Targeted insertion and reporter transgene activity at a gene safe harbor of the 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, School of Medicine & Health Sciences, George Washington University, Washington, D.C. 20037

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

[3] IHPE, Univ Perpignan Via Domitia, CNRS, IFREMER, Univ Montpellier, Perpignan, 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, 12000, Thailand

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

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

[#] Contributed equally

[*] Correspondence: Wannaporn Ittiprasert, wannaporni@gwu.edu; Paul J. Brindley, pbrindley@gwu.edu

Abstract

Identification, characterization, and transgene insertion into a genomic safe harbor site (GSH) is expected to facilitate consistent transgene activity without disruption to the host cell genome while supporting transgene activity. The location of GSH sites was predicted in the human blood fluke, *Schistosoma mansoni* using a computational approach, and after which a CRISPR-focused protocol was developed for insertion of a reporter transgene into a representative GSH. Ribonuclear protein complexes of Cas9 nuclease, three overlapping guide RNAs, and phosphorothioate-modified, double-stranded donor DNA encoding green fluorescent protein driven by a ubiquitously expressed promoter were employed. Gene knockout efficiencies of 24% and reporter knockin transgene fluorescence of >70% of gene-edited schistosome eggs were observed. These outcomes advance functional genomics for schistosomes by providing a tractable path towards transgenic helminths using homology directed repair catalyzed transgene insertion. This approach should be adaptable to helminths generally.

Main

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) technology has revolutionized functional genomics in in biology, medicine, and agriculture[1-3]. Transgenesis approaches are integral in diverse applications including gene therapy, biotherapeutics, deciphering host-pathogen interactions, and enhancements in agricultural production. 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 the helminths, which generally are responsible for neglected tropical diseases that cause substantial mortality and disease[4]. These mostly occur in the global south and are responsible for a disease burden that exceeds that caused by malaria and tuberculosis. In the current post-genomics era, access to CRISPR-based transgenesis protocols for functionally characterizing parasitic helminth putative drug and vaccine targets is a public health research priority. Importantly, progress of this technology in one helminth, for example the schistosome will facilitate its use in other major invertebrate clades found within the 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]. For example, these latter approaches could negatively lead to genetic instability, multi-copy insertion, unstable expression or inactivation of the transgene and interference with the endogenous gene under investigation. In the process of genome editing, the double stranded breaks (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) pathways. 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 donors. 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].

Here, we undertook a computational investigation of potential genome safe harbor sites in *S. mansoni*, and adapted CRISPR/Cas9-based approaches to insert a reporter transgene into one of the predicted GSH situated in euchromatin of chromosome 3, GSH1. The targeted region was free of repetitive sequences and neighboring long non-coding regions, a situation likely to minimize off-target effects of CRISPR/Cas activity. Multiple prototypic adjacent motifs within this region were targeted with two and three programmed single guide RNAs deployed in unison to enhance mutational efficiency and homology directed repair (HDR) in the presence of a phosphorothioate-modified donor DNA template. A knock-in (KI) efficiency of 75% was observed by reporter transgene green protein fluorescence in miracidia within the eggs, when the RNPs and donor transgene were delivered.

Results

Genome safe harbors predicted in the schistosome genome

We undertook a computational search to predict intergenic genome safe harbor (GSH) sites in *S. mansoni* based on accepted criteria, newly introduced principles[13], and genome sequence resources available for this schistosome, that could satisfy benign and stable gene expression. Based on criteria presented in figure 1a and detailed in the Methods section, in brief regions proximal to H3K4me3 marks, regions without H3K27me3, ATAC seq-positive regions, regions free of repetitive sequences, and regions known to be accessible to lentiviral (HIV) insertion, the draft genome of *S. mansoni*, version 9[14-17] included 6,884 regions with enrichment of H3K4me3 in the absence of H3K27me3, in available developmental stages, i.e. positive for H3K4me3 but not K3K27me3. In adult schistosomes, we consistently found 10,533 ATAC-Seq positive regions. There were 4,027 ATAC-Seq regions that overlapped with H3K4me3, and 2,915 regions overlapped with the ATAC-Seq regions and H3K4me3 but not H3K27me3. Forty-two unambiguous HIV integration sites were identified in these regions, and eight genes were < 4 kb upstream or downstream from these integration sites[18]. To identify intergenic GSH, rather than intragenic GSH given the transgene insertion in the former should be safer[13], we located 10,149 intergenic regions[19]. There were 9,985 regions beyond 2 kb upstream of the transcription start site (TSS) and within 11 kb of an HIV integration site, and 8,837 regions outside long non-coding-RNA (lncRNA), which were intersected to 95,587 unique intergenic regions beyond 2 kb of lncRNAs of ≥ 100 bp in length. Next, 200 regions were identified intersecting with the merged ATAC/ H3K4me3 signals and, of these, four were situated ≤ 11 kb from HIV integration sites. These four regions satisfied our criteria. These were termed GSH1 (1,416 bp in length; location, chromosome 3:13380432-13381848), GSH2 (970 bp; chromosome 2: 15434976-15435954), GSH3 (752 bp; chromosome 2: 9689988-9690739), and GSH4 (138 bp; chromosome 3: 13381901-13382038), respectively (Fig. 1), with the naming based in rank order on their size from longest (GSH1) to shortest (GSH4). We note that several protein-coding gene loci were situated in the vicinity of these gene-free GSHs, although these genes were > 2kb distant from any GSH: Smp_052890 [uncharacterized protein], Smp_150460 [copper transport protein], Smp_071830 [uncharacterized protein], Smp_245610 [uncharacterized protein] and Smp_131070 [condensing complex] (Figs. 1b-d). Most of there were as yet uncharacterized

proteins, and may be non-essential genes based on ortholog to essential genes known from model eukaryotes[20].

Efficiency of programmed mutation at GSH1 enhanced by multiple guide RNAs

Having predicted the location of these GSHs, we proceeded to investigate the efficiency of programmed mutation and reporter transgene activity at GSH1, chromosome 3: 13380432-13381848. Multiple and/or overlapping guide RNAs were employed, an approach that enhances KI efficiency in some mammalian cell lines and embryos[10,11]. Among guide RNAs (gRNA) exhibiting on-target specificity for GSH1, three overlapping synthetic guide RNAs, sgRNA1, sgRNA2 and sgRNA3, which lacked self-complementarity and off-target matches to the reference *S. mansoni* genome (Figs. 1b,e, Fig. 2a), were selected from among CRISPR/Cas9 target sites predicted by CHOPCHOP [21,22]. The ribonucleoprotein complexes (RNPs) of Cas9 nuclease and sgRNA were assembled, after which four discrete mixtures of the three RNPs were used. Three of the mixtures included two RNPs, *i.e.* dual RNPs mixtures (RNP1+RNP2, RNP2+RNP3, and RNP1+RNP3) and the fourth included the three RNPs, i.e. triple RNPs (RNP1+RNP2+RNP3).

The mixture of the dual and triple RNPs, in the presence of the donor DNA template encoding EGFP, was delivered to schistosome eggs by electroporation (EP). The transfected eggs were maintained in culture for 15 days, after which EGFP expression was quantified. Thereafter, genomic DNAs and total RNAs were extracted from the eggs. Efficiency of genome editing, both in mock-treated controls and experimental groups, was assessed using DECODR[23] analysis of Sanger sequence chromatograms of amplicons that spanned the CRISPR/Cas9 programmed DSBs. Analysis of PCR products amplified from schistosome genomic DNA using primers (indel-F, indel-R) flanking the DSBs (Fig. 2a) revealed programmed knockout (KO) efficiency, as assessed by indel (insertions-deletions)-bearing alleles resulting from the dual gRNAs, as follows: KO frequencies at GSH1 of 5.4% (range of 0.8-10.4%), 3.6% (range of 1.2-19.3%) and 12.6% (range of 4.9-19.3%) for RNP1+ RNP2, RNP2 + RNP3 and RNP1 + RNP3, respectively (Fig. 2b). The dual gRNA RNPs mixtures induced short deletions of one to several nucleotides in length at the predicted DSB sites for sgRNA1, 2 and/or 3 (Figs. 2c-e). Mutations were not evident in amplicons from the control groups, *i.e.* wild type (no treatment) and mock treatment (not shown). The triple RNPs mixture, where the three sgRNAs shared at least six overlapping nucleotides, induced 23.9% KO (range of 2.4 - 71.9%), higher than any mixture of the dual RNPs (Fig. 2b).

