, C. Integration of retroviral vectors. *Curr Opin Immunol* **24**, 592-597 (2012). https://doi.org:10.1016/j.coi.2012.08.006

38 Mann, V. H., Suttiprapa, S., Skinner, D. E., Brindley, P. J. & Rinaldi, G. Pseudotyped murine leukemia virus for schistosome transgenesis: approaches, methods and perspectives. *Transgenic Res* **23**, 539-556 (2014). https://doi.org:10.1007/s11248-013-9779-3

39 Dai, F. *et al.* Sodium-bile acid co-transporter is crucial for survival of a carcinogenic liver fluke Clonorchis sinensis in the bile. *PLoS Negl Trop Dis* **14**, e0008952 (2020). https://doi.org:10.1371/journal.pntd.0008952

40 Jurberg, A. D. *et al.* The embryonic development of Schistosoma mansoni eggs: proposal for a new staging system. *Dev Genes Evol* **219**, 219-234 (2009). https://doi.org:10.1007/s00427-009-0285-9

41 Rinaldi, G. *et al.* Germline transgenesis and insertional mutagenesis in Schistosoma mansoni mediated by murine leukemia virus. *PLoS Pathog* **8**, e1002820 (2012). https://doi.org:10.1371/journal.ppat.1002820

42 Farbiak, L. *et al.* All-In-One Dendrimer-Based Lipid Nanoparticles Enable Precise HDR-Mediated Gene Editing In Vivo. *Adv Mater* **33**, e2006619 (2021). https://doi.org:10.1002/adma.202006619

43 Kazemian, P. *et al.* Lipid-Nanoparticle-Based Delivery of CRISPR/Cas9 Genome-Editing Components. *Mol Pharm* **19**, 1669-1686 (2022). https://doi.org:10.1021/acs.molpharmaceut.1c00916

44 Prakash, G. *et al.* Microfluidic fabrication of lipid nanoparticles for the delivery of nucleic acids. *Adv Drug Deliv Rev* **184**, 114197 (2022). https://doi.org:10.1016/j.addr.2022.114197

45 Walther, J. *et al.* Impact of Formulation Conditions on Lipid Nanoparticle Characteristics and Functional Delivery of CRISPR RNP for Gene Knock-Out and Correction. *Pharmaceutics* **14** (2022). https://doi.org:10.3390/pharmaceutics14010213

46 Yan, J., Kang, D. D. & Dong, Y. Harnessing lipid nanoparticles for efficient CRISPR delivery. *Biomater Sci* **9**, 6001-6011 (2021). https://doi.org:10.1039/d1bm00537e

47 Diez, B. *et al.* Therapeutic gene editing in CD34(+) hematopoietic progenitors from Fanconi anemia patients. *EMBO Mol Med* **9**, 1574-1588 (2017). https://doi.org:10.15252/emmm.201707540

48 Roquis, D. *et al.* Histone methylation changes are required for life cycle progression in the human parasite Schistosoma mansoni. *PLoS Pathog* **14**, e1007066 (2018). https://doi.org:10.1371/journal.ppat.1007066

49 Smith, J. P. & Sheffield, N. C. Analytical Approaches for ATAC-seq Data Analysis. *Curr Protoc Hum Genet* **106**, e101 (2020). https://doi.org:10.1002/cphg.101

50 Taudt, A., Colome-Tatche, M. & Johannes, F. Genetic sources of population epigenomic variation. *Nat Rev Genet* **17**, 319-332 (2016). https://doi.org:10.1038/nrg.2016.45

51 Suttiprapa, S., Rinaldi, G. & Brindley, P. Genetic manipulation of schistosomes - Progress with integration competent vectors. *Parasitology* **139**, 641-650 (2011). https://doi.org:10.1017/S003118201100134X

52 Pavani, G. & Amendola, M. Targeted Gene Delivery: Where to Land. *Front Genome Ed* **2**, 609650 (2020). https://doi.org:10.3389/fgeed.2020.609650

53 Pavani, G. & Amendola, M. Corrigendum: Targeted Gene Delivery: Where to Land. *Front Genome Ed* **3**, 682171 (2021). https://doi.org:10.3389/fgeed.2021.682171

