Published online 15 June 2018
Nucleic Acids Research, 2018, Vol. 46, No. 18 ***e106***
doi: 10.1093/nar/gky517

Exploiting CRISPR–Cas9 technology to investigate individual histone modifications

Juan-José Vasquez[1,†], Carolin Wedel[1,†], Raul O. Cosentino[1,2,3] and T. Nicolai Siegel[1,2,3,*]

[1]Research Center for Infectious Diseases, University of Würzburg, 97080 Würzburg, Germany, [2]Department of Veterinary Sciences, Experimental Parasitology, Ludwig-Maximilians-Universität München, 80752 Munich, Germany and [3]Biomedical Center Munich, Department of Physiological Chemistry, Ludwig-Maximilians-Universität München, 82152 Planegg-Martinsried, Germany

Received October 16, 2017; Revised April 13, 2018; Editorial Decision May 22, 2018; Accepted May 23, 2018

ABSTRACT

Despite their importance for most DNA-templated processes, the function of individual histone modifications has remained largely unknown because *in vivo* mutational analyses are lacking. The reason for this is that histone genes are encoded by multigene families and that tools to simultaneously edit multiple genomic loci with high efficiency are only now becoming available. To overcome these challenges, we have taken advantage of the power of CRISPR–Cas9 for precise genome editing and of the fact that most DNA repair in the protozoan parasite *Trypanosoma brucei* occurs via homologous recombination. By establishing an episome-based CRISPR–Cas9 system for *T. brucei,* we have edited wild type cells without inserting selectable markers, inserted a GFP tag between an ORF and its 3′UTR, deleted both alleles of a gene in a single transfection, and performed precise editing of genes that exist in multicopy arrays, replacing histone H4K4 with H4R4 in the absence of detectable off-target effects. The newly established genome editing toolbox allows for the generation of precise mutants without needing to change other regions of the genome, opening up opportunities to study the role of individual histone modifications, catalytic sites of enzymes or the regulatory potential of UTRs in their endogenous environments.

INTRODUCTION

Post-translational modifications (PTMs) of histones, e.g. acetylation, methylation or phosphorylation, can serve, either individually or in combination with other PTMs, as binding sites for histone binding proteins. Histone binding proteins, in turn, can induce structural changes in chromatin leading to a multitude of biological processes affecting transcription, replication or DNA repair (1). To obtain insight into the roles of specific PTMs, their distributions have been mapped throughout diverse eukaryotic genomes, from trypanosomes to humans, revealing many conserved features, e.g. an enrichment of histone acetylation at transcription start sites or at sites of double-stranded breaks (DSB) (2–5). In addition, knowledge of the enzymes responsible for adding or removing specific PTMs is steadily increasing, linking PTMs to specific biological events. However, while depletion of one of these enzymes may lead to the loss of a particular histone modification and result in a specific phenotype, it may also impact other histone modifications or non-histone proteins complicating the interpretation of such studies. To date, for most PTMs it is not known whether mutation of a modified histone residue reproduces the phenotype caused by mutation of the modifying enzyme.

The most direct way to understand the role of specific PTMs is to replace the histone residue with one that mimics the modified or unmodified state. While mutational studies have been used to investigate catalytic sites, structural domains or binding sites in thousands of proteins, similar studies are largely missing for histones, with notable exceptions in *Saccharomyces cerevisiae* (6,7) and *Drosophila* (8). In yeast, the role of individual histone acetyl marks has been investigated by replacing the lysine of interest with an arginine or glutamine, to mimic deacetylated and constitutively acetylated lysine, respectively (7). The primary reason for the lack of mutational analyses of histones is that they are encoded by multigene families and that the tools required to simultaneously edit multiple genomic loci with high efficiency and without the need for selectable markers, have not been available. In yeast and *Drosophila,* these problems have been circumvented by the concurrent deletion of endogenous histone genes and expression of exogenous mutated histone genes from a plasmid (6,8). While these studies have yielded valuable information about the role of individual PTMs, a substantial advance in understanding their

*To whom correspondence should be addressed. Tel: +49 89 2180 77098; Fax: +49 89 2180 77093; Email: n.siegel@lmu.de
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.
Present address: Juan-José Vasquez, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

function will require many more such direct loss and gain of function analyses, spanning evolutionarily highly divergent organisms.

During the past few years, targeted nucleases have evolved into powerful tools that allow precise genome manipulation to be performed at high efficiency in organisms ranging from humans (9) to the protozoan parasite *Plasmodium falciparum* (10). Nuclease-induced double-stranded breaks (DSBs) are generally repaired by error-prone non-homologous end joining (NHEJ) or, in the presence of an exogenously introduced repair template, by the high-fidelity homology-directed repair (HDR) pathway. While NHEJ-mediated DNA repair typically leads to small insertions or deletions (INDELs) that can cause frameshift mutations and the inactivation of a particular gene, the HDR-pathway can be exploited to introduce specific genomic alterations. Numerous different targeted nucleases have been used for genome editing, however, the ease of use of the RNA-guided endonuclease Cas9 from the microbial adaptive immune system CRISPR (clustered regularly interspaced short palindromic repeats) makes it particularly suitable for precise genome manipulation. When complexed with CRISPR RNA (crRNA) and a *trans* activating RNA (tracrRNA), Cas9 introduces a DSB in a target sequence that is homologous to the crRNA. The target sequence is referred to as the protospacer sequence and is located next to the protospacer adjacent motif (PAM). To facilitate genome editing, crRNA and tracrRNA are fused and provided as a single guide RNA (sgRNA). CRISPR–Cas9 can thus be used to easily modify virtually any genomic locus as long as it is in close proximity to a PAM by specifying a 20-nt targeting sequence within its sgRNA (11). Given that Cas9 has been successfully simultaneously targeted to multiple genomic loci (12), it should be well suited to precisely edit multicopy genes such as those encoding histones.

In most organisms the NHEJ pathway is more efficient than the HDR pathway, which is an obstacle to the precise editing of multicopy genes. Thus, while Cas9 has been successfully used to disrupt multicopy genes via repair through NHEJ, it has not been used for the precise editing of large multicopy gene arrays. Many protozoa, including the medically relevant parasites *Trypanosoma brucei* and *Trypanosoma cruzi*, the causative agents of sleeping sickness and Chagas disease, have very inefficient NHEJ and microhomology-mediated pathways (13). For *T. brucei* it was observed that in the absence of a repair template the majority of DSBs will result in cell death rather than repair by NHEJ or microhomology (14). We have exploited the absence of efficient NHEJ and microhomology-mediated pathways in *T. brucei* to establish a strategy to allow for very precise genome editing.

The goal of this study was to establish CRISPR–Cas9 technology in *T. brucei* to allow the study of individual histone modifications. Despite its early divergence from other eukaryotes, many general patterns of histone PTMs are conserved in trypanosomes, e.g. just like in most other eukaryotes, the N-terminal tail of histone H4 contains several highly acetylated lysine residues (15).

Here, we describe the development of an episome-based Cas9 genome editing approach that allows marker-free genome editing at nucleotide resolution in *T. brucei*. First, to establish an efficient genome-editing protocol and to compare different transfection strategies, we added a C-terminal GFP tag to the *T. brucei* homologue of SCD6 (suppressor of clathrin deficiency 6, TbSCD6, Tb927.11.550) without changing its UTRs. Next, to determine the robustness of our approach and to exclude the possibility that only a small set of selected genes can be edited, we targeted the genes coding for the histone variants H3.V (Tb927.10.15350) and H4.V (Tb927.2.2670). For these genes, we were able to delete both alleles in one transfection. Finally, we evaluated the feasibility of using our episome-based Cas9 genome editing approach to precisely edit multicopy genes and succeeded in changing an individual codon in histone H4. Importantly, and for the first time in *T. brucei*, all genome edits were performed without the insertion of a resistance marker.

MATERIALS AND METHODS

Generation of the Cas9-episome and sgRNA-episome(SCD6-GFP)

Unless mentioned otherwise, all cloning reactions were performed using the In-Fusion HD Cloning Kit (Clontech) and DNA fragments were amplified using Phusion High-Fidelity DNA Polymerase (Thermo Scientific) according to the manufacturer's instructions. Oligos (Sigma-Adrich) and synthetic dsDNA fragments (Integrated DNA Technologies, IDT) are listed in Supplementary Table S1 and an outline of the cloning steps can be found in Supplementary Figure S1. The Cas9- and sgRNA-episomes (Figure 1A) are based on the pEV_Luc episome described previously (16).

To generate the Cas9-episome, the G418 resistance gene ($G418^R$) was removed from pEV_Luc by digestion with SacI and BamHI and replaced by a puromycin resistance gene (PUR^R). To reinsert sequence motifs required for RNA processing an In-Fusion reaction was performed with three fragments: the digested backbone, a DNA fragment to restore actin 3′-UTR and the PARP 5′-UTR (amplified with ocas_1 and ocas_2 from pEV_Luc), and the PUR^R (amplified with ocas_3 and ocas_4 from pyrFEKO-PUR, Addgene plasmid #24021), generating pEV_Luc_puro. Next, to replace *LUC* with *hSpCas9* the episome was digested with KpnI and SacI. To reinsert sequence motifs required for RNA processing an In-Fusion reaction was performed with four fragments: the digested plasmid backbone, a DNA fragment to the 5′-UTR of the actin gene (amplified with ocas_5 and ocas_6 from pEV_Luc_puro), Cas9 and a DNA fragment to restore the actin 3′-UTR (amplified with ocas_7 and ocas_8 from pEV_Luc_puro).