The CRISPR efficiencies at each target sites varied among the RNP mixtures, as did the indel profiles (Fig. S1). KO efficiency at each target resulting from dual RNPs was lower generally than for the triple guide RNPs. At target site 1, estimated KO rates were 6.9% (4.1-10.4%), 13.8% (6-19.3%) and 26.5% (14-33.2%) using RNP1+RNP2, RNP1+RNP3 and RNP1+RNP2+RNP3, respectively. There were 3.9% (0.8%-5.8%), 1.8% (1.2%-2.4%) and 14.2% (2.4%-16.8%) mutations on target 2 using RNP1+RNP2, RNP2+RNP3 and RNAP1+RNP2+RNP3, respectively. Mutation efficiencies at target 3 using RNP2+RNP3 and RNP1+RNP3 were 13.9% (6%-19.3%) and 11.3% (6%-19.3%). The mutational profile of the indels, as assessed from Sanger sequencing chromatograms of the target amplicons, was mostly deletions rather than insertions (Figs. S1b-e). Notably, deletions up to 115 nt in length were identified among the biological replicates with the triple overlapping RNPs (Fig. 2e, Fig. S1e).

KO efficiency was assessed using at least five independent biological replicates: Combining three overlapping sRNAs induced an aggregate mutation efficiency (23.9%) higher than that obtained with any of the dual RNPs, 5.4%, 3.6%, 12.6% ($P \leq 0.001$, one-way ANOVA). The KO efficiency of the RNP1+RNP3 group was higher than that of either of the other dual RNPs ($P \leq 0.01$) (Fig. 2b).

Three overlapping guide RNAs enhanced efficiency of CRISPR knockin

As multiple sgRNAs with overlapping sequences can enhance CRISPR/Cas9-mediated HDR efficiency[10] and given that triple overlapping sgRNAs performed better than dual gRNAs in initiating programmed mutation at GSH1 in eggs (Fig. 2b), we investigated KI of a reporter transgene at GSH1 with the triple overlapping sgRNA/RNPs mixture (Figs. 3a,b). We employed the gene encoding enhanced green fluorescent protein (EGFP) driven by the promoter of the endogenous *S. mansoni* ubiquitin gene and its cognate terminator region as the repair template for programmed HDR (Fig. 3a). The donor template included symmetrical homology arms specific for GSH1, located on the 5´-flank of target site 1 and the 3´-flanking region of target site 3 (Fig. 3b). The donor template was delivered as linearized long double-strand DNA (lsDNA) of 4,451 bp in length. Aiming to optimize precise and efficient single-copy integration of the donor transgene into GSH1 by HDR, the 5´ termini of the DNA donor amplicons were chemically modified[12] to shield the donor template from multimerization and from integration at the DSB via the non-homologous end-joining (NHEJ) repair pathway (Fig. 3a).

At the outset, we investigated the impact of length of the homology arms (HA) by comparing the performance of the donor repair template bearing homology arms of increasing length of 200, 400 and 600 bp. Dual (RNP1+RNP3) and triple (RNP1+RNP2+RNP3) sgRNA mixtures were used in this investigation. EGFP expression was not evident in eggs electroporated with lsDNA donors with 200 and 400 bp HAs at about day 5 after transfection (not shown). By contrast, we observed a few EGFP-positive eggs, although in < 1% of them, with the lsDNA donor with 600 bp HA (not shown). Subsequently, given this outcome, we changed the focus of the investigation for EGFP expression to transfection with the donor transgene flanked by 600 bp HA using the triple guides RNP mixture (RNP1+RNP2+RNP3) and to monitoring expression of the EGFP transgene for up to 15 days. Thereafter, on examination using spectrally resolved, confocal laser scanning microscopy, the EGFP signal was detected in the eggs of the experimental group, which received the CRISPR materials including the lsDNA donor with 600 bp HA. EGFP signal remained until 15 days, when the experiment ended. EGFP signals were not observed in the negative control groups, although the autofluorescence characteristic of schistosome eggs was apparent[24]. The EGFP signal was detected in the lsDNA donor control (without RNPs) for several days, indicating that extrachromosomal lsDNA expressed EGFP transiently after the transfection.

Next, we investigated the nature of the programmed KI at GSH1 by PCR-based analysis for the presence of the expected amplicons spanning the 5´ and 3´ flanks of the donor transgene i.e., flanking the ubiquitin promoter, EGFP, and the ubiquitin terminator sequences. At the F 5´ flank, we used a forward primer specific for the genome upstream of the 5´ HA paired with a reverse primer specific for the ubiquitin promoter (Fig. 3b). For the 3´ integration junction, were used a reverse primer specific for a site downstream of the 3´ terminus of the HA paired with a forward primer specific for the ubiquitin terminator. Fragments representing the 3´ KI and 5´ KI

integration regions of 983 bp and 728 bp, respectively, were observed in the experiment 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 group, although some variability in transcript abundance among the biological replicates was seen based on the signals obtained for *Sm*GAPDH, which served as the reference gene for integrity of the RNAs (Fig. 3d).

Reporter transgene expressed in schistosome eggs

EGFP positivity and intensity were quantified using spectral laser scanning confocal fluorescence microscopy[24]. Active transgene expression was confirmed within miracidia developing inside transfected eggs (Figs. 4a, b). EGFP appeared to be expressed by numerous diverse cells throughout the developing larva whereas morphological malformation was not observed in transgenic eggs and their enclosed larvae. More intense EGFP fluorescence was consistently recorded and quantified at 509 nm in eggs from the experimental treatment group (Figs. 4b1, 2) than the mock control eggs and in eggs transfected solely with donor template (Figs. 4a1, 2). Subsequently, on day 15 following transfection with the repair template in the presence or absence of the RNPs mixture, we quantified EGFP intensity in eggs that contained a miracidium by normalization with the EGFP signal from lsDNA only transfected eggs. Fluorescence intensity differed markedly between these two groups; by 15 days after transfection, 75% of miracidia in eggs from the experimental group (RNPs/lsDNA) emitted EGFP fluorescence whereas 25% of eggs containing a miracidium transfected with the donor lsDNA repair template only emitted EGFP (Figs. 4c, d) ($P \leq 0.001$; $n = 402$ eggs in experimental group, $n = 397$ eggs in the lsDNA only group, aggregated from four independent biological replicates). In addition, by scoring the intensity of fluorescence at 509 nm[24],the emission wavelength for EGFP, as shown figures 4c1,c3, c4 (green curve), established by subtracting the signal at 509 nm from the autofluorescence spectrum, especially at the egg shell, the EGFP-specific signal in the control lsDNA donor repair template treatment group, mean = 1,290 arbitrary units (AU) [25] (range, 856-1,1713) was lower than the experimental group transfected with the triple, overlapping guide RNPs mixture in the presence of the donor repair template, 6,905 AU (range, 4,973 to 8,963)($P \leq 0.001$) (Fig. S2). Moreover, emission of EGFP was not detected in the control groups, *i.e.* mock-transfected and WT eggs (not shown). Lastly, diverse cells and tissues of the developing miracidium expressed the fluorescence reporter gene, i.e. EGFP was not obviously localized to specific cells (Figs. 4b,c).

Discussion

To advance functional genomics for helminths, here we predicted 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, and quantified transgene activity using confocal imaging of emission spectra specific for EGFP. Programmed editing with RNPs and three overlapping guide RNAs targeting GSH1 delivered 24% mutation efficiency that was accompanied by transgene fluorescence activity in 75% of miracidia in the gene-edited schistosome eggs.