54 Cortes, M. F. *et al.* Community-acquired methicillin-resistant Staphylococcus aureus from ST1 lineage harboring a new SCCmec IV subtype (SCCmec IVm) containing the tetK gene. *Infect Drug Resist* **11**, 2583-2592 (2018). https://doi.org:10.2147/IDR.S175079

55 Mann, V. H., Morales, M. E., Rinaldi, G. & Brindley, P. J. Culture for genetic manipulation of developmental stages of Schistosoma mansoni. *Parasitology* **137**, 451-462 (2010). https://doi.org:10.1017/S0031182009991211

56 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 (Pt 1)**, 29-32 (1997). https://doi.org:10.1017/s0031182097001091

57 Labun, K., Montague, T. G., Gagnon, J. A., Thyme, S. B. & Valen, E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res* **44**, W272-276 (2016). https://doi.org:10.1093/nar/gkw398

58 Hulme, B. J. *et al.* Schistosoma mansoni alpha-N-acetylgalactosaminidase (SmNAGAL) regulates coordinated parasite movement and egg production. *PLoS Pathog* **18**, e1009828 (2022). https://doi.org:10.1371/journal.ppat.1009828

59 Wilfinger, W. W., Mackey, K. & Chomczynski, P. Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity. *Biotechniques* **22**, 474-476, 478-481 (1997). https://doi.org:10.2144/97223st01

60 Domingo, M., Mais, R. F., Weiskopf, R. & Fink, S. Detection of schistosome ova by dark field fluorescence microscopy. *Gastroenterology* **54**, 884-886 (1968).

61 Lu, Z. *et al.* Isolation, enrichment and primary characterisation of vitelline cells from Schistosoma mansoni obtained by the organ isolation method. *Int J Parasitol* **45**, 663-672 (2015). https://doi.org:10.1016/j.ijpara.2015.04.002

62 Collins, J. J., 3rd, King, R. S., Cogswell, A., Williams, D. L. & Newmark, P. A. An atlas for Schistosoma mansoni organs and life-cycle stages using cell type-specific markers and confocal microscopy. *PLoS Negl Trop Dis* **5**, e1009 (2011). https://doi.org:10.1371/journal.pntd.0001009

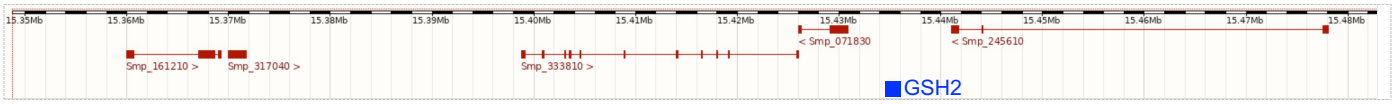
63 Wu, Q., Feng, Z. & Hu, W. Reduction of autofluorescence in whole adult worms of Schistosoma japonicum for immunofluorescence assay. *Parasit Vectors* **14**, 532 (2021). https://doi.org:10.1186/s13071-021-05027-3

