A high-throughput protein tagging toolkit that retains endogenous UTRs for studying gene regulation in Kinetoplastids.

Carla Gilabert Carbajo[1], Xiaoyang Han[1], Bhairavi Savur[1], Arushi Upadhyaya[1], Fatima Taha[1], Michele Tinti[2], Richard J Wheeler[3,4], Phillip A. Yates[5#], & Calvin Tiengwe[1#]

[1]Department of Life Sciences, Imperial College London, United Kingdom
[2]Wellcome Trust Centre for Anti-Infectives Research, University of Dundee, United Kingdom
[3]Peter Medawar Building for Pathogen Research, University of Oxford, United Kingdom
[4]Institute of Immunology and Infection Research, University of Edinburgh, United Kingdom
[5]Department of Chemical Physiology and Biochemistry, Oregon Health & Science University, Portland OR, United States of America.

[#]Co-corresponding authors: yatesp@ohsu.edu and c.tiengwe@imperial.ac.uk

Running title: An endogenous tagging toolkit with UTR retention in Kinetoplastids

Key words: Kinetoplastids, gene regulation, CRISPR/Cas9, molecular tools, T2A peptide, endogenous tagging, untranslated regions (UTRs), differentiation, *Trypanosoma brucei*

Abstract

Kinetoplastid parasites cause diseases that threaten human and animal health. To survive transitions between vertebrate hosts and insect vectors, these parasites rely on precise regulation of gene expression to adapt to environmental changes. Since gene regulation in Kinetoplastids is primarily post-transcriptional, developing efficient genetic tools for modifying genes at their endogenous loci while preserving regulatory mRNA elements is crucial for studying their complex biology.

We present a CRISPR/Cas9-based tagging system that preserves untranslated regulatory elements and uses a viral 2A peptide from *Thosea asigna* to generate two separate proteins from a single transcript: a drug-selectable marker and a tagged protein of interest. This dual-function design maintains native control elements, allowing discrimination between regulation of transcript abundance, translational efficiency, and post-translational events.

We validate the system by tagging six *Trypanosoma brucei* proteins and demonstrate: (i) high-efficiency positive selection and separation of drug-selectable marker and target protein, (ii) preservation of regulatory responses to environmental cues like heat shock and iron availability, and (iii) maintenance of stage-specific regulation during developmental transitions. This versatile toolkit is applicable to all kinetoplastids amenable to CRISPR/Cas9 editing, providing a powerful reverse genetic tool for studying post-transcriptional regulation and protein function in organisms where post-transcriptional control is dominant.

Introduction

Kinetoplastid parasites, including *Trypanosoma brucei*, *T. cruzi*, and *Leishmania species*, cause significant vector-borne diseases in humans and animals. These parasites undergo complex developmental transitions as they switch between vertebrates and insect vectors, adapting to drastically different environments. Unlike most eukaryotes, kinetoplastids rely primarily on post-transcriptional gene regulation, with untranslated regions (UTRs) playing a crucial role in controlling mRNA fate and translation efficiency (1-3). Most protein-coding genes are transcribed by RNA Pol II in polycistronic units, with their expression controlled through mRNA stability and translation, as seen with heat shock protein 70 (Hsp70) regulated by 3′-UTR elements during the stress response (4-6), and the calpain-related proteins CAP5.5V and CAP5.5 during life cycle transitions (7, 8). Only a small subset of genes, including expression site-associated genes (ESAGs), undergo transcriptional regulation by RNA Pol I (9).

A prime example of UTR-mediated regulation is seen in ESAG6 and ESAG7, which encode subunits of the iron transporter, the transferrin receptor (TfR), in *T. brucei.* Their 3' UTRs contain iron-responsive elements that enhance TfR mRNA stability in response to iron starvation (10, 11). Other important examples are procyclin and VSG, the major surface proteins of procyclic and bloodstream forms respectively, which are regulated by elements in their 3' UTRs. Procyclin expression is controlled by several 3' UTR elements, first identified by Hug et al (12), including a conserved 16-mer stem-loop structure (13) and additional regulatory sequences that affect RNA stability and translation (14-16). Conversely, the VSG mRNA 3' UTR contains regulatory elements (17) including a 16-mer sequence essential for mRNA stability (18). More recently, a genome-wide analysis has predicted numerous UTR motifs and identified thousands of 3' UTRs likely to influence mRNA abundance across different life cycle stages (19), highlighting the critical role of UTRs in enabling parasite adaptation.

Preserving UTRs is essential for studying gene function in kinetoplastids, as their disruption can alter expression patterns (10, 11). Several tagging methods have been developed for kinetoplastids, each with trade-offs between efficiency and preservation of regulatory elements. Early approaches, like long primer PCR (20, 21), were efficient, while others required cloning gene fragments for homologous recombination (22). PCR-only tagging in *T. brucei* and similar methods in *Leishmania* enabled scalable tagging but disrupt native UTRs when combined with drug-selectable markers (23). The TrypTag project, which tagged nearly 90% of the *T. brucei*

77 proteome, has been a valuable resource for localisation studies, albeit with the recognised caveat
78 that it replaces UTRs (24). CRISPR/Cas9 improves efficiency and precision but often introduces
79 selection markers that interfere with endogenous regulation (25). Two CRISPR-based marker-
80 free methods have been reported which preserve UTRs but require extensive subcloning by
81 limiting dilution (26, 27). While these methods have advanced functional genomics, most still
82 compromise regulatory elements, highlighting the need for tools that preserve UTR integrity while
83 enabling efficient, high-throughput protein tagging and functional analysis.
84
85 First discovered in the Foot-and-Mouth Disease Virus (FMDV), the viral 2A peptide has become
86 a powerful tool for expressing multiple proteins from a single open reading frame (ORF) (28, 29).
87 Several versions of the 2A peptide have been identified across different viruses, each with similar
88 functionality. These include FMDV (F2A), porcine teschovirus-1 (P2A), and *Thosea asigna* virus
89 (T2A) (30). Interestingly, 2A-like sequences (L2A) have also been found in the L1Tc non-LTR
90 retrotransposons of *T. cruzi*, *T. brucei*, *T. vivax* and *T. congolense,* which represents a rare
91 example of these sequences occurring outside of a non-viral context (31). In general, the T2A
92 peptide, with the sequence -EGRGSLLTCGDVEENPG↓P-, facilitates a mechanism known as
93 "ribosomal skipping," where the glycine-proline bond fails to form during translation, resulting in
94 the production of two distinct proteins from the same transcript (28). This process, often called
95 CHYSEL (cis-acting hydrolase element), “stop-and-go”, “self-processing”, or “co-translational
96 cleavage”, will hereafter be referred to as “cleavage” for simplicity, with instances where the
97 process fails being referred to as “uncleaved” proteins.
98
99 In Kinetoplastids, the T2A peptide has been shown to be highly efficient for co-expressing proteins
100 under endogenous UTRs, enabling the study of post-transcriptional regulatory mechanisms.
101 Some examples include the *L. mexicana* glucose transporter (LmxGT1) (32), the cortical
102 cytoskeletal protein (KHARON1) (33) and the nucleotide transporter NT3 in *L. donovani* (34). The
103 T2A system has also been applied for ectopic expression in *T. cruzi*, improving positive selection
104 of transgenic parasites (35). These applications demonstrate the versatility of the T2A peptide
105 system for studying protein localisation and regulatory mechanisms in kinetoplastids.
106
107 In this study, we combined CRISPR/Cas9 editing with the T2A peptide-based system to test its
108 applicability in *T. brucei* by tagging two secretory proteins (ESAG3, ESAG7) which are under
109 transcriptional and post-transcriptional regulation, and cytosolic Hsp70, part of a multi-gene array
110 whose expression is regulated by its 3' UTR in response to heat shock. We demonstrate efficient