To add an HA-tag and an SV40 NLS to *hSpCas9* CDS, the Cas9 CDS was amplified from pX330 (Addgene plasmid #42230, (17)) using the oligo pair ocas_9/10 from pX330, and ligated into pRPa (18) after both fragments were digested with HindIII and BamHI, yielding pCW20. Next, the tagged Cas9 was excised from pCW20 by digestion with SbfI and AflII and used in the above-mentioned In-Fusion reaction to generate the Cas9-episome (Figure 1A).

To generate the sgRNA-episome, the *LUC* locus and a part of the G418 resistance gene were removed from the pEV_Luc episome by digestion with KpnI and NcoI. Next, the primer pair ocas_11/12 was used to amplify the PARP

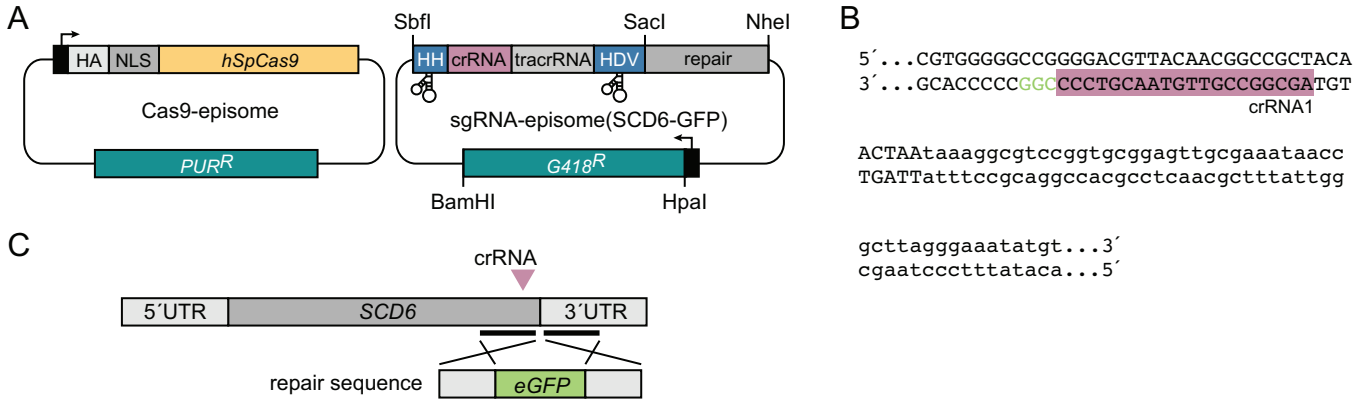

Figure 1. Strategy for episome-based Cas9 genome editing. (**A**) Key features of the Cas9- and the sgRNA-episome. Black arrow: PARP (procyclic acidic repetitive protein) promoter; HA: HA-epitope; NLS: SV40 nuclear localization signal; *hSpCas9*: human codon-optimized *Cas9*; PUR^R: puromycin N-acetyl transferase CDS conferring resistance to puromycin; $G418^R$: neomycin phosphotransferase conferring resistance to G418; HH: hammerhead ribozyme; crRNA: CRISPR-RNA; tracrRNA: *trans* activating RNA; HDV: hepatitis delta virus ribozyme; repair: sequence providing the DNA repair template carrying the desired mutations. (**B**) Sequence of the genomic target on Chr 11, 140,068-140,155. *SCD6* ORF: black, uppercase; *SCD6* 3′ UTR: black, lower case; protospacer: purple highlight; PAM: green. (**C**) Strategy for the marker-free tagging of the *SCD6* ORF. The purple triangle indicates the protospacer location targeted by the crRNA (not drawn to scale). The repair template encompasses the *eGFP* ORF flanked by 220 bp of the 3′-end of the *SCD6* ORF and 220 bp of its 3′-UTR.

5′-UTR and the previously removed fragment of the G418 CDS from pEV_Luc. The PCR product was inserted using the In-Fusion Kit generating pEV-noLuc. Subsequently, to permit insertion of a repair sequence, an SbfI restriction site was introduced downstream of the resistance marker gene by site-directed mutagenesis. To this end, two fragments were amplified from pEV_Luc using ocas_13/14 and ocas_15/16 and fused via PCR using ocas_13/16. Both the fusion PCR product and pEV-noLuc were digested with BamHI and NheI and joined by ligation to generate the sgRNA-episome. The repair sequence containing the 3′-end (220 bp) of the *SCD6* CDS (Tb927.11.550), the *eGFP* CDS, and the 3′-UTR (220 bp) of *SCD6* was synthesized in two fragments (gcas_1, gcas_2) and inserted into the sgRNA-episome after SbfI and NheI digestion. Following digestion of the sgRNA-episome with SbfI and SacI, the sgRNA flanked by ribozymes (gcas_3) was inserted to generate the sgRNA-episome(SCD6-GFP.1) (Figure 1A).

Construction of sgRNA-episomes to generate *ΔH4.V*, *ΔH3.V* and H4R4 cell lines

To generate the *ΔH3.V*, *ΔH4.V* and H4R4 cell lines, the sgRNA-episome(SCD6-GFP.1) was digested with SacI and NheI and the respective repair sequences (see below) were inserted. Next, the episomes were digested with SbfI and SacI and the corresponding sgRNA sequences (gcas_6, gcas_7, gcas_8, Supplementary Table S1) were inserted. Protospacer sequences were chosen to target Cas9 as closely as possible to the start codon of histone *H4.V*, *H3.V* or *H4*.

To disrupt *H3.V* the repair sequence was generated such that it included only the first 27 nt and the last 29 nt of the *H3.V* CDS. The upstream and downstream regions of the repair sequence were amplified from gDNA using the oligo pairs ocas_17/18 and ocas_19/20, respectively. To obtain the complete repair sequence, a fusion PCR of both fragments using the oligos ocas_21 and ocas_20 was performed. The PCR product was inserted into the backbone to yield the sgRNA-episome(ΔH3.V).

To delete *H4.V* the repair sequence was designed such that it did not include any nucleotide of the *H4.V* CDS and was amplified with the oligo pair ocas_22/23 from a synthetic DNA fragment (gcas_9, Supplementary Table S1). The PCR product was inserted into the backbone to yield the sgRNA-episome(ΔH4.V).

To generate the H4R4 cell line a repair sequence containing an arginine codon (CGC) instead of a lysine codon (AAG) in the fifth triplet position of the *H4* CDS was amplified with the oligo pair ocas_24/25 from a synthetic DNA fragment (gcas_10, Supplementary Table S1). In addition, the repair sequence contained a silent mutation (AGG to AAG) in the PAM sequence.

A step-by-step instruction on how to edit a specific gene can be found in the supplementary material.

Construction of sgRNA-episome(ΔG418^R)

The G418 resistance gene from sgRNA-episome(ΔH4.V) was removed by digestion with HpaI and BamHI and replaced by a hygromycin resistance gene, which was amplified with ocas_26 and ocas_27 from pLEW100v5_HYG (kind gift from George Cross, Addgene plasmid #24012). After digestion with SbfI and SacI a sgRNA sequence to target the neomycin resistance gene was inserted (gcas_11 and gcas_12, Supplementary Table S1).

Co-transfection of *T. brucei* Lister 427 procyclic form with episomes

Transfections of *T. brucei* Lister 427 procyclic forms were performed using a Nucleofector (Amaxa) as described previously (19). For the transfection 10 μg of both the Cas9-episome and sgRNA-episome(SCD6-GFP.1/2/3) were added to a cold cuvette (VWR 2 mm, 400 μl), followed by the addition of 2×10^7 cells suspended in 400 μl of transfection buffer (20). Immediately after transfection the cells

were suspended in 5 ml of SDM79 medium (21) (containing 20% fetal calf serum, FCS) and incubated overnight at 27°C. The next day, puromycin and G418 were added to the culture to a final concentration of 1 and 15 μg/ml, respectively. The drug selection was removed 5 days post transfection and fluorescence was determined by microscopy.

Sequential transfection of *T. brucei* Lister 427 procyclic form with episomes

Sequential transfections were performed following the protocol described for the co-transfection except for the following changes: following transfection the cells were suspended in 10 ml of SDM79 medium (containing 20% FCS) and split between two flasks. The next day, puromycin was added to the flasks to a final concentration of 1 μg/ml. The cells that survived the selection process were transfected with the sgRNA-episomes containing the sgRNA and repair sequences for the SCD6-GFP tagging, *H4.V* and *H3.V* deletion or H4R4 substitution following the procedure described above. The next day, G418 was added to a final concentration of 15 μg/ml. After 10 days the cultures were serially diluted using SDM79 (containing 20% FCS and drugs) to 1:1 × 10^1, 1:1 × 10^2, 1:1 × 10^3, 1:1 × 10^4, 1:1 × 10^5 and 1:1 × 10^6 and transferred to 96-well plates, 200 μl per well.

Monitoring of GFP-positive cells

Live microscopy was performed using an EVOS® FL Imaging System (Thermo Scientific) equipped with a Sony ICX285AQ color CCD camera and a LED light cube suitable for the detection of GFP. The cells were observed under both bright field and fluorescence using a 20× objective. In the serially diluted cultures each well was evaluated for the presence or absence of fluorescent cells. A well was considered to contain GFP-positive cells only if the majority of cells were fluorescent.

Evaluation of GFP-positive cells using flow cytometry

A MACSquant analyzer 10 (Miltenyi Biotec) equipped with a laser excitation of 488 nm coupled with 530/30 and 505LP sequential filter was used to evaluate the presence of GFP-positive cells. The photomultiplier voltage was set at 388 V and 100 000 events were measured for 7 samples transfected with the Cas9- and the sgRNA-episomes.