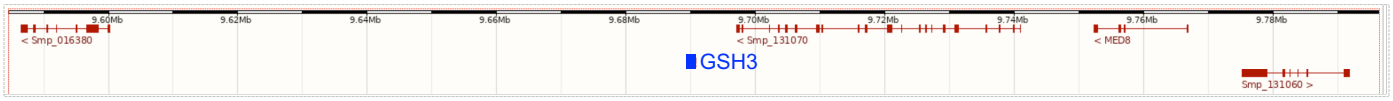
Figure 1

a

Chromosome 3 window

Chromosome 2 windows





b

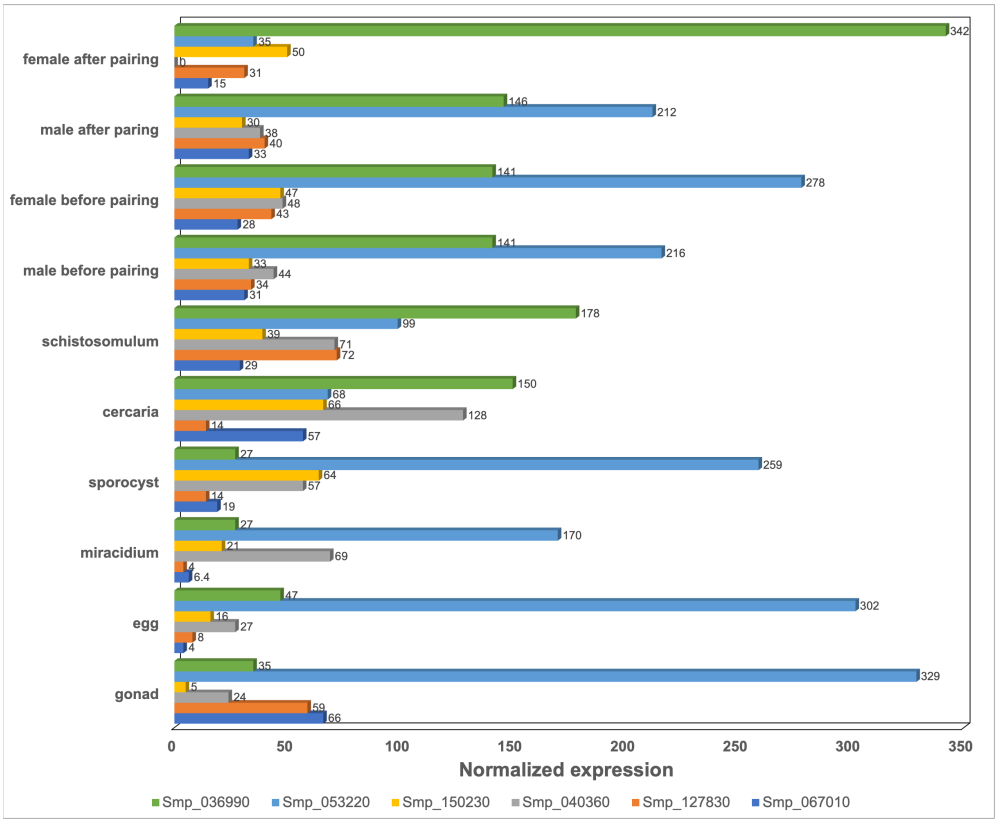


Figure 2

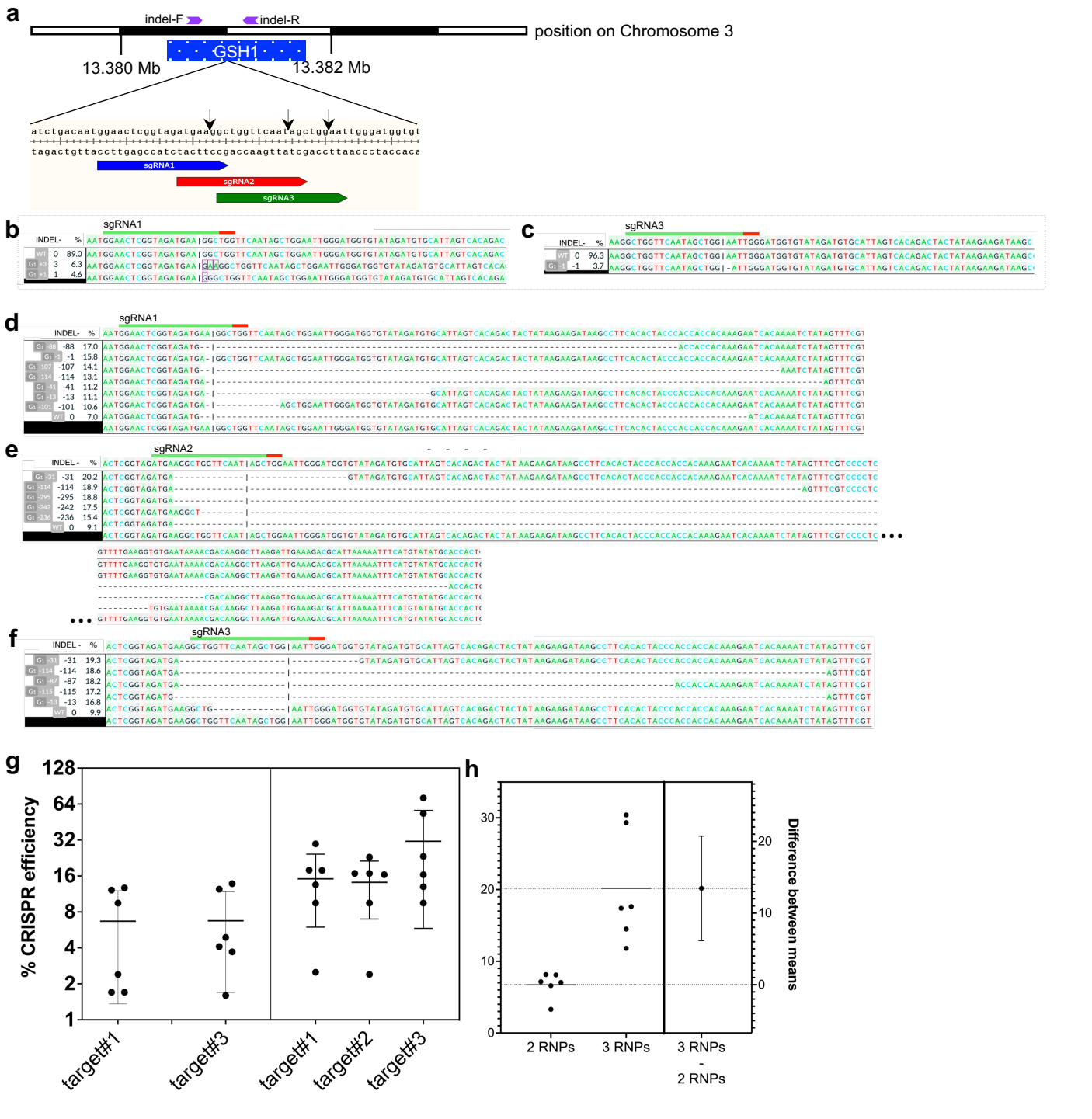

a
indel-F
indel-R
position on Chromosome 3
GSH1
13.380 Mb
13.382 Mb
sgRNA1
sgRNA2
sgRNA3
b
c
d
e
f
g
% CRISPR efficiency
target#1
target#3
target#1
target#2
target#3
h
Difference between means
2 RNPs
3 RNPs
3 RNPs - 2 RNPs