111 T2A peptide cleavage and regulation of tagged mRNA and/or proteins in their native loci. We also
112 constructed new vectors with primer binding sites compatible with the plasmids for PCR-only
113 tagging (pPOT series) (23) and designed an automated primer design tool to facilitate high-
114 throughput tagging while preserving native UTRs. We validated the new vectors on three well-
115 characterised proteins: glycosylphosphatidylinositol phospholipase C (GPI-PLC) (36) and two
116 calpain-related proteins (CAP5.5V and CAP5.5) (7, 8), which show stage-specific expression, and
117 confirmed that this differential expression was maintained in the tagged versions. Together, we
118 provide a versatile tool that can be used at scale to study post-transcriptional and post-
119 translational regulatory mechanisms in Kinetoplastids.
120
121 **Results**
122
123 **Development of a CRISPR/Cas9-compatible T2A peptide tagging system for *T. brucei***
124
125 Our primary goal was to exploit the high efficiency of CRISPR/Cas9 genome editing to insert a
126 "T2A peptide cassette" between the open reading frame (ORF) of a target gene and its native 5'
127 or 3' UTR in *T. brucei*. The cassette contains a drug-selectable marker (drug^R), a 2A peptide
128 sequence (T2A), and a fluorescent protein or an epitope tag (Tag) (**Fig. 1**). We initially constructed
129 donor plasmids in two configurations: drug^R::T2A::Tag for insertion at the 5' end (**Fig. 1A**), and
130 Tag::T2A::drug^R for insertion at the 3' end of the target gene (**Fig. 1B**). This design allowed for
131 positive selection of transgenic parasites while preserving native mRNA processing signals within
132 the UTRs, thus maintaining natural regulation. Available plasmids with fluorescent tags:
133 mNeonGreen (mNG) or mScarlet (mSc), and drug-selectable markers: blasticidin (BSD), or
134 puromycin (PAC) are shown in **Fig. S1**.
135
136 For 5'-terminal tagging, the donor repair template drug^R::T2A::Tag was amplified by PCR using
137 primers containing 25-40 bp homology arms, targeting a protospacer adjacent motif (PAM) site
138 near the start of the ORF. The ATG start codon was removed, but the 5' UTR remained intact.
139 For 3'-terminal tagging, the Tag::T2A::drug^R donor template was amplified using primers with
140 homology arms to integrate the cassette before the stop codon, targeting a PAM site near the end
141 of the ORF. This integration strategy incorporated 57 bp of plasmid backbone sequence between
142 the targeting cassette and the native 3' UTR, though the full-length 3' UTR remained intact
143 downstream of this insertion. In both cases the single-guide RNA (sgRNA) was generated by
144 combining a gene-specific primer encoding a T7 promoter upstream, a 20-bp targeting sequence

145 specifying the target site, and a 3'-end complementary to the sgRNA scaffold, as described in
146 (37).
147 Both sgRNA and donor T2A cassettes were amplified by PCR using well-established protocols
148 described by Beneke et al. (25). The PCR products were pooled for transfection into bloodstream
149 form *T. brucei*, which constitutively express T7 RNA polymerase and Cas9. Cas9 introduces a
150 double strand break at the target site, guided by the sgRNA designed near the PAM site at either
151 the 5' or 3' end of the target gene. We incorporated mutated PAM sites into the repair template to
152 prevent further Cas9 edits post-integration. Once integrated, the T2A peptide sequence mediates
153 co-translational cleavage, allowing for independent expression of the drug resistance marker and
154 the tagged protein of interest in equal amounts (38).
155
156 The donor plasmids described here were originally constructed for *Leishmania*, where we
157 determined that T2A peptide conferred the highest cleavage efficiency among the various 2A
158 peptides tested (**Fig S2**).
159
160 **Test case I: ESAG3 endogenous tagging shows efficient T2A cleavage and robust positive**
161 **selection**
162
163 To assess the efficiency of T2A peptide cleavage in *T. brucei* and demonstrate its utility for protein
164 localisation studies, we targeted the *expression site associated gene 3* (*ESAG3*, gene IDs:
165 Tb427.BES40.10, Tb427.BES40.16). *ESAG3* encodes an N-terminal putative ER targeting signal
166 sequence with a predicted cleavage site between Ala22/Leu23 (SignalP v5.0). A two-component
167 PCR amplicon mix was transfected into *T. brucei* T7/Cas9-expressing cells, consisting of a 20-bp
168 sgRNA targeting adjacent to the PAM site near the stop codon and a donor mNG::T2A::PAC
169 cassette with 40/30 bp homology arms. The stop codon was excluded from repair primers to allow
170 in-frame fusion of mNG at the C-terminus, while synonymous mutations in the PAM site prevented
171 further Cas9 edits.
172
173 Seven days post-transfection, genomic DNA PCR confirmed in-frame fusion of *ESAG3* ORF with
174 mNG (expected size 594 bp), *PAC* with *ESAG3* 3'UTR (211 bp), *ESAG3* ORF with its native UTR
175 (2.5 Kb) (**Fig 2A**). No PCR product was generated using genomic DNA from parental cells
176 confirming specificity. To compare the relative mobility of the ESAG3::mT2A protein, we
177 generated independent cell lines with C-terminally mNG-tagged ESAG3 fused with 6xTy
178 (ESAG3::mTy) using the PCR-only tagging method (37) (**Fig 2B**). Western blot analysis revealed

179 that the ESAG3::mT2A protein (69 kDa) migrated faster than ESAG3::mTy (76 kDa), indicative of
180 efficient T2A cleavage. The expected MW of uncleaved ESAG3::mNG::T2A::PAC is 90 kDa.
181 Since there are two copies of ESAG3 in the active expression site (BES1, Tb427_telo40_v2) in
182 the Lister427 cell line (39), native anti-ESAG3 antibody immunoblots confirmed that tagging was
183 single copy for most clones, as the untagged copy was detectable at the expected MW of ~43
184 kDa (**Fig 2B)**. Flow cytometry showed that 93-95% of cells in the drug-selected clonal population
185 were mNG-positive, confirming robust positive selection of correctly tagged cells (**Fig 2C**).
186 Immunofluorescence microscopy showed a typical ER localisation pattern for ESAG3::mT2A,
187 confirming proper targeting to the secretory pathway (**Fig 2D**).
188
189 These results demonstrate precise successful insertion of the mNG::T2A::PAC cassette at the 3'
190 end of *ESAG3*, with highly efficient positive selection of tagged cells, and successful co-
191 translational T2A cleavage in *T. brucei*. Furthermore, this tagging system enabled us to localise
192 ESAG3 in bloodstream form trypanosomes, for the first time.
193
194 **Test case II: Endogenous T2A tagging of Hsp70 retains mRNA-level heat shock regulation**
195
196 To further demonstrate the robustness of the T2A system, we targeted the cytosolic heat shock
197 protein Hsp70. In *T. brucei*, Hsp70 is encoded by a tandem array of eight nearly identical genes
198 (IDs: Tb427_110128800, Tb427_110128900 … Tb427_110129500) on chromosome 11 (40). In
199 the Tb927 genome assembly, this array is collapsed into a single gene (Tb927.11.11330) (4, 41).
200 Hsp70 mRNA stability is regulated by heat response elements in the 3′-UTR during heat shock
201 (4-6), but corresponding changes at the protein level have not been demonstrated. Our goals
202 were to investigate the behaviour of endogenously T2A-tagged Hsp70 following heat shock and
203 to assess whether our system facilitated single-copy integration within a multi-gene array.
204
205 We designed primers based on the Tb927 genome copy to tag Hsp70 at both the N- and C-termini
206 with mScarlet (mSc) and mNeonGreen (mNG), respectively, under puromycin selection (**Fig. 3A**).
207 The design of sgRNA and donor repair templates followed the strategy used for ESAG3. A GSGS
208 linker was added between the mSc tag and the *Hsp70* ORF for N-terminal tagging, and 25 bp
209 homology arms were used in both repair templates. Post-transfection, locus-specific PCR
210 genotyping of two clones produced products of sizes consistent with the predicted fusions: *Hsp70*
211 ORF with mNG (409 bp), *PAC* with the *Hsp70* 3' UTR (304 bp), and *Hsp70* 5' UTR with *PAC* (339
212 bp) (**Fig. 3A**). No products were generated from genomic DNA purified from parental cells,