Fluorescence microscopy

Microscopy images were taken on a Leica DMI6000B microscope equipped with a Leica CRT6000 illumination system and a color camera Leica DFC630FX. The images were taken with a HCX PL APO oil immersion objective with 100 × 1.47 magnification or with a HCX PL FLUOTAR dry objective with 40 × 0.6 magnification. Linear image processing was done using Leica Application Suite Advance Fluorescence Software. The DAPI signal was detected using an excitation filter 359 nm and emission filter 461 nm (excitation filter BP 360/40 and suppression filter BP 470/40). The eGFP fluorescence signal was detected using an excitation filter 489 nm and an emission filter 508 nm (excitation filter BP 480/40 and suppression filter BP 527/30). Seven clones transfected with the Cas9- and the sgRNA-episomes were analyzed by counting ∼ 100 cells per clone, estimating the number of fluorescent cells over the total number of cells.

Southern blotting

Probes were designed to target the *SpCas9* CDS (nucleotides 1136–1715). β-tubulin was used as a positive control for genomic DNA (nucleotides 967–1088). The probes were synthesized by performing PCRs using the primers ocas_28 and ocas_29 for the *SpCas9* probe, and the primers ocas_30 and ocas_31 for the *β*-tubulin probe.

PCR reactions were performed using the Phusion High-Fidelity DNA Polymerase as described by the manufacturer (Thermo Scientific). The PCR products were run on a 1% agarose gel and the corresponding bands excised from the gel and cleaned up. The probes were labeled with [α-^{32} P]dATP using the DECAprime II Procedure following the manufacturer's instructions (DECAprime™ II Kit Ambion).

For the Southern blotting, genomic DNA was extracted using DNAzol® (Thermo Scientific) from cells sequentially transfected with the Cas9- and the sgRNA-episomes that were kept for 10 days in a flask before transfer to 96-well plates. The drug pressure was removed one month prior to harvest. 30 μg of each gDNA sample were digested with 200 U of KpnI overnight at 37°C while shaking at 80 rpm. The next day, 100 U of KpnI were added and the reaction was incubated for four additional hours to ensure complete digestion. The DNA was precipitated and resuspended in 50 μl of water. In parallel, 30 μg of the Cas9-episome were digested with 40 U of KpnI and precipitated. The digestion of this episome allowed us to have a hybridization control for the Cas9 probe and to perform a semi-quantitative Southern blot by loading different concentrations.

The digested gDNA of wild type cells and sequentially transfected parasites corresponding to 150 million cells were loaded onto a 0.7% agarose gel and run at 30 V overnight. Assuming one cell contained on average one of each episome (22), digested Cas9-episome equivalent to 150 × 10^4, 150 × 10^5 and 150 × 10^6 molecules were loaded to allow quantification of the episome integration frequency. The DNA transfer and hybridization steps were performed according to the Hybond-XL instruction manual (Amersham Biosciences) and (23). Briefly, the membranes were incubated for two hours at 68°C in 50 ml of 1× prehybridization solution (5× SSC, 5× Denhardt's solution, 0.5% (w/v) SDS, 10 μg/ml herring sperm) using glass cylinders in a hybridization oven while rotating. The probes were prepared by resuspending 12.5 μl of them in 100 μl of water, denaturing by heating for 10 min at 100°C and chilling immediately on ice for 2 min. The denatured radiolabeled probes were added directly to the prehybridization solution. The incubation was continued for 12–16 hours at 68°C. After hybridization the membranes were transferred as quickly as possible to another cylinder containing 50 ml of 2× SSC, 0.5% SDS and washed for 5 min at room temperature. After 5 min, the first rinse solution was poured off and 50 ml of 2× SSC, 0.1% SDS were added to the cylinders and the

membranes were incubated for 15 min at room temperature. The rinse solution was replaced with 50 ml of fresh 0.1× SSC, 0.1% SDS and the membranes were incubated for 2 h at 65°C. Then, they were briefly washed with 0.1× SSC at room temperature, dried on blotting paper, and wrapped in plastic foil. The membranes were finally introduced in a cassette with a phosphor imager screen for 1 h. The screen was scanned in a Thyphoon fla7000 (GE Healthcare Life Sciences).

Monitoring of genome editing by PCR and Sanger sequencing

To monitor the genome editing in SCD6-GFP, *ΔH3.V* and *ΔH4.V* cell lines, gDNA from multiple transfected clones and wild type PF cells was extracted from 1×10^6 cells using the Phusion Human Specimen Direct PCR kit (Finnzymes) according to the manufacturer's instruction. Primer sequences are listed in Supplementary Table S1. To detect SCD6-GFP positive cells the oligos ocas_32/33 were used. For the endogenous allele an amplicon size of 1466 bp and for the SCD6-GFP allele an amplicon size of 2186 bp is expected. To detect *ΔH3.V* cells the oligos ocas_34/35 were used. For the endogenous allele an amplicon size of 1572 bp and for the *ΔH3.V* allele an amplicon size of 1208 bp is expected. To detect *ΔH4.V* cells the oligos ocas_36/37 were used. For the endogenous allele an amplicon size of 1174 bp and for the *ΔH4.V* allele an amplicon size of 871 bp is expected. To determine the presence of H4R4 editing events, gDNA was extracted as described above and amplified using the oligos ocas_38/39 (Supplementary Figure S2). The smallest PCR product from each reaction was excised and analyzed by Sanger sequencing. Analyses were repeated at multiple time points.

Whole genome sequencing

The DNeasy Blood and Tissue Kit (Qiagen) was used to extract DNA from wild type and Cas9-edited cells. Six microgram of DNA was suspended in 100 μl of TE buffer (10 mM Tris–HCl, 1 mM EDTA, pH 8.0) and fragmented using a Bioruptor Plus (Diagenode) at the following settings: low-power, 70 cycles of 30 s 'on' followed by 30 s 'off'. Next, fragments between 100 and 300 bp were size-selected by gel electrophoresis and 40 ng were used to prepare indexed DNA libraries as described previously for DNA from ChIP assays (24). After purification, fragment sizes and library quality were analyzed in a Bioanalyzer 2100 (Agilent) and the DNA concentration was determined in a Qubit Fluorometer (Life Technologies). Finally, the libraries were sequenced using the NextSeq 500/550 v2 Kit and a NextSeq500 sequencer (Illumina).

Off-target analysis

To identify off-target cleavage by Cas9 we followed a previously published strategy (10,25). Paired-end reads from wild type, SCD6-GFP and H4R4 (clone B3-3-1) cells were mapped to the *T. brucei* Lister 427 genome version 26, downloaded from TriTrypDB database, using bwa (https://arxiv.org/abs/1303.3997), version 0.7.12-r1039. Sorted bam files were obtained using samtools ((26), version 1.5). VarScan (v2.4.3) software (27) was used to detect INDELs and extract specific INDELs from the edited cell lines by comparing them to the wild type cell line. The command mpileup2indel was used to call INDELs using default options, with the exception of the *P*-value threshold for calling variants, which was set to 0.05 instead of 0.01. Then, the specific INDELs occurring in the edited cell lines were extracted using the 'compare' command of VarScan.

The Protospacer Workbench software suite (28) was used to find putative off-targets from the original seeds, 'AGCG GCCGTTGTAACGTCCC' corresponding to the sgRNA1 targeting *SCD6*, and 'GAAGGGTAAGAAGAGTGGTG' corresponding to the sgRNA1 targeting *H4*. The off-target candidates where identified using the third-party software RazersS3 (29) allowing up to five mismatches, and ranked by the score given by the algorithm developed previously (11). For all off-target candidates, the vicinity (within a 21-nt window) to specific INDELs of the edited cell lines was assessed. The ranked lists of off-targets for the assessed sgRNAs are available in the Supplementary Table S2. In wild-type, SCD6-GFP and H4R4 cells we identified 6216, 4342, 9254 ($P < 0.05$) INDELs, respectively (Supplementary Table S3). The higher amount of INDELs in the H4R4 cells compared to wild type cells may be partially attributed to the sequencing depth. Wild type cells were sequenced to 21.3x coverage, whereas H4R4 cells were sequenced to 29.5x coverage. As we sequence deeper, more INDELs become significant.

Estimation of histone H4 gene array length and number of episomes per cells

Sequence reads from wild type, SCD6-GFP and histone H4R4 (clone B3-3-1) cells mapping to histone *H4* genes in the *T. brucei* Lister 427 genome version 26 were counted with bedtools (30), v2.26.0 using the 'multicov' command, normalized to counts per kilobase and added up. Counts per kilobase were also calculated for four other single-copy genes: (i) *H2A.Z* (Tb427.07.6360), (ii) *RPB9* (Tb427tmp.02.5180), (iii) *PPL1* (Tb427.05.4070) and iv) *XNRE* (Tb427.05.3850), and the average was used as the expected counts per kilobase for a single-copy gene. The ratio between the added-up counts per kilobase of reads mapping the *H4* genes and the average counts per kilobase of the selected single-copy genes was used as an estimation of the histone *H4* gene array length.

The same approach was used to calculate the number of episomal copies per cell except that the reads were mapped to a hybrid genome composed of the *T. brucei* Lister 427 genome version 26 and the Cas9-episome, and the reads mapping to *Cas9* were counted instead of those mapping to histone *H4*.