Figure 3

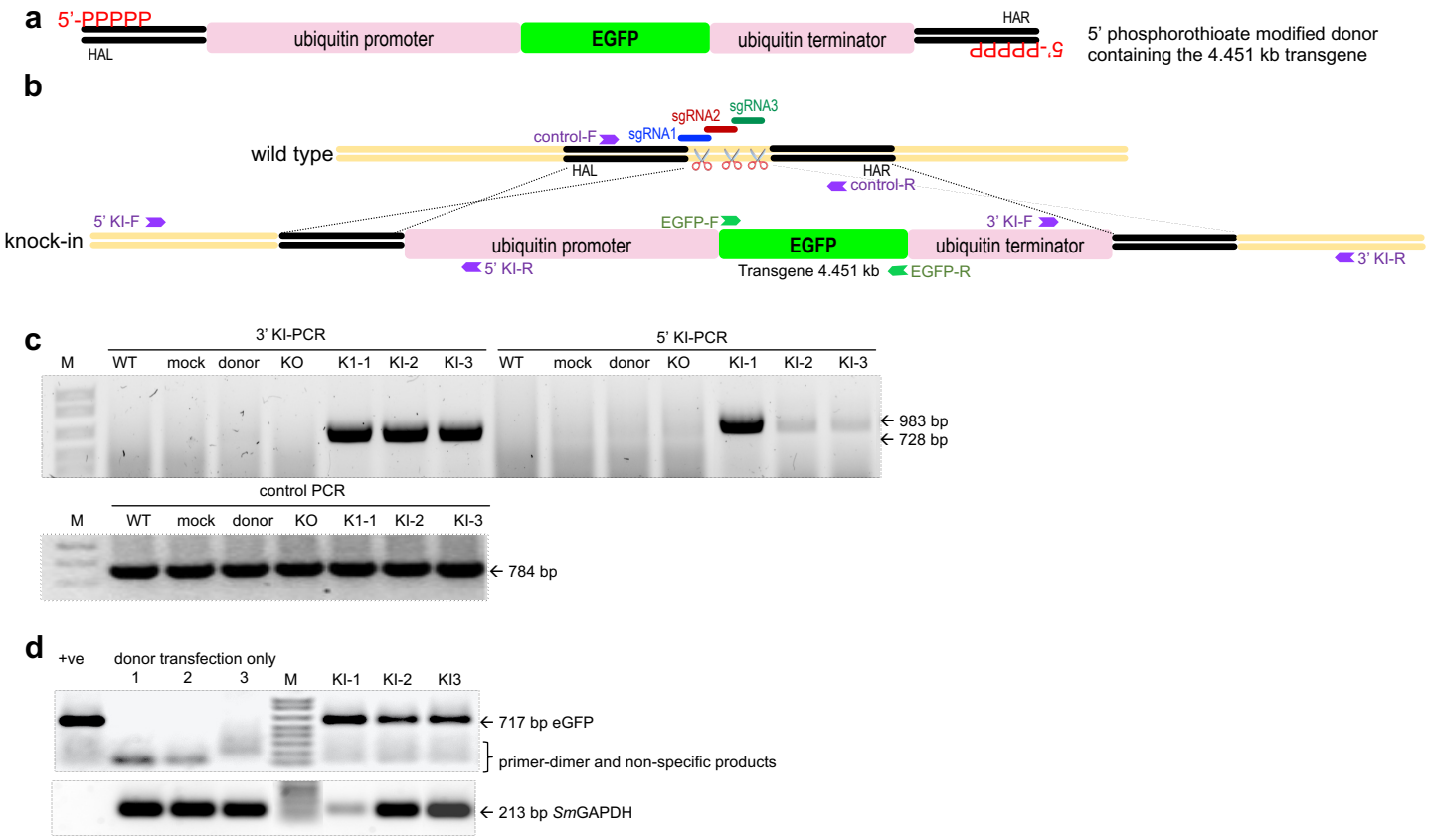