Our criteria to predict GSH in the schistosome included location in euchromatin to avoid silencing of the transgene, unique genome-target sequence to minimize off-target events, avoidance of lncRNA encoding genes, presence of epigenetic marks for open chromatin structure, and the absence of epigenetic marks indicating heterochromatin. We termed the intergenic sites GSH1, -2, -3 and -4, which were located on chromosomes 2 and 3. (*S. mansoni* has seven pairs of autosomes and one pair of sex chromosomes, Z and W). In addition, we assessed the GSH1 locus for CRISPR/Cas9 integration, gene editing and over expression of EGFP. We edited GSH1 using two or three ribonucleoprotein complexes of Cas9 endonuclease with the overlapping guide RNAs. Overlapping CRISPR target sites were selected by the CHOPCHOP web-based tool for CRISPR/Cas9 directed mutagenesis. RNPs of triple overlapping guides delivered significantly higher CRISPR/Cas9 efficiency, and larger sized deletion mutations than RNPs using dual guides. this outcome would also enhance the efficiency of HDR based on reports with rodents, where it has been speculated that shorter rather than longer indels are predominate after Cas9 catalyzed cleavage and the shorter indels enhance HDR[10,26]. Thereafter, we deployed this approach to targeted knockin of a long (4,551 bp) doubled stranded DNA donor template encoding the green fluorescence reporter driven by the *S. mansoni* ubiquitin gene promoter and terminator elements.

These findings indicated that GSH1 represented a promising safe harbor site for forward genetics-focused genomics within this schistosome. Efficient HDR resulted from the combination of multiple target site overlapping RNPs programmed to cleave GSH1 in the presence of a repair template protected by chemical modifications. Notably, 75% of mature eggs exhibited reporter transgene fluorescence in the miracidium developing within the eggshell, and significantly more fluorescence than seen in the control eggs transfected with donor template but not with the RNPs, ~5%. EGFP signals were not present in the control, untreated wild type egg, which by 9-10 days following transfection exhibits minimal background fluorescence[39]. Our approach to circumvent the autofluorescence emitted by schistosome eggs, which can confound detection of GFP transgene activity, used confocal laser spectral imaging and linear unmixing[27,28], an approach to our knowledge previously not used to facilitate quantification of GFP specific emission to resolve overlap between the GFP and endogenous fluorophores of the egg of the helminth.

This study confirmed that GSH1 is a prospective safe harbor site for germline transgenesis, although further validation will surely be informative. Limitations with the other prospective GSHs, GSH2, -3 and 4 included absence of efficient CRISPR/Cas9 sites and predicted off-targets which precluded further analysis,, although comparisons of the utility of all four GSH including with other CRISPR tools such as Cas12a[9] and base editors[29] can be considered. Although distance from known genes was one of our criteria to predict a schistosome GSH, intragenic sites rather than intergenic GSH may in some cases exhibit expedient attributes for functional genomics in *S. mansoni* where detrimental impacts to the fitness the schistosome may be far less consequential, for example for *in vitro*-focused studies. Despite drawbacks for clinical use in human gene therapy[30], several intragenic safe harbors are used in experimental human gene therapy, including *CCR5*, *AAVSI*, and *Rosa26*, and all reside within 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[31]. Intergenic sites nonetheless are inherently safer given coding regions or other elements are not disrupted by transgene insertion[30]. Recently, additional sites, named *Rogi 1* and *Rogi 2*, located

in intergenic regions, conform with additional criteria focusing on enhancement of safety for use in human gene therapy, including location distant from endogenous genes, lncRNAs, enhancers, and telomeres[13].

These findings are consequential in that they advance functional genomics and forward genetics for a hitherto unmet challenge to manipulate a pathogen of global public health significance. They confirm that transgenes can be targeted to safe integration sites to endow individual stages or populations of these pathogens with novel functions, with broad potential for basic and translational studies[32-34]. Whereas these findings deal with somatic transgenesis of the schistosome larva, use of the same approach for transfection of 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] and therefore facilitate germline transgenesis to derive lines of transgenic parasites carrying gain- or loss-of-function mutations. In addition, the editing methods developed can be adapted for knock-out approaches of genes of interest, in schistosomes, and likely other platyhelminths, for which genome sequences are available to be analyzed for prospective genome safe harbors. The information presented here enabled novel insight into efficient transgenesis and forward genetics for *S. mansoni* and will promote forward genetics approaches in functional genomics for parasitic helminths.

Online content

Methods

Computational search for gene safe harbors in *Schistosoma mansoni*

We undertook a genome analysis focusing of intergenic (gene-free) regions to identify prospective GSHs, using similar approaches as those used on the human genome[13]. We aimed to locate a GSH, a site that would facilitate stable expression of the integrated transgene free of interference from the host genome and which, in parallel, integrates and transcribes the transgene without negative consequences or loss of fitness for the host cell[33]. The search for GSHs deployed included several criteria, First, its location should be adjacent to peaks of H3K4me3, a histone modification associated with euchromatin and transcription start sites[35]. Second, it should not be near or not containing H3K27me3 in any developmental stage, a histone mark associated with heterochromatin[35]. Third, as the schistosome genome contains highly repetitive elements[15], the GSH site should be located in a unique tract of the genome sequence. Fourth, it should reside in open, euchromatic chromatin accessible to Tn5 transposase as assessed from ATAC-sequencing, which provides a positive display of transposase integration events[36]; consequently, safe harbor candidate regions should deliver an ATAC-sequence signal. Fifth, in the vicinity of known HIV integration sites, given that HIV integrates preferentially into euchromatin in human cells[37], we anticipated that HIV integration into the schistosome genome may likewise indicate a region of euchromatin (Fig. 1a).

To predict loci conforming to the criteria, pooled ChIP-seq data for H3K4me3 and K3K27me2 from previous studies were aligned against on *S. mansoni* genome data (version 9 on the date of analysis). ATAC-seq was performed as described with modification[19]. Peakcalls of ChIP-seq and ATAC-Seq were done with ChromstaR[13,35,36,38] 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* schistosomula with bedtools intersect. Thereafter, ATAC-seq data from adult male and adult female worms (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. Next, the HIV integration sites were identified by using data from ERR33833.8[51]. Reads were mapped to the lentivirus genome (HIV-1 vector pNL-3, accession AF324493.2) using Bowtie2 with default parameters. 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 for these genes were obtained using the metanalysis tool, https://meta.schisto.xyz/analysis/, of Lu and Berriman[39].

Computational searches that addressed these criteria predicted, *a priori*, gene free (intergenic)-GSH (Fig. 1), given that transgene integration into an existing gene could diminish fitness of the genetically modified cell[40,41]. 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, 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 overlapping with ATAC-seq positive sites with ≤11 kb distance from HIV integration sites also were included. (The HIV-1 genome is ~10 kb in length.) BEDtools were used to delimit 2 kb upstream regions (FlankBed). Annotations of 16,583 lncRNA were pooled from http://verjolab.usp.br/public/schMan/schMan3/macielEtAl2019/files/macielEtAt2019.bed12[42]. 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. 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 which a length ≥ 100 bp were retained. 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.

Schistosome eggs

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 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 of the George Washington University approved the protocol used for maintenance of mice and recovery of schistosomes.

Mice were euthanized six to seven wk after infection, after which schistosomes were recovered by portal vein perfusion with 150 mM NaCl, 15mM sodium citrate, pH 7.0. The liver was resected, homogenized with a. tissue blender, and the homogenate incubated with collagenase at 37°C for 18 h. Thereafter, schistosome eggs from the digested livers were recovered by Percoll gradient centrifugation, as described[43]. Eggs isolated from livers, termed "liver eggs", LE[44], were cultured in 20% inactivated bovine serum, 2% antibiotic/antimycotic at 5% CO_2, 37°C overnight before being subjected to transfection with the CRISPR materials.