Analysis of episome integration

To determine if episomes had integrated into the genome, we mapped the paired-end DNA-reads from wild type, SCD6-GFP and H4R4 cells to the episome-Cas9 and the *T. brucei* genome, as described above. Next, we inspected read-pairs of which one read mapped to the episome and

one to the genome. The number of reads fulfilling these criteria was extremely small for all three cell lines and no evidence was detected suggesting episome integration in the cells transfected with the Cas9-episome (data not shown).

RESULTS

Establishment of an episome-based CRISPR–Cas9 system in *T. brucei*

To enable genome manipulations in wild type cells, Cas9 and sgRNAs were expressed from episomes. Episome-based expression of Cas9 has been reported for other protozoan parasites (10,25,31). Given the large size (>10 kb) of available episomes (22), we established a two-episome system: one episome expressing human codon-optimized *Streptococcus pyogenes* Cas9 containing an HA-epitope tag and an N-terminal SV40 nuclear localization signal and a second episome coding for the desired sgRNA and carrying a donor DNA repair template to ensure efficient and precise DNA repair via the HDR pathway (Figure 1A).

In most CRISPR–Cas9 applications, including studies published for *Plasmodium falciparum* (10) and *Leishmania major* (25), RNA polymerase III promoters (e.g. U6 promoters) have been used to transcribe the sgRNA. The U6 promoter structure found in most organisms allows the transcription of sgRNAs with precise 5′- and 3′-ends, avoiding the incorporation of additional nucleotides that could interfere with normal sgRNA function. However, the *T. brucei* U6 promoter contains an essential element inside the U6 gene, making it unsuitable for the transcription of sgRNAs (32). In addition to U6 promoters, T7 promoters have been used to drive *in vivo* sgRNA transcription (33,34) and are widely used to overexpress genes in *T. brucei* (35). However, reliance on T7 polymerase restricts the use of Cas9-based editing to transgenic cell lines expressing T7 polymerase. Thus, to edit wild type trypanosomes, we employed the endogenous PARP (procyclic acidic repetitive protein) promoter to transcribe the sgRNA. To ensure its correct length, the sgRNA was flanked by a hammerhead ribozyme at its 5′-end and a hepatitis delta virus ribozyme at the 3′-end (36).

To facilitate the establishment of an efficient Cas9-based genome editing protocol in *T. brucei,* we tagged the *T. brucei* homologue of SCD6 with eGFP. Expression of a GFP tag provided a rapid means to assess the efficiency of different transfection strategies; furthermore, SCD6 is highly expressed in *T. brucei* (37) and tolerates a fluorescent protein tag at its 3′-terminus (38).

To insert the *eGFP* ORF between the *SCD6* ORF and its 3′-UTR, a sgRNA was designed to guide Cas9 to introduce a DSB at the 3′-end of the *SCD6* ORF (Figure 1B). Downstream of the sgRNA, the episome contained a repair sequence encompassing the *eGFP* ORF (Figure 1C).

Cas9 allows genome editing at high efficiency

To establish transfection conditions, we first compared the co-transfection of the 'Cas9-episome' and the 'sgRNA-episome(SCD6-GFP.1)' to the sequential transfection of both episomes. While co-transfections yielded several green fluorescent cells (Figure 2A), sequential transfections resulted in a much higher percentage of GFP-positive cells (Figure 2B). No green fluorescent cells were obtained after transfection of a control episome lacking a sgRNA sequence (Figure 3A). As reported previously, the episomes were stably maintained for more than 2 months (22), even after multiple cycles of freezing and thawing. Thus, cells carrying the Cas9-episome can be used as a starting cell line for different genetic manipulations.

Next, we optimized the sequential transfection approach. Cells carrying the Cas9-episome were transfected with the sgRNA-episome(SCD6-GFP.1), diluted and transferred to 96-well plates either immediately or 10 days post transfection (Figure 2B). Immediate dilution and transfer has the advantage that edited cells with a growth phenotype will not be outgrown by wild type cells in which editing did not occur. However, if the editing takes place after dilution and transfer, this will result in a heterogenous population of cells. Heterogeneity may occur at the population level, with editing occurring in some cells and not others, or at the genome level, with some cells carrying homozygous and others carrying heterozygous genomic changes. Thus, late dilution and transfer should increase the homogeneity of a particular clone.

To assess the effect of immediate versus delayed dilution and transfer on homogeneity, we determined the number of GFP-positive and GFP-negative clones for cells that were immediately diluted and transferred and those that were diluted and transferred after 10 days. Delayed dilution and transfer resulted in ∼90% of clones being edited compared to ∼10% after immediate dilution and transfer (Table 1).

To further characterize the SCD6-GFP.1 clones (diluted and transferred after 10 days), we employed fluorescence microscopy, FACS, PCR and Sanger sequencing. Both FACS and imaging (Figure 3A-B and Supplementary Figure S3) of cells from GFP-positive wells indicated that all clones initially classified as GFP-positive contained more than 90% of GFP-positive cells. Thus, Cas9 can be used to perform marker-free genome editing in *T. brucei* with high efficiency.

To assess whether both or just one allele contained a GFP tag, gDNA from several clones was analyzed by PCR. We found a mixture of both outcomes; some clones were homozygous for the SCD6-GFP gene while others were heterozygous (Figure 3C). To understand why only one allele was edited in the heterozygous clones, we sequenced the smaller of the two PCR products, assuming it represented the wild type allele, and found mutations in the PAM or protospacer sequence (Figure 3D). YFP tagging of SCD6 has been reported to lead to a growth defect (37). Thus, while tagging of both alleles in one transfection is possible, our data suggest that if editing events lead to a growth defect, there may be selection for compensatory mutations, such as mutations in the PAM or protospacer sequence that prevent editing.

Cas9-mediated editing leads to negligible off-target effects

The specificity of Cas9 is determined by the 20-nt sequence of the crRNA and the PAM (NGG for SpCas9). However, although the length of the target sequence should be sufficient to ensure targeted cleavage even in large genomes

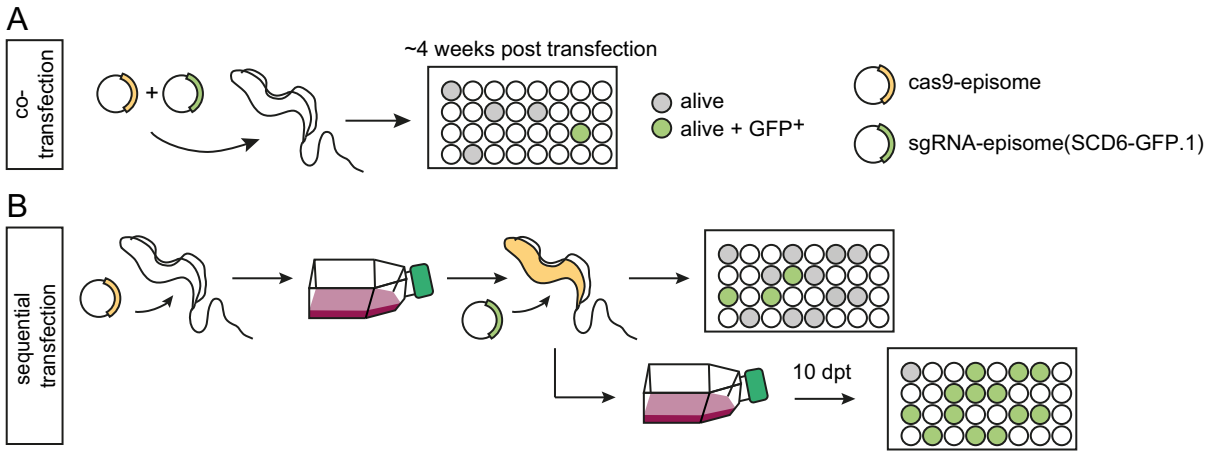

Figure 2. Outline of transfection strategies. (**A**) Co-transfection strategy. (**B**) Sequential transfection strategy.

(∼3Gb), numerous studies have reported that Cas9 can induce off-target mutations, even at sequences that differ by 5 nt from the targeted site (11,39,40). Given the absence of an NHEJ pathway in *T. brucei* (13) and findings using the meganuclease I-*SceI* in *T. brucei* (14), we expect that cleavage at off-target sites, for which no repair template is provided, will not be repaired efficiently and will result in cell death rather than in unintended editing events. Nevertheless, to assess the degree of off-target mutations in *T. brucei*, we sequenced gDNA from wild type cells and a population of cells edited using the sgRNA-episome(SCD6-GFP.1) to 21.3x and 22.9x coverage, respectively. Typically, off-target editing results in small INDELs. Thus, we searched the genome of the edited cells for the presence of INDELs in the vicinity of potential off-target sites that differ by ≤5 nt from the target site as described previously (10,25). Potential off-target sites were identified using the software RazersS3 (29), which is part of the Protospacer Workbench (28), for details see methods. However, no INDELs were identified at potential off-target sites, indicating that off-target editing events are negligible in *T. brucei* (Supplementary Table S2). DNA sequencing also confirmed original reports of ∼1 episome per cell (Supplementary Figure S4).

Cas9 allows multiple genome edits in parallel or in sequence

To increase the versatility of Cas9-mediated editing, we tested strategies to edit multiple genomic loci, e.g. to allow deletion of multiple genes.