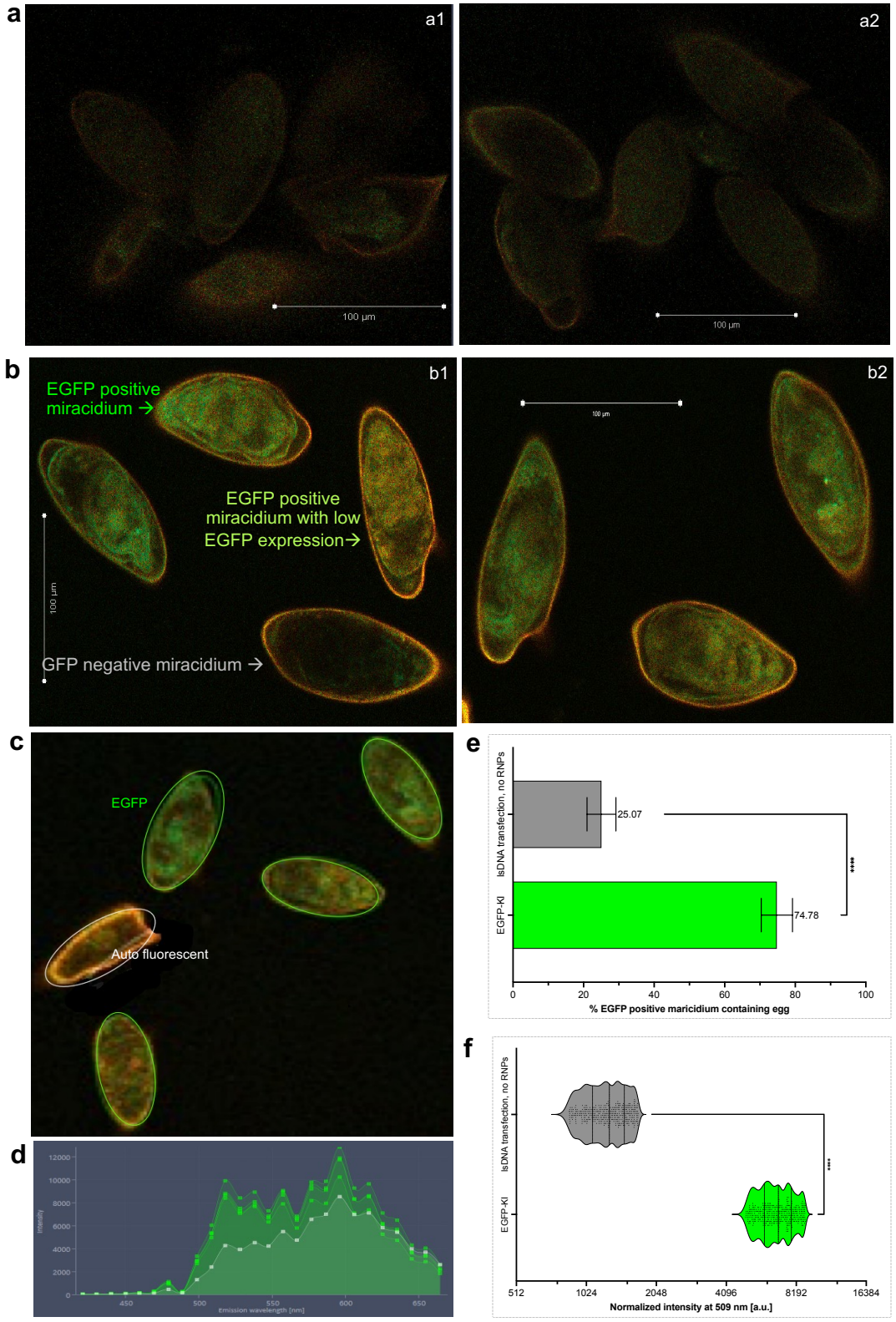
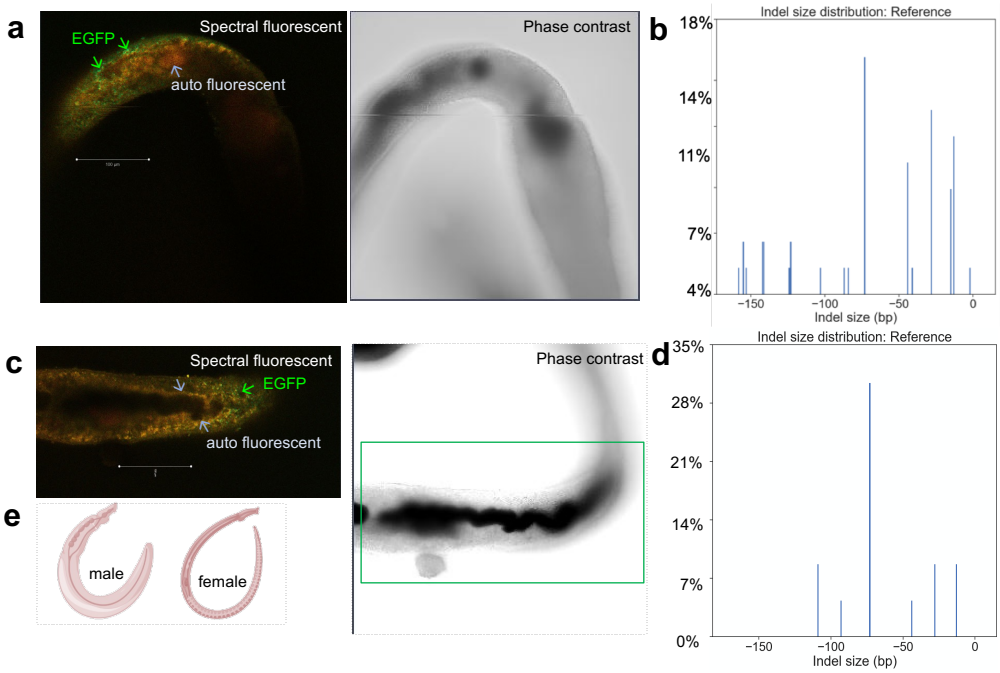