Guide RNAs, ribonucleoprotein complexes

Here, we focused on GSH1, located on *S. mansoni* chromosome 3; 13231955-13233370 (Fig. 1d), an intergenic safe harbor site of 1,416 nt, the longest in length of the four GSH (Figs. 1 b–d). Guide RNAs (gRNA) for GSH1 were designed with the assistance of the CHOPCHOP[21,22,45] tools, using the annotated *S. mansoni* genome[14], to predict target sites, off-targets, and efficiency of CRISPR/Cas9 programmed cleavage. CHOPCHOP predicted three overlapping guide RNAs targeting three cleavage sites withinGSH1 with the three DSBs at six to 12 nt apart from each other. All three overlapping gRNAs lacked off-target sites and lacked self-complementarity. The three gRNAs were located on the forward strand of GSH1 at nucleotide positions 605-624, 617-636, and 623-642, respectively (Fig. 2a). We termed the predicted DSBs as target 1, target 2 and target 3. Synthetic guide RNAs (sgRNA), specifically Alt-R CRISPR-Cas9 sgRNA chemically modified to enhance functional stability, and recombinant *Streptococcus pyogenes* Cas9 nuclease, specifically Alt-R HiFi Cas9, which includes nuclear localization sequences (NLS), were sourced from Integrated DNA Technologies, Inc. (IDT, Coralville, IA). Each ribonucleoprotein complex (RNP) was complexed in the separate tube, with Cas9 and a single sgRNA at 1:1 ratio in 25 µl Opti-MEM; 10 µl of 1 µg/µl sgRNA (Opti-ME as diluent) was mixed with 10 µl of 1 µg/µl Cas9 (in Opti-MEM) by gentle pipetting and incubated for 10 min at 23°C to allow the RNP to assemble.

Doubled stranded DNA donor

A plasmid vector, pUC-Ubi-EGFP-ubi, was constructed by chemical synthesis of the donor transgene and its ligation into pUC (Azenta Life Sciences, Chelmsford, MA). The inserted sequence 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 of the *S. mansoni* ubiquitin promoter (2,056 bp), EGFP (717 bp), and the ubiquitin terminator (578 bp). Plasmid DNA was amplified 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 of 200, 400 or 600 bp in length (Table S2). Thus, three forms of the donor transgene DNA were prepared, with homology arms of increasing length – 200, 400 and 600 bp. The primers employed to amplify the transgene DNA from the pUC-Ubi-EGFP-ubi plasmid were 5×phosphorothioate-modified to enhance stability of the amplified DNA. (5´-modified long dsDNA donor (lsDNA) enhances HDR and favors efficient single-copy integration by its retention of a monomeric conformation[12].) (Fig. 3a). PCRs were carried out in reaction volumes of 50 µl in 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 of 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.

Amplificons 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 transgene stored at -20°C until used.

Transfection of schistosome eggs

Ten thousand eggs (LE) of *S. mansoni* were washed three times with chilled (4°C) 1×PBS before transfer into a chilled electroporation cuvette (4 mm electrode gap, BTX, Holliston, MA) with Opti-MEM as the electroporation buffer. Each 25 µl of RNP along with the lsDNA donor were immediately added into the cuvette with the schistosome eggs, to a total cuvette volume of 300 µl with Opti-mem: specifically, for the dual guide RNA/RNPs, group 1) 25 µl RNP1+25 µl RNP2, group 2) 25 µl RNP2+25 µl RNP3, and group 3) 25 µl RNP1+25 µl RNP3.. In groups with the lsDNA, 10 µg of this donor DNA was added into the cuvette before bringing the final volume to 300 µl/cuvette. Transfection of schistosome eggs with CRISPR materials was undertaken using square wave electroporation (Electro SquarePorator ECM 830, BTX), with a single 125 volt pulse for 20 ms[8,9,46]. The transfected eggs 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 (Biometer 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) 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.). Yields and purity were assessed quantified by spectrophotometry (NanoDrop One Spectrophotometer, Thermo Fisher Scientific), based on the ratios of absorbance at 260/280 and at 260/230 nm[47].

Analysis of CRISPR on-target efficiency

Amplicons of GSH1 spanning the programed DSBs were obtained using population genomic DNA (above) and primers termed 'indel-F and indel-R primers' that cover the region flanking expected double strand break of all the CRISPR target sites (Fig. 2a, Table S1). Amplification products were purified (NucleoSpin Gel and PCR Cleanup and gel extraction kit, Takara; cat no. 740609) 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[23] at default parameters with single and multiple CRISPR target analysis (Figs. 2b-2f, S1a). The pattern of indels in was confirmed by TIDE analysis of Sanger sequence reads [48] (Figs. S1b-e).

Detection of integration of the transgene into the schistosome genome

Integration of the donor transgene at GSH1 was analyzed by PCR, using GoTaq G2 DNA polymerase (Promega, Madison, WI) and two pairs of primers; one primer located on the GSH1

using specific primers upstream (5' KI-F) or downstream (3' KI-R) of the homology arms paired with primers specific for the transgene (5' KI-R or 3' KI-F) (Fig. 2b, Table S2), as described[9]. The PCR cycling conditions were 95°C, 2 min, 40 cycles 94°C, 15 sec, 58°C 30 sec, 72°C, 60 sec, and the amplification products were separated by electrophoresis and stained with ethidium bromide. The expected product sizes for the 5' and 3' integration site-specific amplicons were 728 and 983 bp, respectively, and the amplification control (amplified by indel-F and indel-R primers), expected product size, 764 bp (Fig. 2c).

Quantification of transgene expression in schistosome eggs

To examine the mRNA expression of EGFP, total RNAs extracted from LE were exposed to DNase to eliminate residual genomic DNAs and donor lsDNA donor were transcribed into cDNA using the 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) served as the reference gene. The specific primer for GAPDH-specific oligos: GAPDH-F; 5´-atgggacatttccaggcgag-3´, GAPDH-R; 5´-ccaacaacgaacatgggtgc-3´, expected amplicon of 285 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 examined, as above.

Quantification of fluorescence by spectral imaging and linear unmixing

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 a spectral META detector at 16 channels, scanning at 477 nm through 638 nm, simultaneously. A hurdle for the detection of EGFP using fluorescence microscopy is presented by the autofluorescence known to originate from the schistosome eggshell[49,50,51], with vitelline cells the likely origin of this emission[52]. To surmount this hurdle, the EGFP spectrum emitted from within each eggshell was investigated by selecting the area to be examined, specifically the entire miracidium, and collecting multispectral images of the miracidium within the eggshell using the LSM Image Examiner. The images collected were assessed for EGFP by subtracting regions emitting autofluorescence from the fluorescence signal collected for entire surface area of the egg: Specifically, the total EGFP intensity at 509 nm[52] was calculated using the Zeiss Zen (black edition) software module from ~400 eggs (~100 eggs from each of four biological replicates) in each of the control and experimental groups.

Databases. The sequence reads are available at GenBank BioProject XXX, SRA reads YYY to ZZZ.

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). This study is set within the framework of the « Laboratoire d'Excellence (LabEx) » TULIP (ANR-10-LABX-41) with the support of LabEx CeMEB, an ANR « Investments d'avenir » program (ANR- 10-LABX-04-01) and the Environmental Epigenomics Core Service at IHPE.

Author contributions

PB, WI, CG and CGG conceived the study. CC and CG identified gene safe harbors, MFM and TQ contributed promoter/ terminator cloning and analysis, RR optimized conditions for knock-in of donor transgenes, and PW and WB performed mutation analyses. WI planned and performed the programmed gene editing, knockout and transgene knock-in analysis. VM, AM, LM, SS and MM-K maintained the schistosome life cycle. MFM, AP and WI analyzed confocal micrographs and fluorescence. WI, CG and PB drafted the manuscript. All authors contributed to the writting and all approved the final version of the paper.

Competitive interests

The authors declare no competing interests.