213 confirming locus-specific modification.
214
215 Western blot analysis with anti-Hsp70 antibodies showed two bands in all selected clones: one at
216 ~70 kDa (untagged endogenous Hsp70) and another at ~100 kDa, (tagged Hsp70: 101 kDa for
217 Hsp70::mT2A and 98 kDa for T2A-mSc::H). The higher MW bands were absent in parental cells
218 (**Fig. 3B**), indicating successful tagging. Flow cytometry revealed that ~99% of cells were positive
219 for either mNG or mSc (**Fig 3C**), and live-cell microscopy confirmed the characteristic cytosolic
220 localisation of Hsp70::mT2A (**Fig. 3D**), indicating that T2A tagging did not cause Hsp70 mis-
221 localisation.
222
223 To assess whether expression of mT2A-tagged Hsp70 tracked with the untagged allele following
224 heat shock, we performed immunoblotting with anti-mScarlet, anti-mNG, and anti-Hsp70
225 antibodies. We observed a slight reduction in both tagged and endogenous Hsp70 protein levels
226 after heat shock (**Fig. 3E**). However, qRT-PCR analysis showed a 2-fold upregulation of Hsp70
227 mRNA (**Fig. 3F**), suggesting a typical heat shock response at the mRNA level (6). We attribute
228 the reduction in protein levels to either global translational arrest or cell stress resulting from the
229 heat shock treatment (42).
230
231 Overall, our results confirmed UTR-dependent *Hsp70* mRNA regulation after heat shock (4), but
232 not at the protein level - a pattern consistent with established models in *T. brucei* where selective
233 mRNA stabilisation enables continued Hsp70 synthesis during stress-induced translational
234 repression (5, 6). We successfully achieved efficient single-copy tagging of cytosolic Hsp70 at
235 both the N- and C-termini within a complex multi-gene array. The apparent lower expression of
236 tagged Hsp70 compared to the wild type likely indicates that only one of the eight nearly identical
237 gene copies was tagged, accounting for approximately 1/16th of the total signal in Western blots
238 of a diploid genome. We confirmed this using whole genome sequencing analysis and found no
239 evidence of deletions of any Hsp70 genes in the array (**Fig. S3**).
240
241 **Test case III: T2A-tagged ESAG7 preserves iron-dependent regulation and function**
242
243 In functional studies, it is essential that tagging does not disrupt protein function or its regulatory
244 elements at the RNA level. In *T. brucei,* ESAG7 (E7) exemplifies this dual requirement: at the
245 protein level, it forms a 1:1 stoichiometric complex with ESAG6 to create a functional transferrin
246 receptor (TfR) for iron uptake (43). At the RNA level, its expression is regulated by iron-responsive

elements in its 3' UTR that control mRNA stability in response to iron availability (44). Tagging approaches that replace the native 3' UTR can disrupt this iron-dependent regulation, potentially altering the expression levels and stoichiometry of ESAG6 and ESAG7, thereby interfering with TfR function (10, 11).

To investigate regulation and function of ESAG7, we generated mNG::T2A- and mNG::6xTy-tagged ESAG7 cell lines (E7::mT2A and E7::mTy, respectively) (**Fig 4A**). E7::mT2A tagging retained the native *E7* 3' UTR, while the E7::mTy tagging replaced the 3' UTR with the *PFR2* intergenic region. Hsp70 (H::mT2A) was also tagged as a control. All analyses were performed after treatment with the iron chelator Deferoxamine (DFO), which induces iron starvation. qRT-PCR analysis showed 2-fold increased mNG transcript levels in E7::mT2A cells following DFO treatment, while E7::mTy showed a reduction, and H::mT2A cells remained unchanged (**Fig 4B**). This pattern was reflected at the protein level, as Western blotting confirmed increased E7::mT2A protein abundance, with E7::mTy and H::mT2A remaining unaffected (**Fig 4C**). Flow cytometry further supported these findings, showing a 3-fold increase in mNG expression in E7::mT2A cells, while no significant changes were observed in E7::mTy or H::mT2A (**Fig 4D**).

The functional implications of these findings were assessed by analysing native TfR levels and transferrin (Tf) uptake. Both qRT-PCR and Western blot analyses revealed increased TfR levels in parental, H::mT2A, and E7::mT2A cells after DFO treatment, but showed a significant reduction in E7::mTy cells (**Fig 4E and 4F**). Tf uptake assays demonstrated significant increased uptake in E7::mT2A cells, confirming functional TfR activity, whereas E7::mTy cells showed no Tf uptake, indicating a disruption of iron-dependent regulation and receptor function (**Fig 4G**).

Collectively, these data demonstrate that mNG::T2A-tagged ESAG7, when retaining its native 3' UTR, maintains iron-dependent regulation and function. In contrast, replacing the 3' UTR with the *PFR2* intergenic region disrupts both iron-responsive regulation and TfR function, likely by perturbing the coordinated expression of ESAG7 with its heterodimeric partner ESAG6 required for functional receptor assembly. This validates our T2A-tagging approach as an effective tool for studying post-transcriptional iron regulation and hence likely other regulatory systems in *T. brucei*.

Confirmation of site-specific integration and in-frame fusion of the T2A gene tagging by whole genome sequencing

281 Finally, to confirm site-specific integration and in-frame fusion of mNG::T2A with the target ORFs,
282 we performed paired-end whole genome sequencing (WGS) of the five tagged cell lines at 30X
283 genome coverage. The reads were mapped to either the wild-type genomic locus or a manually
284 reconstructed locus containing the mNG::T2A or mNG::6xTy tagging cassettes fused to the target
285 genes (**Fig. S4**). In all cases, WGS confirmed that the donor repair cassettes were inserted in-
286 frame with the target gene in tagged cells, with no frameshifts or premature stop codons observed,
287 confirming site-specific in-frame fusion. Collectively, combining data from WGS, integration
288 PCRs, and Western blot analyses, we conclude that mNG::T2A fused to target ORFs
289 predominantly as single-copy integrations, and that T2A peptide-mediated cleavage occurs with
290 high efficiency. Collectively these results show that this approach is highly suitable for epitope-
291 tagging secretory and cytosolic proteins at their endogenous loci in *T. brucei*.
292
293 **Dedicated modular plasmid templates for high-throughput CRISPR/Cas9 T2A tagging: The**
294 **pRExT2A system**
295
296 To enhance the efficiency and versatility of T2A tagging in high-throughput applications, we
297 developed the pRExT2A system - plasmids for *R*egulated co-*Ex*pression mediated by *T2A*
298 peptide) - a set of dedicated plasmid templates optimised for *T. brucei*. This new design addresses
299 two key limitations of the initial T2A tagging system where: (i) the coding sequences for *mNG*,
300 *mSc*, and $drug^R$ were optimised for *Leishmania* codon usage, which differs from *T. brucei* and
301 could lead to inefficient translation and reduced expression levels; and (ii) the addition of 57 bp to
302 the 3' UTR during C-terminal tagging, which could potentially disrupt native mRNA processing
303 signals and affect gene regulation.
304
305 To overcome these issues, we designed pRExT2A to be fully compatible with the pPOT (plasmids
306 for PCR-only tagging) platform by incorporating the coding sequences for drug resistance markers
307 and fluorescent proteins from pPOT into our new plasmids (23). Our goal was to integrate the
308 efficiency, modularity, and cloning-free advantages of pPOT while ensuring that primer usage is
309 compatible across both systems.
310
311 **Features and primer design of the pRExT2A system:** The pRExT2A system features modular
312 N- and C-terminal tagging donor plasmids encoding either a 3xTy-tagged fluorescent protein or a
313 3xTy epitope tag, linked via the T2A peptide to a 3xTy-tagged $drug^R$ (**Fig. 5**). For N-terminal
314 tagging, the constructs follow the format: $drug^R$::3xTy::T2A::3xTy::FP or $drug^R$::3xTy::T2A::3xTy