a
5'-PPPPP
HAL
ubiquitin promoter
EGFP
ubiquitin terminator
HAR
5'-PPPPP
5' phosphorothioate modified donor containing the 4.451 kb transgene
b
wild type
control-F
sgRNA1
sgRNA2
sgRNA3
HAL
HAR
control-R
knock-in
5' KI-F
EGFP-F
3' KI-F
ubiquitin promoter
EGFP
ubiquitin terminator
5' KI-R
Transgene 4.451 kb
EGFP-R
3' KI-R
c
3' KI-PCR
5' KI-PCR
M
WT
mock
donor
KO
KI-1
KI-2
KI-3
983 bp
728 bp
control PCR
784 bp
d
+ve
donor transfection only
1
2
3
M
KI-1
KI-2
KI3
717 bp eGFP
primer-dimer and non-specific products
213 bp SmGAPDH

Figure 4


a
a1
100 µm
a2
100 µm
b
b1
EGFP positive miracidium →
EGFP positive miracidium with low EGFP expression→
GFP negative miracidium →
100 µm
b2
100 µm
c
EGFP
Auto fluorescent
d
Intensity
Emission wavelength [nm]
e
lsDNA transfection, no RNPs
EGFP-KI
25.07
74.78
****
% EGFP positive maricidium containing egg
f
lsDNA transfection, no RNPs
EGFP-KI
****
Normalized intensity at 509 nm [a.u.]