Figure legends

Figure 1. Bioinformatic identification of sites of genomic safe harbors and their CRISPR programmed targets in the genome of *Schistosoma mansoni*

Panel **a.** Table shows criteria and rationale used to computationally predict genome safe harbors (GSHs). **b-d.** Chromosomal locations and lengths of four candidate GSH sites (blue boxes), located on chromosome 2 (GSH2 and GSH3) and chromosome 3 (GSH1, GSH4) of *S. mansoni*. In addition, coding sequences of genes identified in the vicinity of the four (gene-free) GSHs are shown, which were mostly uncharacterized proteins (red boxes with connected red line). Black and white scale bars at the top of each panel each represent one megabase pairs in length. **e.** Features of the GSHs including chromosomal location, length, programmed CRISPR/Cas9 targets, guide RNAs, and their predicted on- and off-target specificities from the CHOPCHOP design toolbox.

Figure 2. Efficiency of programmed knockout enhanced using three overlapping guide RNAs. Panel **a**. Schematic map of sites of the overlapping guide RNAs (blue, red and green arrows for target 1, -2 and -3, respectively) within GSH1 (yellow box), along with primer locations for indel analysis (purple arrows). The black arrows indicate the site programmed double stranded break (DSB) programmed by sgRNA1, 2 and 3. **b**. Efficiency of CRISPR knockout at GSH1 in eggs of *S. mansoni*, as assessed with the Deconvolution of Complex DNA Repair (DECODR) algorithm using distance, following transfection with dual and triple overlapping guide RNPs; RNP1+RNP2 (blue dots), RNP2+RNP3 (red), RNP1+RNP3 (green), and RNP1+RNP2+RNP3 (purple). Significantly higher CRISPR efficiency obtained with the three overlapping guide RNPs, mean = 23.6%, than the other groups ($P<0.001$). Among the groups transfected with dual RNPs, efficiency obtained with the RNP1+RNP3 treatment group, mean = 12.6%, was significantly higher than either of the other groups, RNP1+RNP2 at 5.4% and RNP2+RNP3 at 3.6% ($P<0.01$; one-way ANOVA with 95% confidence intervals, five or six independent biological replicates; GraphPad Prism). **c-e**. Representative alleles, in the schistosome genome, bearing indels at the target site in GSH1 (panel a) following transfection with dual guide RNPs, as a gauge of efficiency CRISPR-catalyzed gene editing. The reference wild type (WT) allele is shown above the mutant (KO) allele. KO alleles were identified during examination of nucleotide sequences for small deletions, 1-6 nt in length (- dash) or insertion/substitution (black boxes) The vertical black line boxes show targeted PAM (protospacer adjacent motif) trinucleotides. **f**. A representative example of an KO allele bearing a large sized deletion resulting from transfection with the triple guide RNPs.

Figure 3. Targeted transgene insertion and expression in the schistosome egg. Programmed CRISPR/Cas9 insertion (knock-in, KI) of a donor repair template of 4.451 kb in length, encoding an EGFP transgene driven by the *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 bp upstream of the position of sgRNA1 and the homology arm on the right (HAR) was sited 600 bp of downstream of the prototypic adjacent motif of sgRNA 3. **b.** Schematic illustration of the wild type (WT) and mutant (KI) alleles after

double stranded breaks programmed by the overlapping sgRNAs 1, 2 and 3 (scissors). PCR primer sites are indicated by the purple arrows. **c.** Targeted KI of the transgene detected by genomic PCR analysis using 5´KI (728 bp) or 3´KI (983 bp) specific pairs of primers. Negative controls for KI included WT, mock transfected, donor only treatment or KO (only RNPs transfection) groups, none of which had been transfected with RNPs. Lanes KI-1, KI-2 and KI-3 show amplicons from three independent biological replicates of programmed KI of the donor transgene. Other lanes show the outcome of RT-qPCRs from schistosome RNA with donor DNA electroporation (without CRISPR materials - nuclease or guide RNAs). The integrity of the genomic DNAs was confirmed by the presence of the expected amplicon of 784 bp in all lanes of the gel labeled "control PCR" using control primer (Fig. 3b).

Primer-dimer and/or non-specific PCR band(s) from DNA donor transfected-eggs were seen ≤ 100 bp in size. **d.** Expression of the EGFP transcript (expected amplicon size, 717 bp) as assessed by RT-qPCR following programmed KI of the transgene into GSH1. The integrity of the RNAs was assessed by analysis of transcripts of the reference gene, *Sm*GAPDH, with an expected amplicon size of 285 bp. Double-stranded DNA donor was used as the positive PCR template for EGFP. 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.

Figure 4. Markedly higher numbers of EGFP-positive eggs following programmed knock-in at GSH1 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 only with the lsDNA donor repair template; **a1, a2,** representative images from biological replicates. **b.** Eggs expressing EGFP in 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 KI. Eggs expressed EGFP observed from day 5 until day 15 when the experiment was terminated and taken the micrographs. **c.** Linear mixing analysis revealed EGFP fluorescence in miracidia within the eggshell in these representative micrographs; **c1**, plot showing the emission spectrum of EGFP (green curve) with a peak signal at 509 nm, and, also showing the autofluorescence background (purple curve). **d.** To assess expression of EGFP, eggs were counted (scored) as EGFP-positive when >20% of the surface area of the miracidium (within the eggshell) was EGFP positive; representative EGFP-positive eggs are shown in panels c3 and c4 (green eggs). On day 15 following transfection, eggs were scored and counted for EGFP positivity in each of four independent biological replicates (~100 eggs from each replicate) of KI (RNPs + lsDNA) and lsDNA only eggs. About 25% of the eggs in the lsDNA only control group were scored as EGFP-positive (range, 19 to 32%), whereas 75% of eggs the KI group expressed EGFP (range, 68 to 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. There was marked decrease of background autofluorescence in eggs from WT and mock-transfection groups after the linear unmixing normalization of the fluorescence spectrum.

Supporting information

Figure S1. a. CRISPR efficiency (percentage) on individual target sites as estimated using the Deconvolution of Complex DNA Repair (DECODR) algorithm using distance from dual

gRNAs; RNP1+RNP2 (blue dots), RNP2+RNP3 (green), RNP1+RNP3 (yellow) and triple gRNAs; RNP1+RNP2+RNP3 (pink dots), plotted from 5-6 independent biological replicates. The % indels on each target varied among RNP combinations:| transfection with triple RNPs lead to highest CRISPR efficiencies on each target site 1, 2 and 3, 26.5%, 14.2% and 31.2%, respectively, compared with the dual RNP groups. The % indels resulting by dual RNPs on target 1 were 6.7% and 13.8%; target 2, 3.9% and 1.7%, and target 3, 5.4% and 11.3%. **b-e,** examples of indel patterns resulting from the dual RNPs and triple RNPs. The graphs plotted by DECODR software confirmed the indel profiles similar to those obtained with TIDE (supplement figures 1b-1e).

Figure S2. Emission spectral intensity for eggs, scanned from 477 to 638 nm, 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).

Table S1. Primer list used in PCRs to amplify DNA donor, investigate EGFP transcript, programmed knockout and knockin.