315 (**Fig. 5A**), while for C-terminal tagging the orientation was as follows: FP::3xTy::T2A::drugR::3xTy
316 or 3xTy::T2A::drug^{R}::3xTy (**Fig. 5B**).
317 The primer design workflow remains consistent with our previous protocol (**Fig. 1**), but we
318 modified the primer binding sites to match the pPOT system. This allows users to use identical
319 primer pairs and sgRNAs across both systems. Although an improvement over the original T2A
320 tagging cassettes, this design introduces a short 20 bp scar into the 5' or 3' UTR after integration
321 - depending on whether N- or C-terminal tagging is used (**Fig. 5A** and **5B**). To facilitate high-
322 throughput applications, we developed an automated primer design tool
323 [https://github.com/zephyris/scarlesstagging] which generates primers compatible with both
324 pRExT2A and pPOT systems. This tool significantly simplifies experimental setup and minimises
325 design errors. To facilitate adoption, we provide pre-designed primers (including scarless options)
326 for all *T. brucei* genes in TREU927, Lister 427 strains, and selected species, available at:
327 https://github.com/zephyris/scarlesstagging/tree/main/examples/py/tritrypdb-genomes-v68.
328
329 **Dual functionality of the 3xTy-tagged drug^{R}:** A key feature of the pRExT2A system is the
330 inclusion of a 3xTy-tagged drug^{R} which serves two functions: as a selection marker controlled by
331 the endogenous UTRs of the target gene and an internal reference for measuring relative protein
332 abundance before and after stimulus exposure (**Fig. 5C)**. Since the drug^{R} and the target protein
333 are translated from the same mRNA but yield distinct proteins, changes in mRNA abundance or
334 translational efficiency affect both proteins proportionally. Comparing protein levels using Western
335 blotting or flow cytometry reveals whether regulation occurs at the transcript level (where both
336 proteins change similarly) or through differential protein stability (where the target protein changes
337 independently of the drug^{R}) (**Fig. 5C** and **5D**). This design enables researchers to determine the
338 level of regulation by measuring mRNA abundance (using qRT-PCR) and protein levels within
339 the framework of a single experiment. A constant protein/drug^{R} indicates regulation at the mRNA
340 or translational level, while changes in ratios from baseline reflect differential protein stability (**Fig**
341 **5D**), as exemplified by T2A-based reporter systems in *Leishmania* (32, 34).
342
343 **Modularity of the pRExT2A System**: We developed plasmid versions with various drug-
344 selectable markers, including blasticidin (BSD), hygromycin (HYG), and puromycin (PAC), as well
345 as two fluorescent proteins, mNeonGreen (mNG) and mScarlet (mSc) (**Table 1**). This modular
346 design allows for easy swapping of tags and drug^{R} markers, providing flexibility to accommodate
347 different experimental needs.
348

349 **The pRExT2A system retains life cycle stage-specific gene regulation and confirms UTR-**
350 **mediated control**
351 To evaluate the ability of the pRExT2A system to preserve critical regulatory elements, we tested
352 three developmentally regulated genes in *T. brucei*: glycosylphosphatidylinositol phospholipase
353 C (GPI-PLC), which is expressed exclusively in bloodstream forms via 3' UTR regulatory
354 elements, and two paralogous calpain-related proteins, CAP5.5V (bloodstream-specific) and
355 CAP5.5 (procyclic-specific) (7, 8, 45). These genes provide ideal test cases for assessing the
356 system's ability to maintain post-transcriptional regulatory mechanisms during *in vitro*
357 differentiation.
358
359 **GPI-PLC example:** GPI-PLC (Tb927.2.6000) is expressed exclusively in bloodstream forms via
360 regulatory elements in its 3' UTR, and it's enzymatic activity is essential for VSG shedding during
361 the early stages of differentiation to procyclic forms (45, 46). To track its expression during
362 bloodstream to procyclic form differentiation, we generated C-terminal mNG::3xTy fusions (GPI-
363 PLC::mT2A::3xTy) using the pRExT2A system, under hygromycin selection (**Fig. 6A)**. Live
364 imaging showed GPI-PLC::mT2A::3xTy predominantly localised to the flagella membrane (**Fig.**
365 **6B**), consistent with native GPI-PLC antibody staining (47) and GPI-PLC-eYFP localisation (48).
366 This pattern differs from the dual flagella/cell body distribution seen with pPOT tagging where the
367 3'-UTR was replaced (25). Western blot analysis using validated anti-GPI-PLC antibodies (47),
368 showed expression of both native GPI-PLC (~41 kDa) and GPI-PLC::mT2A::3xTy (G::mT2A; ~73
369 kDa) specifically in bloodstream form cells (**Fig 6C, left**). The stoichiometric ratio between tagged
370 and wild-type alleles indicated single-copy integration in the diploid genome, and efficient T2A
371 cleavage was confirmed by detection of separate 3xTy::HYG product (42 kDa) in anti-Ty Western
372 blots (**Fig 6C, right**). Although the tagged and untagged alleles in the G::mT2A cell line showed
373 lower signal intensity with the anti-GPI-PLC antibody compared to the parental cell line (22, 48),
374 the tagged protein maintained correct developmental regulation and localisation. Importantly, flow
375 cytometry showed that GPI-PLC::mT2A expression dropped to near-background levels after
376 differentiation to procyclic forms (**Fig 6D, 6E**), indicating that the pRExT2A system preserves the
377 native UTR-mediated regulation of GPI-PLC expression. The distinct localisation patterns
378 between our UTR-preserved and UTR-replaced constructs (25) highlight the importance of
379 maintaining endogenous UTRs for proper protein expression.
380
381 **CAP5.5V and CAP5.5 examples:** CAP5.5V (Tb927.8.8330) and CAP5.5 (Tb927.4.3950) play
382 crucial roles in maintaining cell morphology by organising the subpellicular cytoskeleton in

383 bloodstream and procyclic forms, respectively (8). In the TrypTag project, CAP5.5V showed no
384 expression in procyclic forms when tagged at the N-terminus, but C-terminal tagging, which
385 replaced the native 3' UTR, led to inappropriate expression in procyclic forms, demonstrating the
386 critical role of UTR-mediated regulation (49). To prevent such dysregulation, we preserved the
387 native UTRs while tagging CAP5.5V and CAP5.5 at their C-termini using the pRExT2A system
388 (**Fig. 7A)**.
389
390 Live-cell microscopy confirmed the characteristic cytoskeletal localisation of CAP5.5V across the
391 cell body (excluding the flagellum) in bloodstream cells (**Fig. 7B**), while CAP5.5 expression was
392 not detectable above background, consistent with its known stage-specific expression. Anti-Ty
393 immunoblots revealed stoichiometric expression of the tagged proteins: CAP5.5V::3xTy::T2A
394 (104 kDa), CAP5.5V::mNG::3xTy::T2A (131 kDa), CAP5.5::3xTy::T2A (100 kDa), and
395 CAP5.5::mNG::3xTy::T2A (127 kDa), as well as the cleaved 3xTy::PAC product (~25 kDa),
396 indicating efficient T2A processing (**Fig. 7C**). After differentiation to procyclic forms, flow
397 cytometry showed the expected stage-specific regulation: CAP5.5V fluorescence decreased,
398 while CAP5.5 signal increased (**Fig. 7D** and **7E**), corresponding to the previously reported 19-fold
399 upregulation of CAP5.5V mRNA in bloodstream forms and 5-fold enrichment of CAP5.5 mRNA in
400 procyclic forms (8). This differential expression of the two paralogs before and after differentiation
401 validates the ability of the pRExT2A system to preserve complex stage-specific gene expression
402 patterns mediated by endogenous UTRs.
403
404 **Discussion**
405
406 Kinetoplastids such as *T. brucei* serve as powerful models for understanding fundamental
407 biological processes, particularly the diversity and evolution of eukaryotic gene control (1, 3). The
408 genetic toolbox for *T. brucei* is extensive, with existing tools for genome-wide conditional
409 expression and localisation screens (24, 50, 51). However, efficiently tagging proteins in their
410 native regulatory context, with positive selection for high throughput studies, remains challenging.
411
412 Our new CRISPR/Cas9-T2A system using the pRExT2A plasmids adds to this toolbox, enabling
413 precise, efficient, scalable, stage-specific expression while retaining native mRNA processing
414 signals, specifically native spliced leader addition sites (SLAS), native polyadenylation sites
415 (PAS), native untranslated regions (UTRs), and any regulatory elements contained within these
416 regions (1, 19). Importantly, the tagged proteins tested in this system retain correct localisation