Supplement figure 1



Supplement figure 2

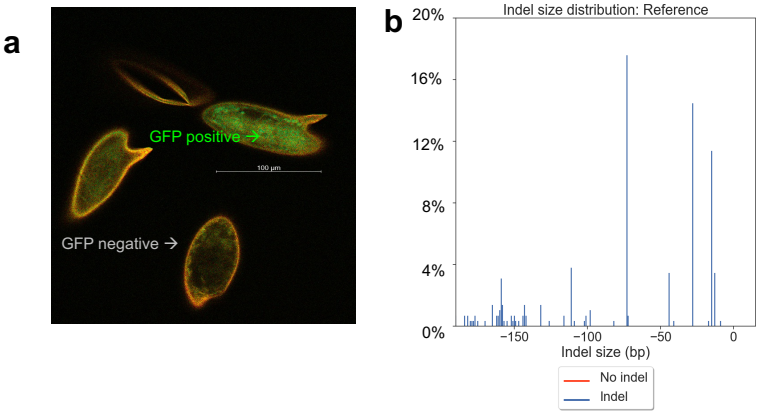


Table 1. A Table shows GSH criteria and rationale to computationally predict GSH sites in the *S. mansoni* genome. Candidate GSH sites indicating the chromosome locations and lengths of six candidate intragenic GSH sites and four candidate intergenic GSH sites.

GSH criteria	Predicted *S. mansoni* GSH		
	Accession no.	Protein coding	Location on the genome
Intragenic GSH			
1) Close to peaks of H3H4me3, a histone modification that is associated with euchromatin and transcription start sites	• Smp_053220 • Smp_150230	• aldo keto reductase family, member B4 • metal tolerance protein C3	• Chromosome 4: 1,244,090-1,260,835 forward strand (16.745 Kb) • Chromosome 3: 27,720,233-27,752,200 forward strand (31.967 Kb)
2) Do not contain H3K27me3, a histone modification that is associated wit heterochromatin in any of the life stages	• Smp_040360 • Smp_127830	• endoplasmic reticulum Golgi intermedia • actin protein ARP2:3 complex subunit	• Chromosome 1:76,379,080-76,396,355 forward strand (17.275 Kb) • Chromosome 7:5,605,098-5,634,189 reverse strand (29.091 Kb)
3) Deliver an ATAC-seq signal	• Smp_067010	• RUN domain containing protein 1	• Chromosome 4:22,148,377-22,197,632 reverse strand (49.255 Kb)
4) Possibility for viral integration site as in human cell line shown preferentially into euchromatin	• Smp_036990	• endophilin III	• Chromosome 4:4,502,471-4,557,287 reverse strand (54.816 Kb)
	Gene ID	Protein coding nearby	Location on the genome
Intergenic GSH			
1) Unique sequences	GSH1	• n/a; Smp_052890 • copper transport protein atox1-related; Smp_150460	• Chromosome 3:13380432-13381848 (1,416 bp)
2) Locate outside annotated genes and long non-coding RNA	GSH2	• 6-phosphogluconate dehydrogenase; Smp_33810 • n/a; Smp_071830, Smp_245610	• Chromosome 2:15434976-15435945 (970 bp) • Chromosome 2:9689988-9690739 (752 bp)
3) Locate outside putative promotor regions where more than 2Kb upstream of the TTS			
4) Close to peaks of H3K4me3 in all parasite stages	GSH3	• cytohesin-related guanine nucleotide-exchange protein; Smp_016380 • condensin complex subunit 1; Smp_131070	• Chromosome 3:13381901-13382038 (138 bp)
5) Do not contain h3K27me3			
6) Overlapping ATAC-sequence positive sites	GSH4	• Smp_052890, Smp_150460	
7) Within 11 Kb of HIV integration site			