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 Aznauryan, E. *et al.* Discovery and validation of human genomic safe harbor sites for gene and cell therapies. *Cell Rep Methods* **2**, 100154 (2022). https://doi.org:10.1016/j.crmeth.2021.100154

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

15 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

16 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

17 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

18 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

19 Augusto, R. C. *et al.* A simple ATAC-seq protocol for population epigenetics. *Wellcome Open Res* **5**, 121 (2020). https://doi.org:10.12688/wellcomeopenres.15552.2

20 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

21 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

22 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

23 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

24 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

25 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

26 Aoshima, T. *et al.* Modification of improved-genome editing via oviductal nucleic acids delivery (i-GONAD)-mediated knock-in in rats. *BMC Biotechnol* **21**, 63 (2021). https://doi.org:10.1186/s12896-021-00723-5

27 Dickinson, M. E., Bearman, G., Tille, S., Lansford, R. & Fraser, S. E. Multi-spectral imaging and linear unmixing add a whole new dimension to laser scanning fluorescence microscopy. *Biotechniques* **31**, 1272, 1274-1276, 1278 (2001). https://doi.org:10.2144/01316bt01

28 Lee, H., Kim, J., Kim, H. H., Kim, C. S. & Kim, J. Review on Optical Imaging Techniques for Multispectral Analysis of Nanomaterials. *Nanotheranostics* **6**, 50-61 (2022). https://doi.org:10.7150/ntno.63222

29 Newby, G. A. *et al.* Base editing of haematopoietic stem cells rescues sickle cell disease in mice. *Nature* **595**, 295-302 (2021). https://doi.org:10.1038/s41586-021-03609-w

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

31 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

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

33 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

34 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

35 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

36 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218 (2013). https://doi.org:10.1038/nmeth.2688

37 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

38 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

39 Lu, Z. & Berriman, M. Meta-analysis of RNA-seq studies reveals genes responsible for life stage-dominant functions in Schistosoma mansoni. *bioRxiv*, 308189 (2018). https://doi.org:10.1101/308189

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

41 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

42 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

43 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

44 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

45 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

46 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

47 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

48 Brinkman, E. K. *et al.* Easy quantification of template-directed CRISPR/Cas9 editing. *Nucleic Acids Res* **46**, e58 (2018). https://doi.org:10.1093/nar/gky164

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

50 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

51 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

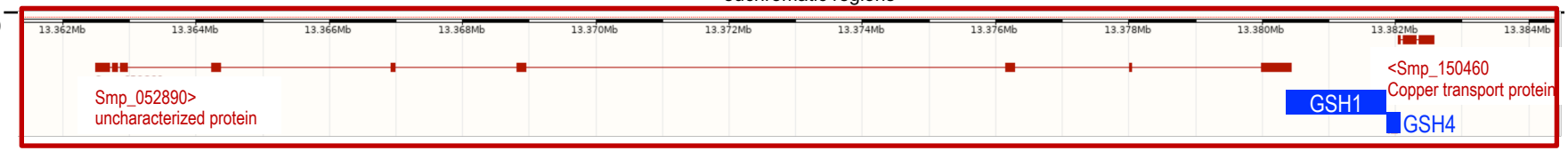
52 Wu, Q., Feng, Z. & Hu, W. Reduction of autofluorescence in whole adult worms of Schistosoma japonicum for immunofluorescence assay. *Parasit Vectors* **14**, 532 (2021). https://doi.org:10.1186/s13071-021-05027-3

Figure 1

a

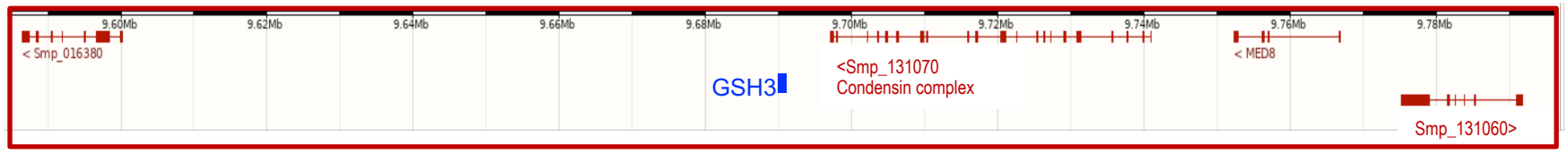
GSH criteria	Rationale
• Close to peaks of H3K4me3 in all developmental stages of the parasite	• a histone modification site that is associated with euchromatin and transcription start sites for highly expression of transgene
• Do not contain H3K27me3 in any developmental stage	• a histone modification that is associated with heterochromatin in any of the life stages to avoid silencing genes to be integrated
• Unique sequences	• Avoid CRISPR/Cas off target
• Euchromatin accessible to Tn5 integration	• Positive display of integration events
• Located outside annotated genes and long non-coding RNA, outside putative promotor regions where *more than 2 kb upstream* of the TTS and within 11 kb of HIV integration site	• Avoid gene disruption during cell differentiation • Avoid disrupting essential function of endogenous gene(s) • HIV integration sites known from *S. mansoni* genome to additionally support predictions of euchromatic regions

b



c

d



e

S. mansoni GSH genome, version 9	chromosome	location	length (nt)	CRISRP/Cas9 target sequence(s) predicted by CHOPCHOP
GSH1	3	13231955-13233370	1,416	Efficient CRISPR/Cas9, overlapping gRNAs without predicted off-targets: GGAACTCGGTAGATGAAGGC[TGG] ATGAAGGCTGGTTCAATAGC[TGG] GCTGGTTCAATAGCTGGAAT[TGG]
GSH2	2	13399775-13400744	967	One region only contains CRISPR/Caa9 target with PAM, CAGCATTACCTGTGACTTAC[TGG]G[TGG]
GSH3	2	7708870-7709621	752	Modest CRISPR/Cas9 efficiency (<35%) overlapping gRNAs without predicted off-targets : TGTAAAATGGCAACAGTTCT [TGG] ATGGCAACAGTTCTTGGTGT [TGG] ACAGTTCTTGGTGTTGGACC [TGG]
GSH4	3	13233423-13233560	138	Only one CRISPR/Cas9 target was predicted with very low CRISPR efficiency. Several CRISPR/Cas9 targets were predicted but these contain off targets on chromosome 1, 2, 3, 5 and ZW