417 and differential expression in the tested life cycle stages. We demonstrate the versatility of this
418 system by tagging proteins involved in key pathways, including GPI-PLC - which is critical for
419 membrane lipid remodelling (52), CAP5.5V and CAP5.5, which are involved in cytoskeletal
420 organisation (7, 8), and ESAG7, a subunit of the transferrin receptor responsible for iron uptake
421 (53). The high efficiency allows for tagging up to 10 proteins per week in bloodstream form
422 parasites, as the protocols for generating repair templates are well-established (25). A beginner's
423 guide for using this toolkit is outlined in **Fig. 8**.
424
425 A key innovation of our system is the dual functionality of the drug resistance marker as both a
426 selection tool and an expression control (**Fig. 5**). This feature allows users to distinguish between
427 transcript abundance/translation efficiency and post-translational events by comparing the levels
428 of the resistance marker to the tagged protein of interest. This approach has proven valuable in
429 related kinetoplastids, such as *L. mexicana*, where T2A-based tagging showed that life cycle-
430 dependent expression of LmxGT1 is controlled by changes in protein stability rather than mRNA
431 levels (32). In contrast, in *L. donovani*, T2A applications showed that upregulation of LdNT3 is
432 driven by increased transcript stability and translational efficiency, not changes in protein turnover
433 (34).
434
435 Using this system in *T. brucei*, we can systematically study proteins involved in nutrient transport
436 and their regulation across different stages of the life cycle. For example, this system can enable
437 investigation into the multi-layered control of iron response genes by examining multiple proteins
438 identified from transcriptomics and proteomics screens (10). This approach can help distinguish
439 between transcript abundance, translation efficiency, and post-translational regulatory events
440 (**Fig. 5D**), while characterising iron response elements (IREs) in a high-throughput format.
441 Furthermore, combining our system with single-molecule RNA FISH enables simultaneous
442 analysis of RNA processing dynamics, offering insights into gene regulation at all levels in a single
443 experiment.
444
445 Another advantage of this system is its compatibility with the pPOT platform, enabling seamless
446 integration into existing workflows and allowing for direct comparisons with protein expression
447 levels where exogenous UTRs from the pPOT system serve as controls (25). For instance, we
448 show that tagging ESAG7 by swapping UTRs diminished native regulation and function (**Fig. 4).**
449 The modular architecture of the pRExT2A offers flexibility for easy swapping of tags, drug

450 resistance markers, or regulatory elements without the need for extensive re-cloning. This
451 adaptability facilitates rapid development of new constructs for diverse experimental applications.
452
453 A key strength of this system is its ability to track stage-specific protein expression before and
454 after differentiation, as demonstrated for the dynamic regulation of cytoskeletal proteins CAP5.5V,
455 and CAP5.5. By tagging multiple components with different fluorophores, we can visualise
456 complex assembly and disassembly in real-time during differentiation, revealing both the timing
457 of protein interactions and factors affecting complex formation. For example, tagging components
458 of mitochondrial respiratory complexes alongside iron-sulfur cluster assembly proteins could allow
459 tracking of respiratory complex assembly relative to transferrin receptor (TfR) downregulation.
460 Such experiments could reveal how iron is mobilised from storage proteins to newly synthesised
461 mitochondrial complexes and how iron availability affects these metabolic transitions.
462
463 Our T2A system has several technical considerations. First, CRISPR-based integration leaves a
464 small (20 bp) scar sequence that could theoretically affect gene expression, though we observed
465 no impacts during validation experiments. To address this potential limitation, we provide pre-
466 designed primers for scarless integration, although these remain to be tested. Second, genome-
467 wide analysis of PAM sequence requirements in 9,666 *T. brucei* genes shows that 87.6% have
468 suitable N-terminal PAM sites and 79.3% have C-terminal sites. Given the GC content of *T. brucei*
469 TREU927 (45.47%) and Lister427-2018 (43.67%), NGG PAM sites theoretically occur every 8-10
470 bp, providing multiple targeting options near gene termini (54, 55). Switching termini usually
471 resolves targeting challenges when optimal sites are unavailable. Third, the system predominantly
472 tags one gene copy, which can be challenging when targeting genes within multigene arrays that
473 share identical UTRs, such as *Hsp70*. For such cases, specific targeting can be achieved by
474 manually designing primers to regions that are not identical within the UTRs.
475
476 Fourth, when tagging developmentally regulated genes, users should evaluate available mRNA
477 and protein data to guide experimental design. Our data with CAP5.5 demonstrates successful
478 tagging despite rapid protein turnover in bloodstream forms ($t_{1/2}$ = 2.3 h versus median $t_{1/2}$ = 5.6
479 h) (56). The relatively stable mRNA half-life (33 min vs. median BSF mRNA $t_{1/2}$ of 13 min) and
480 detectable ribosome association enable sufficient expression for drug selection (57, 58).
481 However, genes with highly unstable mRNAs in the chosen life cycle stage might be challenging
482 to tag, as drug resistance and protein expression are controlled by the same regulatory elements.
483 We recommend consulting published mRNA stability and ribosome-profiling data when designing

484 T2A tagging strategies for stage-specific genes (2, 57, 58), keeping in mind that some genes may
485 be successfully tagged despite low predicted expression.
486
487 For protein detection, the shared 3xTy epitope tag between the drug resistance marker and target
488 protein creates specific constraints. While Western blotting can distinguish these proteins by size,
489 immunofluorescence cannot separate their signals. We recommend either live imaging or using
490 anti-mNeonGreen antibodies for localisation studies and choosing drug resistance markers of a
491 different molecular weight to your target protein. Additionally, the limited dynamic range of
492 mNeonGreen may not detect subtle expression changes, such as responses to nutrient
493 starvation. To overcome this, we are developing luciferase-based vectors that offer enhanced
494 sensitivity for monitoring dynamic responses to environmental signals.
495
496 Our T2A system has the potential to offer new insights into the complex biology of these parasites
497 and contribute to our broader understanding of eukaryotic gene regulation, particularly in systems
498 where post-transcriptional control plays a dominant role. By preserving native regulatory contexts,
499 this approach could accelerate our understanding of molecular mechanisms underlying parasite
500 adaptation and survival.
501
502 **Materials and methods**
503
504 ***T. brucei* cell culture and generation of constitutive Cas9 expressing cells**
505 A tetracycline inducible single-marker bloodstream form Lister 427 strain *T. brucei brucei* was
506 used for all experiments (59). The cells were transfected with a humanised *Streptococcus*
507 *pyogenes* Cas9 plasmid (pTB011) that integrates upstream of the first β-tubulin gene in the tubulin
508 array and is under Pol II transcription (37). Cas9 expression was confirmed by Western blot using
509 anti-Cas9 antibody (Abcam). The cells were cultured at 37°C, 5% CO_2 in HMI-9 medium
510 supplemented with 20% heat inactivated Foetal Bovine Serum, 100 U/ml penicillin and 100 µg/ml
511 streptomycin (Gibco). Cell lines were propagated with the appropriate selection drugs where
512 necessary: 2.5 µg/ml G418 (parental) and 10 µg/ml Blasticidin for Cas-expressing cells and 0.2
513 µg/ml Puromycin for reporter or epitope-tagged cell lines.
514
515 **Construction of T2A peptide-based plasmid templates for endogenous tagging**
516 The "donor vectors" for T2A peptide-based tagging were originally designed as "SfiI cassettes" to
517 fit into our multi-fragment assembly method for generating targeting constructs for integration via