The episome used in this study has been reported to be highly stable and not to integrate into the genome (22). Our study confirmed these observations. We found cells to remain resistant to G418 even when they had been cultured for 2 months without drug pressure, suggesting that the episome had been efficiently maintained. To evaluate the frequency of episome integration, we performed Southern blotting using DNA from a population of cells edited with the sgRNA-episome(SCD6-GFP.1) and found no evidence for integration (Supplementary Figure S4A). Analysis of gDNA-seq data from the same population suggested the presence of ∼1.1–1.5 episomes per cell (Supplementary Figure S4B) as published previously (22) and yielded no indication that episome integration had taken place. Next, to investigate whether episomes could be actively removed, for example once editing had occurred, we decided to generate a second sgRNA-episome expressing a sgRNA recruiting Cas9 to cut the first sgRNA-episome at the neomycin resistance gene. In the absence of a repair template the cut should not be repaired and the episome should be lost (Figure 4A).

To test this approach, we first generated sgRNA-episomes to permit deletion of the histone variants *H3.V* or *H4.V*. Deletion of *H3.V* and *H4.V* occurred in both alleles in 100% and 66% of the clones, respectively (Figure 4B). Sanger sequencing of PCR products from three *ΔH3.V* and three *ΔH4.V* clones confirmed that the edited regions exactly matched the provided DNA repair template (Figure 4C). Thus, for genes that can be deleted without inducing a growth defect, both alleles were efficiently edited in a single transfection.

In a second step, to perform additional genome edits, we replaced the resistance marker $G418^R$ of the sgRNA-episome with a hygromycin resistance gene (Hyg^R). In addition, we inserted a sgRNA targeting the $G418^R$ gene present in the sgRNA-episome used to delete *H3.V* during the first round of genome editing. As expected because the newly constructed sgRNA-episome lacked a template sequence to facilitate repair of the $G418^R$ gene (Figure 4A), cells transfected with sgRNA($\Delta G418^R$) regained sensitivity to G418 treatment, indicating a loss of the first sgRNA-episome carrying the $G418^R$ gene. Thus, the efficient removal of episomes can allow multiple editing events.

Cas9 can be used to change single residues in multicopy histone arrays

The goal of this study was to determine whether CRISPR–Cas9 technology could be used to study the role of individual histone residues. As a proof of principle, we replaced the lysine at position 4 on histone H4 (H4K4) with an arginine (H4R4) to mimic the constitutively non-acetylated state (Figure 5A). In wild type *T. brucei* H4K4 is acetylated in 80% of the histones (15) and in cells lacking the histone acetyltransferase 3 (*ΔHAT3* cells) H4K4ac levels drop to

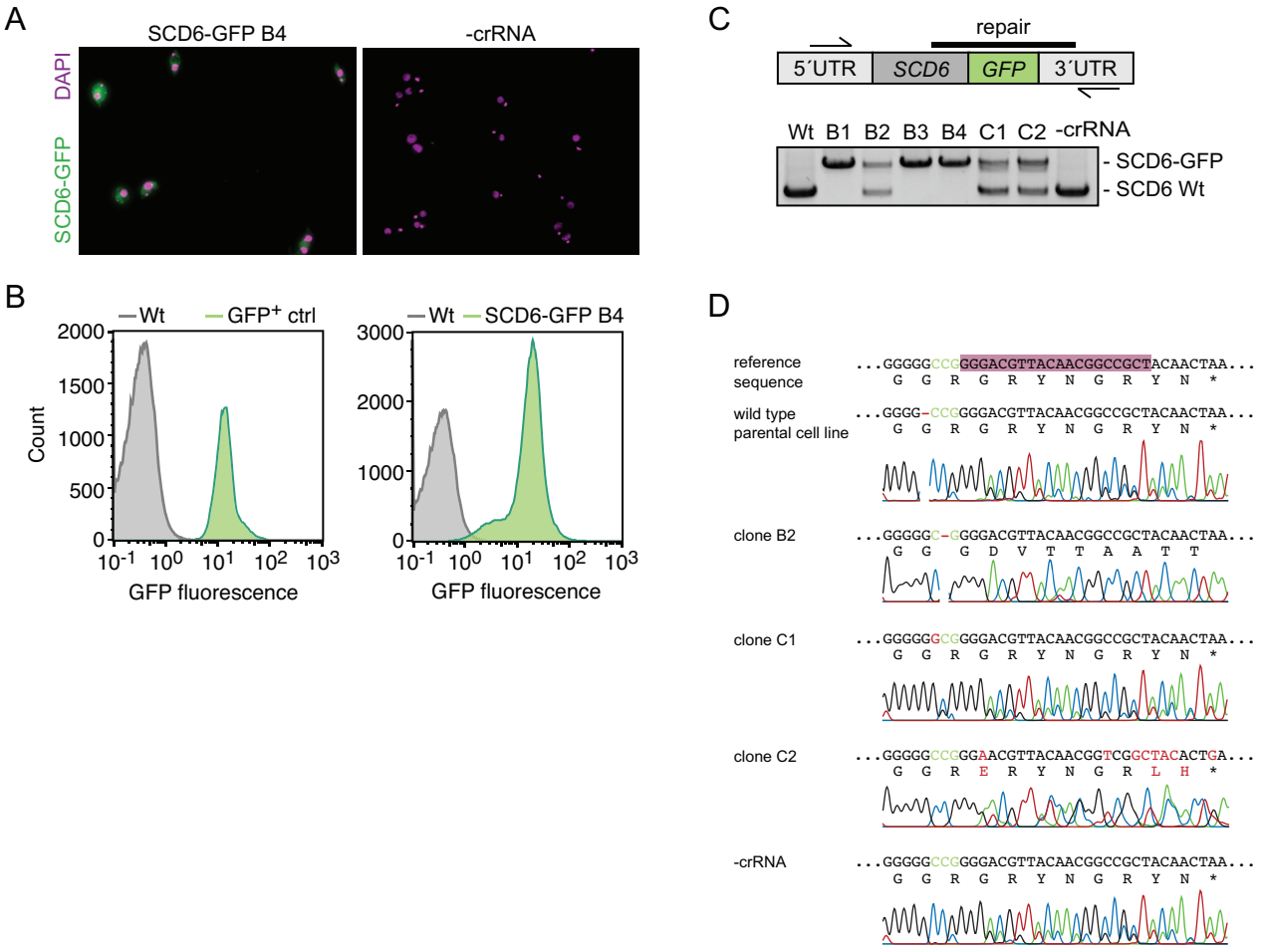

Figure 3. Cas9 allows marker-free genome editing at high efficiency. Unless indicated otherwise, cells were sequentially transfected with the Cas9-episome and the sgRNA-episome(SCD6-GFP.1) and diluted 10 days after transfection. Cells were only analyzed if <20 wells per 96-well plate contained living cells. (**A**) Fluorescence microscopy analysis of cells transfected with sgRNA-episome(SCD6-GFP.1, clone B4) or a sgRNA-episome lacking a sgRNA (-crRNA). (**B**) Flow cytometry analysis of edited cells. Left panel: wild type cells (gray) and cells in which SCD6 was GFP-tagged using the pMOT system (55) (green). Right panel: wild type cells (gray) and a representative clone of cells transfected with the sgRNA-episome(SCD6-GFP.1, clone B4). (**C**) PCR-based analysis of wild type cells, six clones of cells transfected with sgRNA-episome(SCD6-GFP.1) and cells transfected with an sgRNA-episome lacking a sgRNA (-crRNA). Top panel: outline indicating primer binding sites (half arrows). Bottom panel: agarose gel revealing the presence of homozygous (one PCR product) or heterozygous (two PCR products) tagging events. (**D**) Sanger-sequencing results of the *SCD6* gene from wild type cells, the non-edited *SCD6* allele of three heterozygous clones and cells transfected with a sgRNA-episome lacking a sgRNA (-crRNA).

5–10% with no growth defect (41,42). Based on the analysis of deep sequencing data of gDNA, we estimated the number of histone H4 genes to be ∼43 (21.5 per homologous chromosome).

Since we suspected that editing is a dynamic process that proceeds over weeks, we designed one sgRNA to target the multicopy histone H4 array (Figure 5B) and monitored the replacement of H4K4 with H4R4 over a period of 5 months (Figure 5C).

The first copies of histone H4R4 were detected after 3 months in one out of three analyzed clones (Figure 5C). To obtain clonal populations, cells were diluted again and transferred to a new 96-well plate. Analysis of the subclones revealed a strong increase in the percentage of H4R4 (Figure 5C). Deep sequencing of the gDNA from the Cas9-edited cells indicated that ∼90% of H4K4 was replaced with H4R4 (Figure 5C). At 29x coverage genome sequencing, no INDELs were detected in the vicinity of putative off-target sites (Supplementary Table S2) and we found the number of *H4* copies to drop from ∼43 to ∼26 (Figure 5D). Thus, in the absence of an efficient NHEJ pathway Cas9 technology can be used for the precise editing of histone gene arrays.

DISCUSSION

The availability of versatile and reliable genetic tools is one of the key reasons why *T. brucei* has become one of the most important model protozoa. For a long time, it has been relatively easy to generate knockout *T. brucei* parasites, to overexpress specific genes and to exploit the endogenous RNAi pathway to deplete transcripts.

The goal of this study was to further extend the available toolbox and to exploit the absence of efficient NHEJ to develop a strategy to study individual histone modifications. The approach presented here enables, for the first time, genome editing in *T. brucei* without the need to insert a selectable marker, editing of individual nucleotides and of

Table 1. Comparison of immediate versus delayed dilution and transfer

Dilution	Immediate dilution and transfer			Delayed dilution and transfer (10 dpt)		
	Wells with living parasites	Wells with GFP-positive parasites*	% of wells with GFP-positive parasites	Wells with living parasites	Wells with GFP-positive parasites*	% of wells with GFP-positive parasites
1:10	41	5	12	96	ND	ND
1:100	7	0	0	96	ND	ND
1:1000	0	0	0	84	75	89
1:10 000	0	0	0	51	45	88
1:100 000	0	0	0	18	17	84
1:1 000 000	0	0	0	19	17	89

*Wells were classified as GFP-positive if the majority of cells appeared to express GFP. Analysis was performed using an EVOS FL imaging system (Invitrogen). Note, the numbers listed in this table are derived from the original protocol optimization. The numbers of wells with living parasites obtained at the different dilutions varied strongly among transfections.