Figure 2

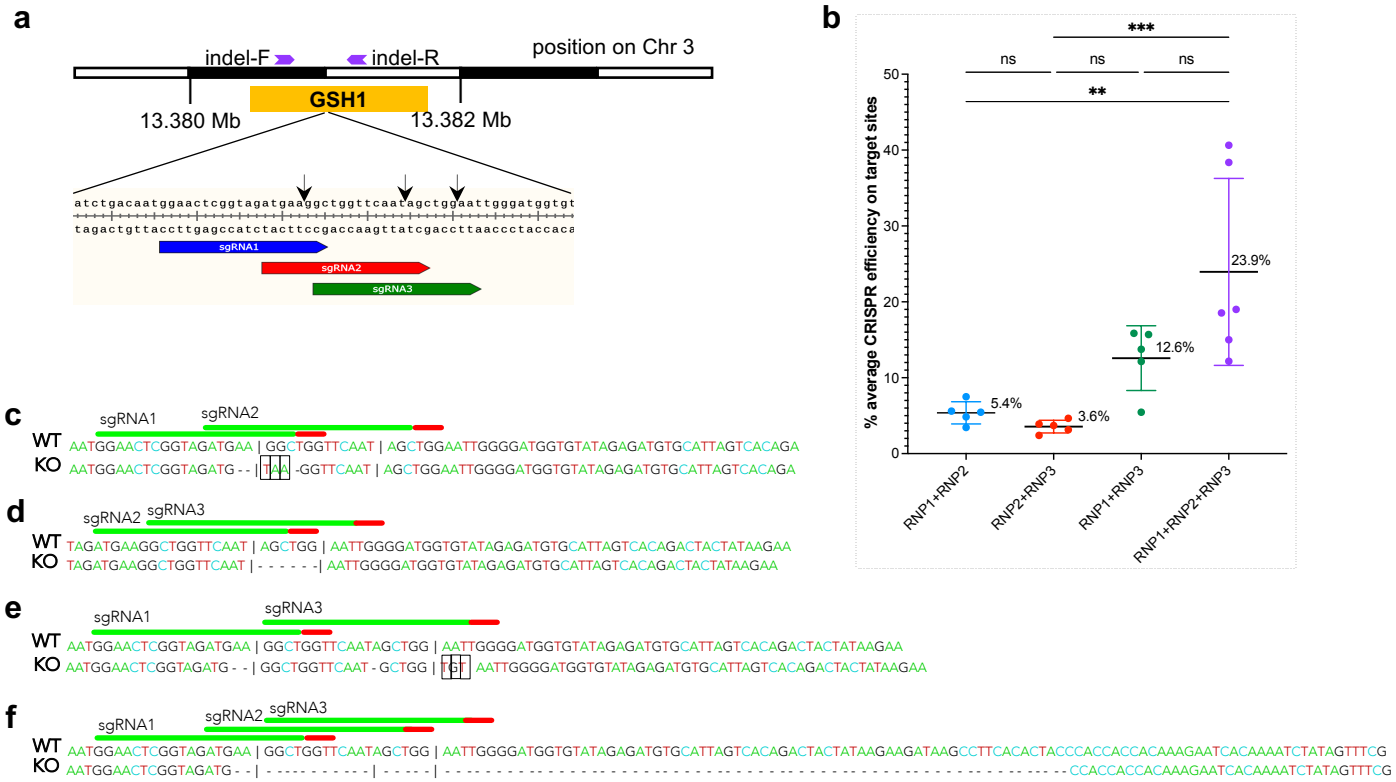

a
indel-F
indel-R
position on Chr 3
GSH1
13.380 Mb
13.382 Mb
sgRNA1
sgRNA2
sgRNA3
b
% average CRISPR efficiency on target sites
ns
ns
ns
***
**
5.4%
3.6%
12.6%
23.9%
RNP1+RNP2
RNP2+RNP3
RNP1+RNP3
RNP1+RNP2+RNP3
c
sgRNA1
sgRNA2
WT
KO
d
sgRNA2
sgRNA3
WT
KO
e
sgRNA1
sgRNA3
WT
KO
f
sgRNA1
sgRNA2
sgRNA3
WT
KO

Figure 3

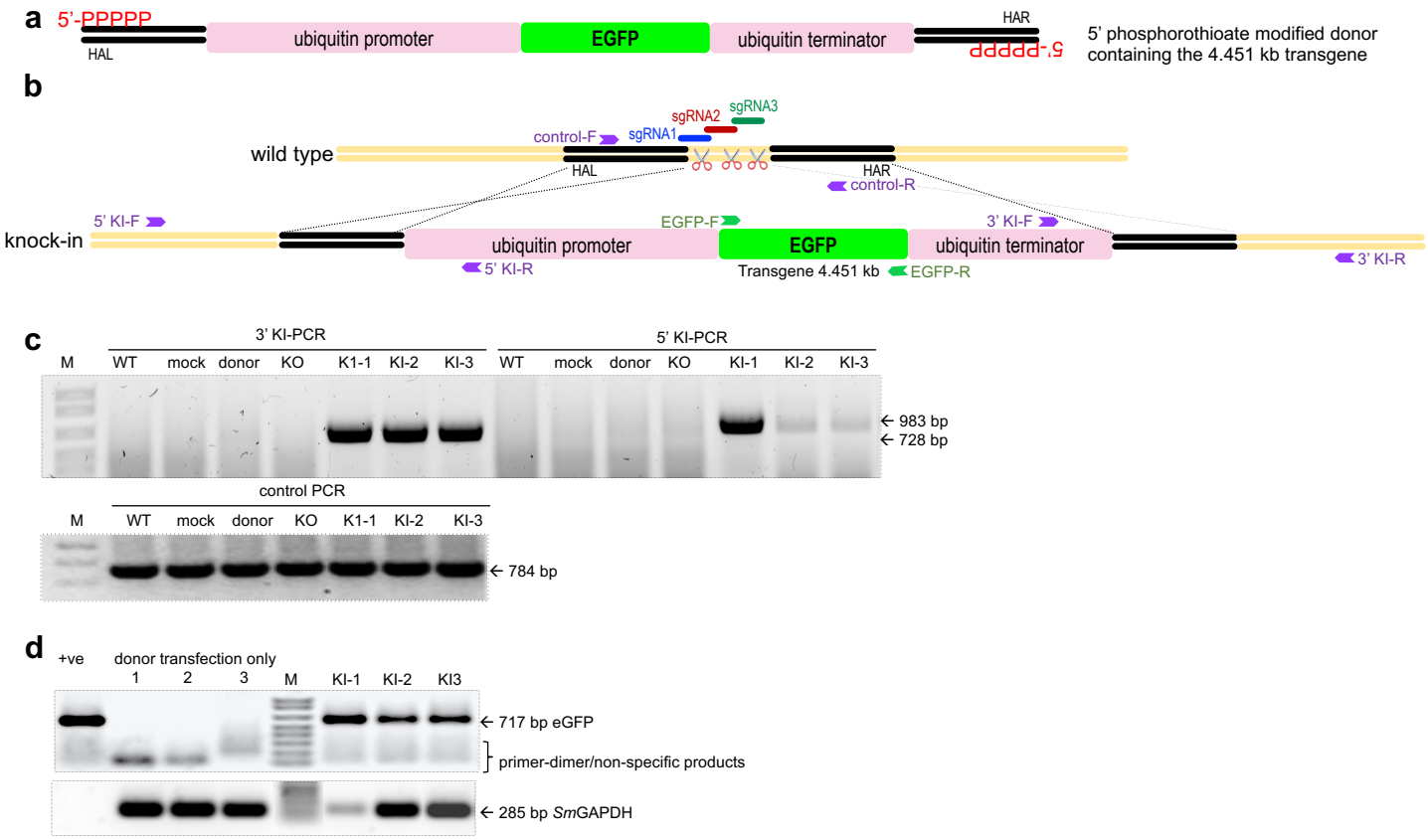
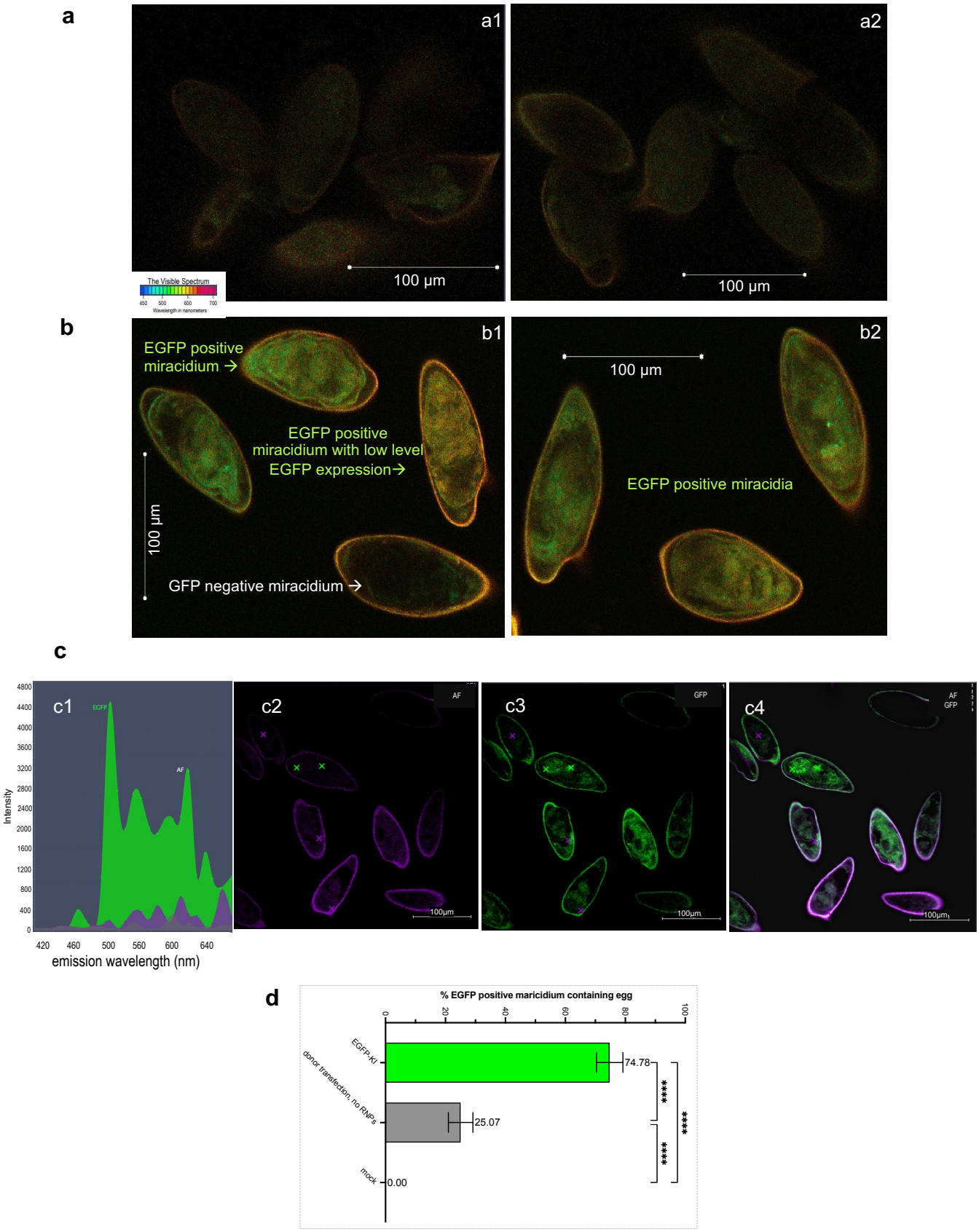
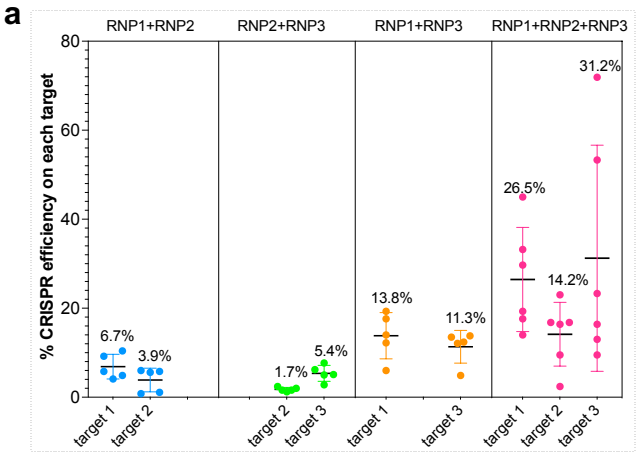
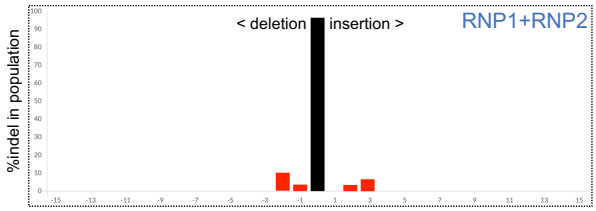