518 homologous recombination (60). The PAC::T2A::mNG and mNG::T2A::PAC tagging cassettes
519 were synthesised in full by GenScript. The PAC and mNeon genes were encoded with a
520 *Leishmania donovani* codon bias (i.e., predominantly G or C in the wobble position), and a nine
521 amino acid flexible linker was included upstream of the T2A peptide sequence to facilitate efficient
522 co-translational cleavage (38) The mScarlet (61) gene was PCR amplified from
523 pCytERM_mScarlet_N1 (Addgene_85066) and inserted into N-terminal and C-terminal "Empty
524 T2A::Drug^R cassettes as described in the **Fig. S1**. All pRExT2A plasmids were synthesised by
525 Genscript using sequences from the pPOT series which have been optimised for *T. brucei* (25).
526
527 **sgRNA and repair template design and assembly**
528 All oligonucleotides for generation of donor DNA repair templates and for analyses of edited
529 genomic loci were synthesised by Life Technologies (Invitrogen, UK) and presented below or in
530 the supplementary material. sgRNA primers targeting specific genomic loci were selected by
531 choosing a 20 bp sequence adjacent to a Protospacer Adjacent Motif (PAM, 5'-NGG) site that is
532 either close to the start codon (for 3' targeting) or close to the STOP codon (for 5' targeting).
533 SgRNA primers were cross-checked for any off-target binding. PCR amplification of all sgRNAs
534 followed the protocol published Beneke et al using 2 μM of primer G00 (sgRNA scaffold) as
535 template and a gene-specific sgRNA primer (37). The resulting sgRNA PCR product and
536 matching PCR-amplified repair template were pooled and transfected into cells that express T7
537 polymerase for T7 RNAP-driven transcription (59).
538
539 The repair template primers were designed to amplify the fluorescent TAG::T2A::drug^R with 25 -
540 40 bp homology arms targeting the gene of interest and the adjacent UTR (Fig 1). The homology
541 arms were selected such that full length endogenous 5' and 3' UTR sequences remained intact
542 following repair. The STOP and START codons were omitted for either C-terminus or N-terminus
543 tagging, respectively. To disrupt the PAM sequence or sgRNA binding regions within the repair
544 templates synonymous mutations or an *Xba*I site was introduced in the repair template primer,
545 highlighted in bold in the primer sequences (26). For PCR amplification of repair templates, 30 ng
546 of plasmid was used with gene-specific forward and reverse primers (**Table 2**).
547
548 **PCR of repair templates and analyses of genomic integrations**
549 All PCR assays for amplification of repair templates were performed in a 50 μl reaction volume
550 using Q5 High-Fidelity 2X MasterMix (New England Biolabs NEB Cat #M0515) according to
551 manufacturer's instructions. The reagent components were assembled on ice and transferred to

552 a preheated thermal cycler at 98.0°C. The annealing temperature for each primer pair was
553 calculated using the NEB Tm calculator online tool. To validate correct integration of tags at the
554 target loci, DNA was extracted from wild type and transfected cells using a DNeasy blood and
555 tissue Kit (Qiagen). All PCR assays for genomic DNA analysis of edited loci were performed in a
556 25 μl reaction volume using Hot Start Taq 2X Master Mix (NEB, Cat #M0496). Five microlitres
557 of all PCR reactions were resolved on a 1% agarose gel to verify the presence of correct size
558 PCR products. For high-throughput applications, we recommend well-established workflows
559 published by Beneke et al 2017 (25).
560
561 **Transfection of pooled PCR-amplified sgRNA and donor repair templates.** 1-3 x 10^7 Lister
562 427 bloodstream cells were harvested and resuspended in 1x TbBSF Roditi buffer premixed with
563 50 µL of pooled PCR reactions (25 μl repair template fragment + 25 μl sgRNA; ~25 μg total DNA
564 each) without purification. The cell suspension in a final volume of 200 μl was electroporated
565 using an Amaxa Nucleofector IIb (Lonza), program X-001. Cells were recovered for 6h at 37
566 ^{0}C/5% CO_2 before selection with appropriate antibiotics and distributed in 24 well plates (2 ml in
567 each well). Positive clones were selected after 5 – 7 days and analysed by flow cytometry and
568 Western blots. Correct integration of repair templates at target loci were confirmed by locus-
569 specific PCR and genotyped by performing whole genome sequencing at
570 BGI Genomics (China). Locus-specific PCR primers are detailed in **Table S1**.
571
572 **Western blotting and antibodies.** Total protein was extracted from wild type and transgenic
573 cells, washed in 1x Phosphate buffered saline, lysed in 1x SDS loading buffer, separated by SDS
574 PAGE, transferred to PVDF membranes and immunoblotted with appropriate antibodies for 1 hr
575 at room temperature: anti-ESAG3 (rabbit, 1:2500), anti-TfR (rabbit, 1:2500), anti-Hsp70 (rabbit,
576 1:5000) (62), anti-mNG (Mouse monoclonal antibody [32F6], ChromoTek 1:1000), anti-RFP
577 (Mouse monoclonal antibody [6G6], ChromoTek 1:1000) and anti-2A (monoclonal clone 3H4,
578 Merck, 1:1000). The membranes were blocked in 5% milk, washed 3x in 1xPSBT after antibody
579 incubation and incubated for 1 hr at room temperature with the appropriate horseradish
580 peroxidase (HRP)-conjugated secondary antibodies (anti-mouse IgG, 1:5000; and anti-rabbit IgG,
581 1:5000 dilution), developed after 1 min incubation with Super Signal ECL reagent (Thermo Fisher
582 Scientific), and imaged using a ChemiDoc Gel Imaging MP System (Bio-Rad).
583
584 **Quantitative real time PCR (qRT-PCR).** All qRT-PCR analyses used ~1–5 x 10^7 cells total, and
585 RNA was extracted from cell pellets using Qiagen RNeasy Extraction Kit. qRT-PCR from total

586 RNA was performed using Luna Universal One-step kit (NEB) per manufacturer's instructions.
587 All reactions were performed in technical duplicates and specificity of each reaction was checked
588 by ensuring single peaks from melt curves. Fold change was calculated by comparative C_T
589 method using the Applied Biosciences software. Relative transcript levels were compared to
590 endogenous control, *ZFP3* (Tb927.3.720, nts 241-301) relative to untreated controls. Primer
591 sequences are available in supplementary material.
592
593 **Heat shock and iron starvation inductions**. Iron starvation was induced for 5 hr at cell densities
594 $5x10^5$ cells/ml with 25 mM Deferoxamine (Sigma Aldrich) in HMI-9 medium. For heat shock,
595 Hsp70 tagged cells were maintained at 37 °C or transferred to 42 °C for one hour in HMI-9 medium
596 at same cell densities £ $5x10^5$ cells/ml. In both cases, cells were harvested and analysed by qRT-
597 PCR, flow cytometry or Western blots.
598
599 **Immunofluorescence microscopy and DAPI staining.** Immunofluorescence assays were
600 performed as previously described in (63). αBiP (Rabbit) antibody dilution was 1:2000, anti-mNG
601 (Mouse monoclonal antibody [32F6], ChromoTek 1:100). Slides were washed with 1x PBS then
602 incubated with secondary antibody for 1 hr at room-temperature: Alexa-488 conjugated a-mouse
603 (1:500) and Alexa-594 conjugated a-rabbit (1:500) (Molecular Probes, USA). Slides were washed
604 in 1x PBS and mounted in VectaShield mounting medium (Vector Laboratories) containing 4′,6-
605 diamidino-2-phenylindole (DAPI), sealed with coverslips and imaged on a Zeiss AxioImager M2
606 with an ORCA Flash 4 camera.
607
608 **Flow cytometry.** Live cells were harvested, washed in 1x PBSG, resuspended in 1 mL 1x PBSG,
609 and analysed with an LSR Fortessa cell analyser using BD FACS Diva software (BD Biosciences).
610 For uptake of Alexa-647 transferrin (Tf-647), the cells were pre-incubated with serum-free HMI9
611 supplemented with 0.5 μg/mL BSA at 37 ^{0}C for 10 minutes. Tf-647 was added at 50 μg/mL and
612 incubated for 30 minutes. The cells were washed twice in 1x PBSG and analysed by flow
613 cytometry. Post-acquisition analysis of flow cytometry data was performed with FlowJo v10
614 software (BD Biosciences).
615
616 **Whole Genome Sequencing.** Genomic DNA was extracted from wild type and Cas9-edited cells
617 using the DNeasy Blood and Tissue Kit (Qiagen). Paired-end whole genome sequencing was
618 performed by BGI Genomics (China) on a DNBseq platform with 100bp read length (PE100) for
619 five tagged cell lines - ESAG3::mNG::T2A, Hsp70::mNG::T2A, ESAG7::mNG::T2A,