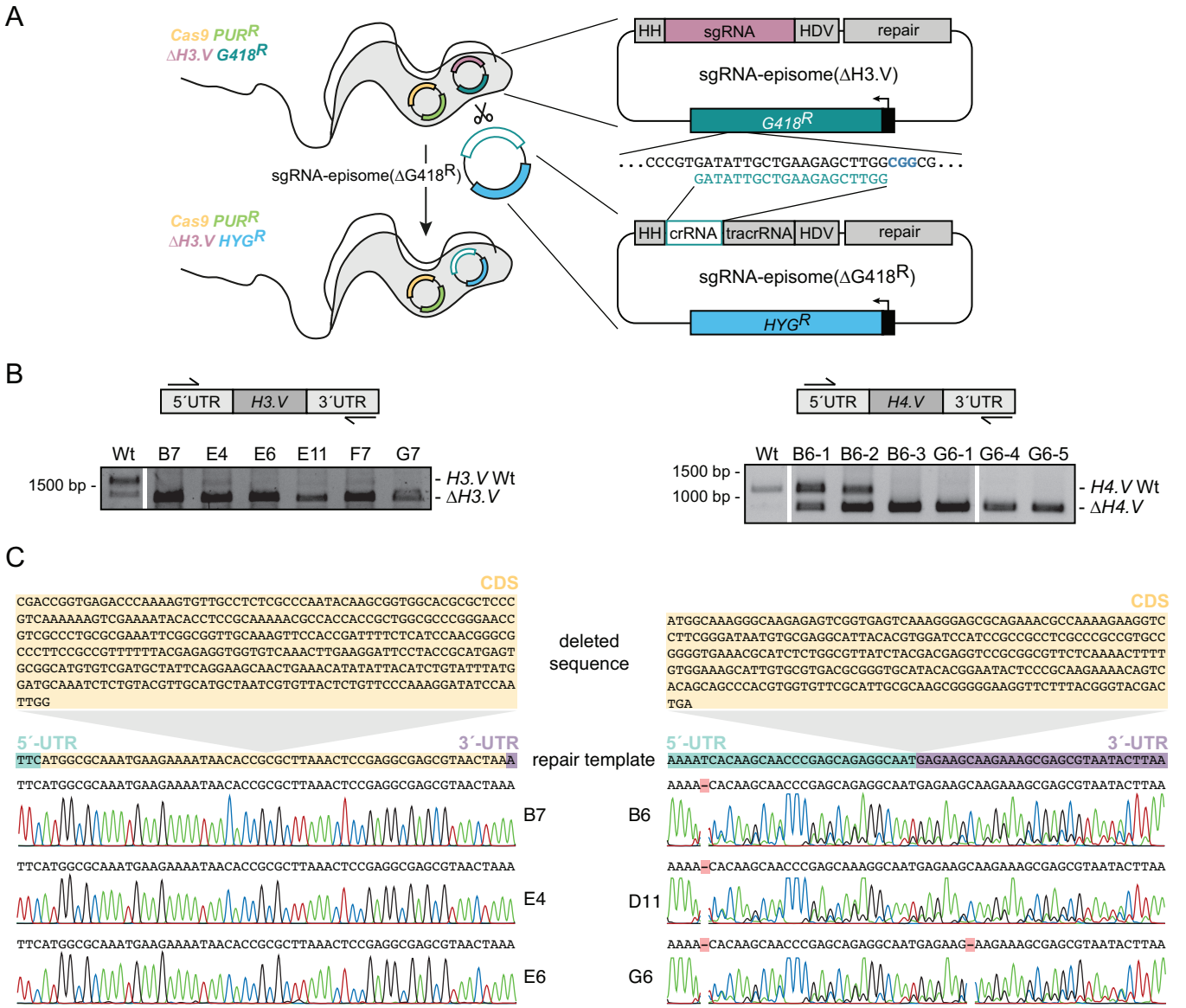

Figure 4. Cas9 allows multiple genome edits. (**A**) Strategy used for the removal of the sgRNA-episome. (**B**) PCR-based analysis of cells sequentially transfected with the Cas9-episome and sgRNA-episome(ΔH3.V) or sgRNA-episome(ΔH4.V). Top panel: outline indicating primer binding sites (half arrows). Bottom panel: agarose gel revealing presence or absence of *H3.V* (left) or *H4.V* (right). Homozygous deletions of *H4.V* were only observed after the culture had been diluted a second time. (**C**) Sanger sequencing demonstrated that all editing events exactly matched the repair template. The deleted CDSs of *H3.V* and *H4.V* are represented on top and highlighted in orange. The repair templates are shown in the middle (5′-UTR in green, 5′-end of the *H3.V* CDS in orange and 3′-UTR in purple. The sequencing chromatograms are depicted at the bottom.

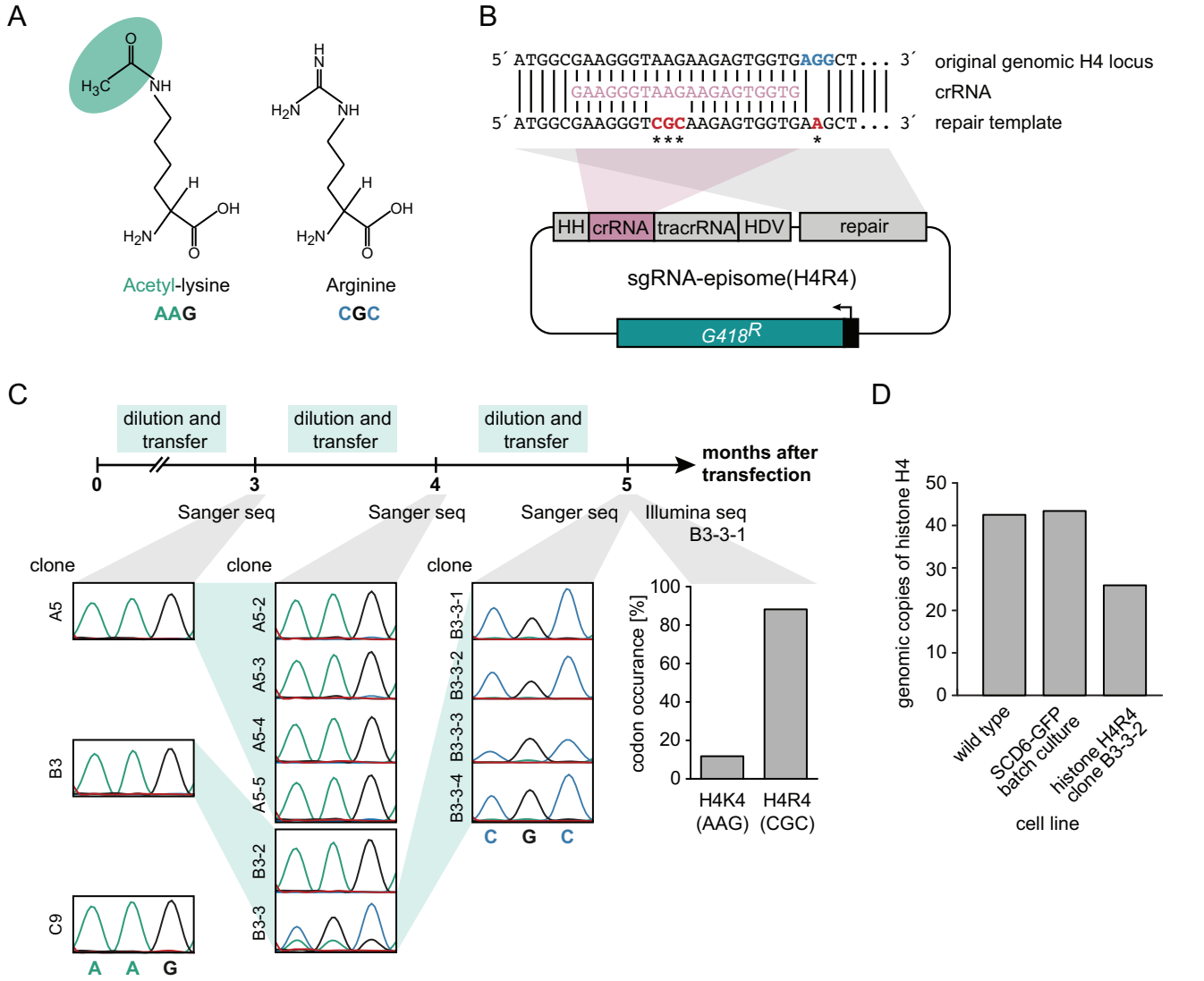

Figure 5. Cas9 allows editing of multicopy histone arrays. (**A**) Structure of acetylated lysine and arginine. (**B**) Sequences of the crRNA(H4R4) and the relevant regions of the H4R4 repair template. (**C**) Sanger-sequencing chromatograms from gDNA of cells transfected with the sgRNA-episome(H4R4). gDNA was extracted from multiple clones at different time points and revealed an accumulation of edited cells coding for H4R4 over time. Illumina-based gDNA-seq of clone B3-3-1 to 29x coverage indicated that after 5 months ~90% of histone H4 transcripts code for H4R4. (**D**) gDNA-seq-based quantification of histone H4 copy number (for details see methods section).

multicopy genes, three features that we consider major improvements over currently available approaches. While use of CRISPR–Cas9 has recently been described in *T. brucei* (34), the strategy aimed to enable high-throughput analyses and cannot be used for the three applications listed above, instead it provides a clever strategy for the rapid deletion of genes.