a
5'-PPPPP
HAL
ubiquitin promoter
EGFP
ubiquitin terminator
HAR
5'-PPPPP
5' phosphorothioate modified donor containing the 4.451 kb transgene
b
wild type
control-F
sgRNA1
sgRNA2
sgRNA3
HAL
HAR
control-R
knock-in
5' KI-F
EGFP-F
3' KI-F
ubiquitin promoter
EGFP
ubiquitin terminator
5' KI-R
Transgene 4.451 kb
EGFP-R
3' KI-R
c
3' KI-PCR
5' KI-PCR
M
WT
mock
donor
KO
K1-1
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/non-specific products
285 bp SmGAPDH

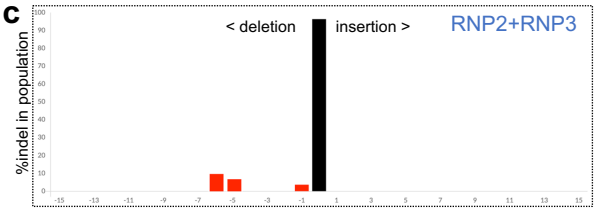
Figure 4

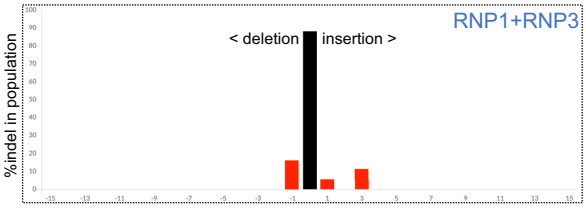

a
a1
a2
100 µm
100 µm
The Visible Spectrum
450 500 600 700
Wavelength in nanometers
b
b1
b2
EGFP positive miracidium →
EGFP positive miracidium with low level EGFP expression→
GFP negative miracidium →
100 µm
100 µm
EGFP positive miracidia
c
c1
c2
c3
c4
EGFP
AF
GFP
Intensity
emission wavelength (nm)
100µm
d
% EGFP positive maricidium containing egg
EGFP-KI
donor transfection, no RNPs
mock
74.78
25.07
0.00
****
****
****

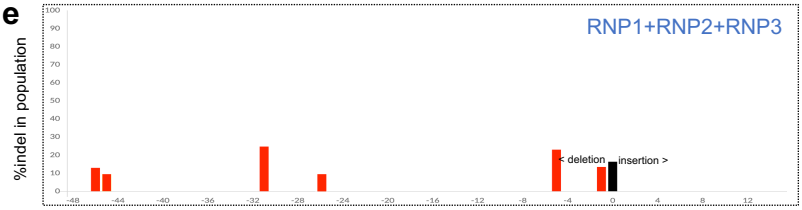
Supplement figure 1



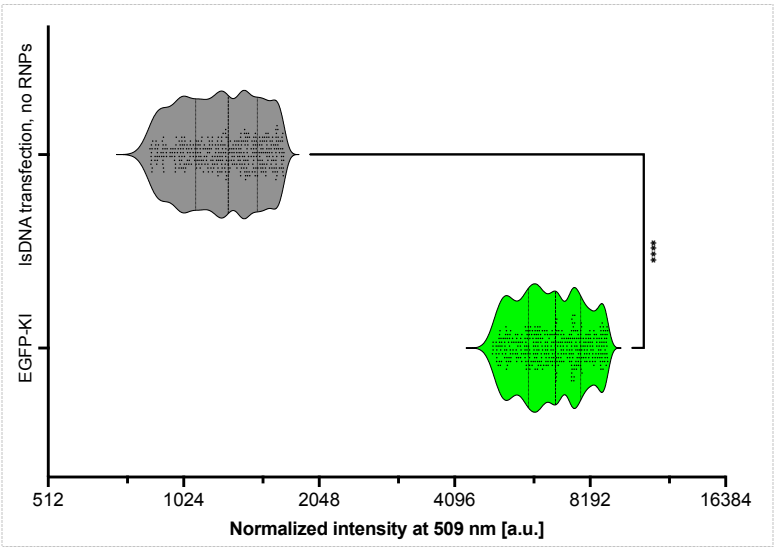








Supplement figure 2



Supplement Table 1. Primer list used in PCRs to amplify DNA donor, investigate EGFP transcript, programmed knockout and knockin

Primer	5’ to 3’ sequence
indel-F	TGTTATCGTCCGTCGCTTCA
indel-R	GCGTTCAAACATTGCCCACT
600 bp HA-F	AGATTGCTAGAAATTTATGAAAG
600 bp HA-R	ACTGCCGAATTTATAATATTTGG
400 bp HA-F	GATGAGTGCATCGATCGATTAC
400 bp HA-R	GAATGATGTTGGAAGTCAGA
200 bp HA-F	CGCTTCAATATTGTTTTTTGC
200 bp HA-R	CACACCTTCAAAACAATGTTTC
EGFP-F	ATGGTGAGCAAGGGCGAGG
EGFP-R	CTTGTACAGCTCGTCCATGCC
SmGAPDH-F	ATGGGACATTTCCAGGCGAG
SmGAPDH-R	CCAACAACGAACATGGGTGC
5’ KI-F	AGGGTTTTGGTTCGCAGGAT
5’ KI-R	CGGAGACAATCTGGAAAGGTCA
3’ KI-F	GTCCGCGTAATCGTCGTTACTA
3’ KI-R	GTGGTTCCATACTATGCAGTTTCC