620 ESAG3::mNG::Ty, and ESAG7::mNG::Ty, - and wild-type (untagged). Libraries were prepared
621 using a BGI genomics custom short-insert DNA library construction protocol including DNA
622 fragmentation, size selection, end repair and 'A' tailing, adapter ligation, and PCR amplification.
623 Raw reads were filtered using BGI genomics custom pipeline SOAPnuke (64) to remove adapter
624 sequences, contamination and low-quality reads (parameters: -n 0.001 -l 10 -q 0.4 --adaMR 0.25
625 --ada_trim --minReadLen 100), yielding approximately 6.3-6.4 million clean reads per sample
626 (1.26-1.29 GB data). Quality control of raw reads was evaluated and showed consistently high
627 quality across all samples (97.99-98.16% bases ≥Q20 and 43.48-43.73% GC content). Clean
628 reads were mapped to the *T. brucei* Lister427-2018 reference genome using Rsubread. For
629 analysing the repetitive Hsp70 array, we employed systematic stringency controls in our mapping
630 approach: from very high stringency (unique mapping only, no mismatches allowed, no alternate
631 mapping allowed) to low stringency (allowing reads to map to multiple locations with up to 3
632 mismatches). Coverage statistics were calculated using GenomicAlignments in R.
633
634 **Development of an automated primer design tool.** To facilitate use of this tool, we developed
635 code for automatic design of primers. This was implemented as a Python module, using sequence
636 and metadata retrieval from TriTrypDBv68 (65) to allow design of primers from a gene ID or list
637 of gene IDs as found in TriTrypDBv68. For ease of use, we also wrote an iPython/Jupyter
638 Notebook that runs in Google Collaboratory. Anyone can visit this notebook (available at:
639 https://github.com/zephyris/scarlesstagging) and design primers without any need for coding
640 ability or installing software on a computer.
641
642 The automated primer design attempts to design primers which make absolutely no change to
643 the 5` (for N terminal tagging) or 3` (for C terminal tagging) UTRs. Without a CRISPR- mediated
644 double strand break, this could be done by providing homology arms for the homologous repair
645 construct which are, for N terminal tagging, the end of the 5` UTR and start of the ORF or, for C
646 terminal tagging, the end of the ORF (without the stop codon) and the start of the 3` UTR. For
647 CRISPR-assisted homologous recombination, a double strand cut site defined by a protospacer
648 adjacent motif (PAM) site of GG must be selected near the start (for N) or end (for C) of the ORF.
649 However, any such site would also cut the repair construct. Therefore, we search for protospacer
650 adjacent motif (PAM) sites within the ORF which can be modified to remove the PAM without
651 changing the encoded amino acid. This search range is defined by the number of bases typically
652 available for primer synthesis (100 bases) minus the length of the template plasmid annealing
653 site (template dependent, around 18 bases) and the length of homology (25 bases), searching

654 within the start or end of the ORF for N or C terminal tagging respectively. The PAM site closest
655 to the start (for N terminal) or end (for C) of the ORF is selected. The final primer sequences for
656 the generation of the sgDNA are then constructed for the automatically selected PAM site, and
657 the forward and reverse primer sequences for the homologous repair construct constructed for
658 the user-selected template plasmid and automatically selected PAM site. The reverse (for N
659 terminal) and forward (for C) primers contain a single point mutation in the ORF which removes
660 the PAM site and so prevents CRISPR-mediated cutting.
661
662 **Data Availability:** The whole genome sequencing data generated in this study have been
663 deposited in NCBI Sequence Read Archive under BioProject accession PRJNA1202750. Raw
664 reads for all tagged cell lines (ESAG3::mT2A, ESAG3::mTy, ESAG7::mT2A, ESAG7::mTy,
665 HSP70::mT2A) and parental controls are included. These data support the integration analysis
666 presented in Figure S3.
667
668 All pRExT2A plasmid templates for N- and C-terminal tagging described in this study have been
669 deposited at Addgene (https://www.addgene.org/browse/article/28252442/).
670
671 **Code availability:** The automated primer design tool for our tagging system is freely available
672 at https://github.com/zephyris/scarlesstagging, along with pre-designed primers for *T. brucei*
673 (TREU927, Lister 427)) and selected species, available at:
674 https://github.com/zephyris/scarlesstagging/tree/main/examples/py/tritrypdb-genomes-v68.
675
676 **Acknowledgements:** pCytERM_mScarlet_N1 was a gift from Dorus Gadella (Addgene plasmid
677 # 85066; http://n2t.net/addgene:85066 ; RRID:Addgene_85066). The authors are grateful to
678 Professor Jay Bangs (University at Buffalo, USA) for generous anti-TfR and anti-Hsp70
679 antibodies, and Professor Keith Gull, FRS for his insightful discussions and advice on gene choice
680 for developmental regulation and experimental design.
681
682 **Funding:** This work was funded by a Wellcome Trust and Royal Society Sir Henry Dale
683 Fellowship (208780/Z/17/Z) to CT, and a NIAID, National Institutes of Health Grant R03AI137636
684 to PAY. Richard Wheeler is supported by a Wellcome Trust Sir Henry Dale Fellowship
685 (211075/Z/18/Z)
686

687 **References**
688

689 1. Clayton C. Regulation of gene expression in trypanosomatids: living with polycistronic
690 transcription. Open biology. 2019;9(6):190072.

691 2. Fadda A, Ryten M, Droll D, Rojas F, Färber V, Haanstra JR, et al. Transcriptome-wide
692 analysis of trypanosome mRNA decay reveals complex degradation kinetics and suggests a
693 role for co-transcriptional degradation in determining mRNA levels. Mol Microbiol.
694 2014;94(2):307-26.

695 3. Luzak V, Osses E, Danese A, Odendaal C, Stricker SH, Haanstra JR, et al. SLAM-seq
696 reveals independent contributions of RNA processing and stability to gene expression in African
697 trypanosomes. bioRxiv. 2024:2024.06.18.599538.

698 4. Lee MG, Polevere RI, Van der Ploeg LHT. Evidence for segmental gene conversion between
699 a cognate hsp70 gene and the temperature-sensitively transcribed hsp70 genes of
700 *Trypanosoma brucei*. Mole Biochem Parasitol. 1990;41:213-20.

701 5. Häusler T, Clayton C. Post-transcriptional control of hsp70 mRNA in Trypanosoma brucei.
702 Mol Biochem Parasitol. 1996;76(1-2):57-71.

703 6. Droll D, Minia I, Fadda A, Singh A, Stewart M, Queiroz R, et al. Post-transcriptional
704 regulation of the trypanosome heat shock response by a zinc finger protein. PLoS Pathog.
705 2013;9(4):e1003286.

706 7. Hertz-Fowler C, Ersfeld K, Gull K. CAP5.5, a life-cycle-regulated, cytoskeleton-associated
707 protein is a member of a novel family of calpain-related proteins in *Trypanosoma brucei*. Mol
708 Biochem Parasitol. 2001;116:25-34.

709 8. Olego-Fernandez S, Vaughan S, Shaw MK, Gull K, Ginger ML. Cell morphogenesis of
710 Trypanosoma brucei requires the paralogous, differentially expressed calpain-related proteins
711 CAP5.5 and CAP5.5V. Protist. 2009;160(4):576-90.

712 9. Günzl A, Kirkham JK, Nguyen TN, Badjatia N, Park SH. Mono-allelic VSG expression by
713 RNA polymerase I in Trypanosoma brucei: expression site control from both ends? Gene.
714 2015;556(1):68-73.