The CRISPR–Cas9-mediated approach developed in this study allows genome editing without the need to select for transcription of a resistance marker, a feature that opens unique opportunities to study the effect of transcription on chromatin structure and nuclear organization. It is becoming increasingly clear that chromatin structure and nuclear organization can affect transcription and vice versa (43). Thus, selection for the expression of a resistance marker gene inserted into a previously non-transcribed region of the genome may cause changes in chromatin structure and nuclear organization that in turn affects the expression of surrounding genes, complicating downstream analyses. The ability to manipulate the genome without the need to insert a selectable marker can be used to delete specific non-transcribed DNA sequence elements. Subsequently, ChIP-seq or FISH analyses may be used to study the effect of these deletions on chromatin structure or nuclear localization without the confounding effect of induced transcription of a selectable marker.

Furthermore, the precision of the outlined Cas9-mediated genome editing approach can be exploited to edit single nucleotides *in vivo*. This feature will be useful to unravel the role of individual amino acids within the active site of an enzyme or to investigate the importance of specific DNA motifs. Unusually for a eukaryote, expression of

RNA pol II-transcribed genes is regulated predominantly post-transcriptionally in *T. brucei*, at the level of RNA maturation (44), RNA stability (45) and translation (46–48). Numerous studies have identified distinct DNA sequence motifs that can affect the different steps (49–54). Using the episome-based CRISPR–Cas9 approach outlined here, it is now possible to make precise mutations in the UTR or to change the codon usage of a gene *in vivo*, an invaluable tool to study regulatory motifs that may help explain the large range in transcript-specific RNA stabilities and translational efficiencies. In addition, it will provide new options for one of the most commonly performed genetic manipulations, the addition of an epitope tag to a protein of interest. In this study we have demonstrated that Cas9-mediated genome editing can be used to insert the sequence coding for such a tag precisely between ORF and UTR. Previously, *in situ* tagging approaches developed for *T. brucei* involved the replacement of the endogenous UTR with one from a highly expressed gene, e.g. aldolase, tubulin or actin (55,56). In most cases the replacement of the UTR led to an increase in protein levels, which facilitates protein detection by immunofluorescence or western blotting, but is likely to result in an unnatural protein distribution across the cell. Thus far, the least 'disruptive' genome-editing approach developed for *T. brucei* takes advantage of the ability to remove *loxP*-flanked resistance marker cassettes by the transient expression of a CRE recombinase. While this strategy leaves only a 34-bp *loxP* 'scar' after excision of the *loxP*-flanked region (57), like all other currently available approaches, the insertion of the *loxP*-flanked resistance marker requires the selection for active transcription of the resistance marker gene. Thus, the approach presented here is not only less disruptive to the local transcriptional landscape, it is also more precise than other currently available approaches.

Finally, we have demonstrated that Cas9 technology can be used to edit multicopy genes in *T. brucei*. The *T. brucei* genome contains at least 194 gene families with ≥5 members, representing around 20% of the ∼9000 genes (58). The most widely employed approach to study multicopy genes is RNAi (59). While RNAi has been an extremely useful tool to study gene functions in *T. brucei*, its use can be limited by incomplete transcript depletion and off-target effects. For example the high degree of sequence similarity between the canonical histone H4 and the histone variant H4.V makes it impossible to efficiently specifically deplete one protein. In contrast, the CRISPR–Cas9 approach presented here can be used to edit multicopy genes without detectable off-target effects.

As a proof of principle, we have replaced a specific lysine residue of histone H4 with an arginine to mimic the constitutively non-acetylated state. Histone modifications are of great interest in *T. brucei* as they play important roles in DNA repair (5,60) and transcription (3,61). However, given that the responsible histone modifying enzymes are often unknown or tend to have multiple target sites, it has been difficult to unequivocally link individual histone modifications to specific biological functions. Recently we developed a quantitative proteomic-based approach to identify the target residues of histone acetyl transferases and deacetylases (42) and an MNase-ChIP-seq protocol for the detailed genome-wide mapping of individual histone modifications (24). In combination with these tools, the Cas9-based approach presented here represents an invaluable collection that should help unravel the role of histone modifications in trypanosomes. A better understanding of the function of individual histone modifications in this early branching eukaryote should shed light on the evolution of histone modification-mediated mechanisms.

DATA AVAILABILITY

The gDNA-seq data used for the off-target analysis and to estimate the histone *H4* gene array length have been deposited in EBI's European Nucleotide Archive and are accessible through ENA Study accession number PRJEB21645 (http://www.ebi.ac.uk/ena/data/view/PRJEB21645).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank all members of the Siegel laboratory, Christian Janzen and José Juan Lopez Rubio for valuable discussions and for sharing plasmids. We thank Stan Gorski for critical reading of the manuscript and Tobias Hertlein for the FACS-based analysis of GFP-positive cells.

FUNDING

Young Investigator Program of the Research Center of Infectious Diseases (ZINF) of the University of Würzburg, Germany; The German Research Foundation DFG [SI 1610/3-1]; George Forster fellowship from the Humboldt Foundation (to R.O.C.); European Research Council Starting Grant (3D_Tryps 715466). Funding for open access charge: European Research Council Starting Grant [3D_Tryps 715466].
Conflict of interest statement. None declared.

REFERENCES

1. Ruthenburg,A.J., Li,H., Patel,D.J. and Allis,C.D. (2007) Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell. Biol.*, **8**, 983–994.
2. Bernstein,B.E., Kamal,M., Lindblad-Toh,K., Bekiranov,S., Bailey,D.K., Huebert,D.J., McMahon,S., Karlsson,E.K., Kulbokas,E.J., Gingeras,T.R. *et al.* (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, **120**, 169–181.
3. Siegel,T.N., Hekstra,D.R., Kemp,L.E., Figueiredo,L.M., Lowell,J.E., Fenyo,D., Wang,X., Dewell,S. and Cross,G.A.M. (2009) Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. *Genes Dev.*, **23**, 1063–1076.
4. Price,B.D. and D'Andrea,A.D. (2013) Chromatin remodeling at DNA double-strand breaks. *Cell*, **152**, 1344–1354.
5. Glover,L. and Horn,D. (2014) Locus-specific control of DNA resection and suppression of subtelomeric VSG recombination by HAT3 in the African trypanosome. *Nucleic Acids Res.*, **42**, 12600–12613.
6. Kayne,P.S., Kim,U.J., Han,M., Mullen,J.R., Yoshizaki,F. and Grunstein,M. (1988) Extremely conserved histone H4 N terminus is dispensable for growth but essential for repressing the silent mating loci in yeast. *Cell*, **55**, 27–39.

7. Megee,P.C., Morgan,B.A., Mittman,B.A. and Smith,M.M. (1990) Genetic analysis of histone H4: essential role of lysines subject to reversible acetylation. *Science*, **247**, 841–845.
8. Pengelly,A.R., Copur,O., Jackle,H., Herzig,A. and Muller,J. (2013) A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. *Science*, **339**, 698–699.
9. Mali,P., Yang,L., Esvelt,K.M., Aach,J., Guell,M., DiCarlo,J.E., Norville,J.E. and Church,G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823–826.
10. Ghorbal,M., Gorman,M., Macpherson,C.R., Martins,R.M., Scherf,A. and Lopez-Rubio,J.J. (2014) Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR–Cas9 system. *Nat. Biotechnol.*, **32**, 819–821.
11. Hsu,P.D., Scott,D.A., Weinstein,J.A., Ran,F.A., Konermann,S., Agarwala,V., Li,Y., Fine,E.J., Wu,X., Shalem,O. *et al.* (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.*, **31**, 827–832.
12. Cong,L. and Zhang,F. (2015) Genome Engineering Using CRISPR–Cas9 System. *Methods Mol. Biol.*, **1239**, 197–217.
13. Kelso,A.A., Waldvogel,S.M., Luthman,A.J. and Sehorn,M.G. (2017) Homologous recombination in protozoan parasites and recombinase inhibitors. *Front. Microbiol.*, **8**, 1716.
14. Glover,L., Alsford,S. and Horn,D. (2013) DNA break site at fragile subtelomeres determines probability and mechanism of antigenic variation in african trypanosomes. *PLoS Pathog.*, **9**, e1003260.
15. Janzen,C.J., Fernandez,J.P., Deng,H., Diaz,R., Hake,S.B. and Cross,G.A.M. (2006) Unusual histone modifications in *Trypanosoma brucei*. *FEBS Lett.*, **580**, 2306–2310.
16. Patnaik,P.K., Fang,X. and Cross,G.A. (1994) The region encompassing the procyclic acidic repetitive protein (PARP) gene promoter plays a role in plasmid DNA replication in Trypanosoma brucei. *Nucleic Acids Res.*, **22**, 4111–4118.
17. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.
18. Alsford,S., Kawahara,T., Glover,L. and Horn,D. (2005) Tagging a T. brucei RRNA locus improves stable transfection efficiency and circumvents inducible expression position effects. *Mol. Biochem. Parasitol. Mol. Biochem. Parasitol.*, **144**, 142–148.
19. Burkard,G., Fragoso,C.M. and Roditi,I. (2007) Highly efficient stable transformation of bloodstream forms of Trypanosoma brucei. *Mol. Biochem. Parasitol.*, **153**, 220–223.
20. Burkard,Schumann, Jutzi,G. and Roditi,I. (2011) Genome-wide RNAi screens in bloodstream form trypanosomes identify drug transporters. *Mol. Biochem. Parasitol.*, **175**, 91–94.
21. Brun,R. and Schonenberger,M. (1979) Cultivation and in vitro cloning or procyclic culture forms of *Trypanosoma brucei* in a semi-defined medium. *Acta Trop.*, **36**, 289–292.
22. Patnaik,P.K., Kulkarni,S.K. and Cross,G.A.M. (1993) Autonomously replicating single-copy episomes in *Trypanosoma brucei* show unusual stability. *EMBO J.*, **12**, 2529–2538.
23. Sambrook,J. and Russell,D.W. (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY.
24. Wedel,C. and Siegel,T.N. (2017) Genome-wide analysis of chromatin structures in Trypanosoma brucei using high-resolution MNase-ChIP-seq. *Exp. Parasitol.*, **180**, 2–12.
25. Sollelis,L., Ghorbal,M., MacPherson,C.R., Martins,R.M., Kuk,N., Crobu,L., Bastien,P., Scherf,A., Lopez-Rubio,J.J. and Sterkers,Y. (2015) First efficient CRISPR–Cas9-mediated genome editing in Leishmania parasites. *Cell Microbiol.*, **17**, 1405–1412.
26. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
27. Koboldt,D.C., Zhang,Q., Larson,D.E., Shen,D., McLellan,M.D., Lin,L., Miller,C.A., Mardis,E.R., Ding,L. and Wilson,R.K. (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.*, **22**, 568–576.
28. MacPherson,C.R. and Scherf,A. (2015) Flexible guide-RNA design for CRISPR applications using Protospacer Workbench. *Nat. Biotechnol.*, **33**, 805–806.
29. Weese,D., Holtgrewe,M. and Reinert,K. (2012) RazerS 3: faster, fully sensitive read mapping. *Bioinformatics*, **28**, 2592–2599.
30. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features *Bioinformatics*, **26**, 841–842.
31. Peng,D., Kurup,S.P., Yao,P.Y., Minning,T.A. and Tarleton,R.L. (2015) CRISPR–Cas9-Mediated Single-Gene and gene family disruption in trypanosoma cruzi. *MBio*, **6**, e02097-14.
32. Nakaar,V., Günzl,A., Ullu,E. and Tschudi,C. (1997) Structure of the *Trypanosoma brucei* U6 snRNA gene promoter. *Mol. Biochem. Parasitol.*, **88**, 13–23.
33. Wagner,J.C., Platt,R.J., Goldfless,S.J., Zhang,F. and Niles,J.C. (2014) Efficient CRISPR–Cas9-mediated genome editing in Plasmodium falciparum. *Nat. Methods*, **11**, 915–918.
34. Beneke,T., Madden,R., Makin,L., Valli,J., Sunter,J. and Gluenz,E. (2017) A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids. *R. Soc. Open Sci.*, **4**, 170095.
35. Wirtz,E., Hartmann,C. and Clayton,C. (1994) Gene expression mediated by bacteriophage T3 and T7 RNA polymerases in transgenic trypanosomes. *Nucleic Acids Res.*, **22**, 3887–3894.
36. Gao,Y. and Zhao,Y. (2014) Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *J. Integr. Plant Biol.*, **56**, 343–349.
37. Krüger,T., Hofweber,M. and Kramer,S. (2013) SCD6 induces ribonucleoprotein granule formation in trypanosomes in a translation-independent manner, regulated by its Lsm and RGG domains. *Mol. Biol. Cell*, **24**, 2098–2111.
38. Kramer,S., Queiroz,R., Ellis,L., Webb,H., Hoheisel,J.D., Clayton,C. and Carrington,M. (2008) Heat shock causes a decrease in polysomes and the appearance of stress granules in trypanosomes independently of eIF2(alpha) phosphorylation at Thr169. *J. Cell Sci.*, **121**, 3002–3014.
39. Fu,Y., Foden,J.A., Khayter,C., Maeder,M.L., Reyon,D., Joung,J.K. and Sander,J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.*, **31**, 822–826.
40. Pattanayak,V., Lin,S., Guilinger,J.P., Ma,E., Doudna,J.A. and Liu,D.R. (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.*, **31**, 839–843.
41. Siegel,T.N., Kawahara,T., Degrasse,J.A., Janzen,C.J., Horn,D. and Cross,G.A.M. (2008) Acetylation of histone H4K4 is cell cycle regulated and mediated by HAT3 in *Trypanosoma brucei*. *Mol. Microbiol.*, **67**, 762–771.
42. ElBashir,R., Vanselow,J.T., Kraus,A., Janzen,C.J., Siegel,T.N. and Schlosser,A. (2015) Fragment ion patchwork quantification for measuring Site-Specific acetylation degrees. *Anal. Chem.*, **87**, 9939–9945.
43. Therizols,P., Illingworth,R.S., Courilleau,C., Boyle,S., Wood,A.J. and Bickmore,W.A. (2014) Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science*, **346**, 1238–1242.
44. Fadda,A., Ryten,M., Droll,D., Rojas,F., Farber,V., Haanstra,J.R., Merce,C., Bakker,B.M., Matthews,K. and Clayton,C. (2014) Transcriptome-wide analysis of trypanosome mRNA decay reveals complex degradation kinetics and suggests a role for co-transcriptional degradation in determining mRNA levels. *Mol. Microbiol.*, **94**, 307–326.
45. Manful,T., Fadda,A. and Clayton,C. (2011) The role of the 5′-3′ exoribonuclease XRNA in transcriptome-wide mRNA degradation. *RNA*, **17**, 2039–2047.
46. Vasquez,J.J., Hon,C.C., Vanselow,J.T., Schlosser,A. and Siegel,T.N. (2014) Comparative ribosome profiling reveals extensive translational complexity in different Trypanosoma brucei life cycle stages. *Nucleic Acids Res.*, **42**, 3623–3637.
47. de Freitas Nascimento,J., Kelly,S., Sunter,J. and Carrington,M. (2018) Codon choice directs constitutive mRNA levels in trypanosomes. *Elife*, **7**, e32467.
48. Jeacock,L., Faria,J. and Horn,D. (2018) Codon usage bias controls mRNA and protein abundance in trypanosomatids. *Elife*, **7**, e32496.
49. Horn,D. (2008) Codon usage suggests that translational selection has a major impact on protein expression in trypanosomatids. *BMC Genomics*, **9**, 2.
50. Hehl,A., Vassella,E., Braun,R. and Roditi,I. (1994) A conserved stem-loop structure in the 3′ untranslated region of procyclin mRNAs regulates expression in *Trypanosoma brucei*. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 370–374.

51. Furger,A., Schurch,N., Kurath,U. and Roditi,I. (1997) Elements in the 3′ untranslated region of procyclin mRNA regulate expression in insect forms of *Trypanosoma brucei* by modulating RNA stability and translation. *Mol. Cell. Biol.*, **17**, 4372–4380.
52. Hotz,H.R., Hartmann,C., Huober,K., Hug,M. and Clayton,C. (1997) Mechanisms of developmental regulation in *Trypanosoma brucei*: a polypyrimidine tract in the 3′-untranslated region of a surface protein mRNA affects RNA abundance and translation. *Nucleic Acids Res.*, **25**, 3017–3026.
53. Siegel,T.N., Tan,K.S. and Cross,G.A.M. (2005) Systematic study of sequence motifs for RNA *trans* splicing in *Trypanosoma brucei*. *Mol. Cell. Biol.*, **25**, 9586–9594.
54. Mayho,M., Fenn,K., Craddy,P., Crosthwaite,S. and Matthews,K. (2006) Post-transcriptional control of nuclear-encoded cytochrome oxidase subunits in *Trypanosoma brucei*: evidence for genome-wide conservation of life-cycle stage-specific regulatory elements. *Nucleic Acids Res.*, **34**, 5312–5324.
55. Oberholzer,M., Morand,S., Kunz,S. and Seebeck,T. (2006) A vector series for rapid PCR-mediated C-terminal in situ tagging of *Trypanosoma brucei* genes. *Mol. Biochem. Parasitol.*, **145**, 117–120.
56. Dean,S., Sunter,J., Wheeler,R.J., Hodkinson,I., Gluenz,E. and Gull,K. (2015) A toolkit enabling efficient, scalable and reproducible gene tagging in trypanosomatids. *Open Biol.*, **5**, 140197.
57. Scahill,M.D., Pastar,I. and Cross,G.A.M. (2008) CRE recombinase-based positive-negative selection systems for genetic manipulation in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **157**, 73–82.
58. Barry,J.D., McCulloch,R., Mottram,J.C. and Acosta-Serrano,A. (2007) *Trypanosomes: After the Genome*. Horizon Bioscience, Norfolk.
59. Ngo,H., Tschudi,C., Gull,K. and Ullu,E. (1998) Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 14687–14692.
60. Glover,L. and Horn,D. (2012) Trypanosomal histone gammaH2A and the DNA damage response. *Mol. Biochem. Parasitol.*, **183**, 78–83.
61. Wright,J.R., Siegel,T.N. and Cross,G.A.M. (2010) Histone H3 trimethylated at lysine 4 is enriched at probable transcription start sites in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **136**, 434–450.