715 10. Gilabert Carbajo C, Cornell LJ, Madbouly Y, Lai Z, Yates PA, Tinti M, et al. Novel aspects
716 of iron homeostasis in pathogenic bloodstream form Trypanosoma brucei. PLoS Pathog.
717 2021;17(6):e1009696.

718 11. Benz C, Lo W, Fathallah N, Connor-Guscott A, Benns HJ, Urbaniak MD. Dynamic
719 regulation of the Trypanosoma brucei transferrin receptor in response to iron starvation is
720 mediated via the 3'UTR. PLoS one. 2018;13(12):e0206332.

721 12. Hug M, Carruthers VB, Hartmann C, Sherman DS, Cross GAM, Clayton C. A possible role
722 for the 3'-untranslated region in developmental regulation in *Trypanosoma brucei*. Mol Biochem
723 Parasitol. 1993;61:87-96.

724 13. Hehl A, Vassella E, Braun R, Roditi I. A conserved stem-loop structure in the 3'
725 untranslated region of procyclin mRNAs regulates expression in Trypanosoma brucei.
726 Proceedings of the National Academy of Sciences. 1994;91(1):370-4.

727 14. Hotz HR, Hartmann C, Huober K, Hug M, Clayton CE. Mechanisms of developmental
728 regulation in *T. brucei*: a polypyrimidine tract in the 3'-untranslated region of a surface protein
729 mRNA affects RNA abundance and translation. Nucleic Acids Res. 1997;25:3017-26.

730 15. Schürich N, Furger A, Kurath U, Roditi I. Contributions of the procyclin 3' untranslated
731 region and coding region to the regulation of expression in bloodstream forms of *Trypanosoma*
732 *brucei*. Mol Biochem Parasitol. 1997;89:109-21.

733 16. Furger A, Schürch N, Kurath U, Roditi I. Elements in the 3' untranslated region of procyclin
734 mRNA regulate expression in insect forms of Trypanosoma brucei by modulating RNA stability
735 and translation. Mol Cell Biol. 1997;17(8):4372-80.

736 17. Berberof M, Vanhamme L, Tebabi P, Pays A, Jefferies D, Welburn S, et al. The 3'-terminal
737 region of the mRNAs for VSG and procyclin can confer stage specificity to gene expression in
738 *Trypansoma brucei*. EMBO J. 1995;14:2925-34.

739 18. Ridewood S, Ooi CP, Hall B, Trenaman A, Wand NV, Sioutas G, et al. The role of genomic
740 location and flanking 3'UTR in the generation of functional levels of variant surface glycoprotein
741 in Trypanosoma brucei. Mol Microbiol. 2017;106(4):614-34.

742 19. Trenaman A, Tinti M, Wall RJ, Horn D. Post-transcriptional reprogramming by thousands of
743 mRNA untranslated regions in trypanosomes. Nature Communications. 2024;15(1):8113.

744 20. Shen S, Arhin GK, Ullu E, Tschudi C. In vivo epitope tagging of *Trypanosoma brucei* genes
745 using a one step PCR strategy. Mol Biochem Parasitol. 2001;113:171-3.

746 21. Oberholzer M, Morand S, Kunz S, Seebeck T. A vector series for rapid PCR-mediated C-
747 terminal in situ tagging of Trypanosoma brucei genes. Mol Biochem Parasitol. 2006;145(1):117-
748 20.

749 22. Kelly S, Reed J, Kramer S, Ellis L, Webb H, Sunter J, et al. Functional genomics in
750 Trypanosoma brucei: a collection of vectors for the expression of tagged proteins from
751 endogenous and ectopic gene loci. Mol Biochem Parasitol. 2007;154(1):103-9.

752 23. Dean S, Sunter J, Wheeler RJ, Hodkinson I, Gluenz E, Gull K. A toolkit enabling efficient,
753 scalable and reproducible gene tagging in trypanosomatids. Open Biol. 2015;5(1):140197.

754 24. Billington K, Halliday C, Madden R, Dyer P, Barker AR, Moreira-Leite FF, et al. Genome-
755 wide subcellular protein map for the flagellate parasite Trypanosoma brucei. Nat Microbiol.
756 2023;8(3):533-47.

757 25. Beneke T, Madden R, Makin L, Valli J, Sunter J, Gluenz E. A CRISPR Cas9 high-
758 throughput genome editing toolkit for kinetoplastids. Royal Society open science.
759 2017;4(5):170095.

760 26. Kovarova J, Novotna M, Faria J, Rico E, Wallace C, Zoltner M, et al. CRISPR/Cas9-based
761 precision tagging of essential genes in bloodstream form African trypanosomes. Mol Biochem
762 Parasitol. 2022;249:111476.

763 27. Vasquez J-J, Wedel C, Cosentino RO, Siegel TN. Exploiting CRISPR–Cas9 technology to
764 investigate individual histone modifications. Nucleic Acids Res. 2018;46(18):e106-e.

765 28. Donnelly ML, Luke G, Mehrotra A, Li X, Hughes LE, Gani D, et al. Analysis of the
766 aphthovirus 2A/2B polyprotein 'cleavage'mechanism indicates not a proteolytic reaction, but a
767 novel translational effect: a putative ribosomal 'skip'. Journal of General Virology.
768 2001;82(5):1013-25.

769 29. Ryan MD, King AM, Thomas GP. Cleavage of foot-and-mouth disease virus polyprotein is
770 mediated by residues located within a 19 amino acid sequence. Journal of General Virology.
771 1991;72(11):2727-32.

772 30. Garry AL, Martin DR. The 2A Story: The End of the Beginning. In: Ph DMS, editor. Beyond
773 the Blueprint - Decoding the Elegance of Gene Expression. Rijeka: IntechOpen; 2024. p. Ch. 7.

774 31. Heras S, Thomas M, Garcia-Canadas M, Felipe Pd, Garcia-Perez J, Ryan MD, et al. L1Tc
775 non-LTR retrotransposons from Trypanosoma cruzi contain a functional viral-like self-cleaving
776 2A sequence in frame with the active proteins they encode. Cellular and Molecular Life
777 Sciences CMLS. 2006;63:1449-60.

778 32. Rodriguez-Contreras D, Aslan H, Feng X, Tran K, Yates PA, Kamhawi S, et al. Regulation
779 and biological function of a flagellar glucose transporter in Leishmania mexicana: a potential
780 glucose sensor. FASEB journal : official publication of the Federation of American Societies for
781 Experimental Biology. 2015;29(1):11-24.

782 33. Tran KD, Rodriguez-Contreras D, Vieira DP, Yates PA, David L, Beatty W, et al. KHARON1
783 mediates flagellar targeting of a glucose transporter in Leishmania mexicana and is critical for
784 viability of infectious intracellular amastigotes. J Biol Chem. 2013;288(31):22721-33.

785 34. Licon MH, Yates PA. Purine-responsive expression of the Leishmania donovani NT3 purine
786 nucleobase transporter is mediated by a conserved RNA stem–loop. J Biol Chem.
787 2020;295(25):8449-59.

788 35. Niemirowicz GT, Carlevaro G, Campetella O, Bouvier LA, Mucci J. A versatile 2A peptide-
789 based strategy for ectopic expression and endogenous gene tagging in Trypanosoma cruzi.
790 Heliyon. 2024;10(2):e24595.

791 36. Webb H, Burns R, Ellis L, Kimblin N, Carrington M. Developmentally regulated instability of
792 the *GPI-PLC* mRNA is dependent on a short-lived protein factor. Nucleic Acids Res.
793 2005;33:1503-12.

794 37. Beneke T, Madden R, Makin L, Valli J, Sunter J, Gluenz E. A CRISPR Cas9 high-
795 throughput genome editing toolkit for kinetoplastids. R Soc Open Sci. 2017;4(5):170095.

796 38. de Felipe P, Luke GA, Hughes LE, Gani D, Halpin C, Ryan MD. E unum pluribus: multiple
797 proteins from a self-processing polyprotein. Trends Biotechnol. 2006;24(2):68-75.

798 39. Hertz-Fowler C FLM, Quail MA, Becker M, Jackson A, Bason N, Brooks K, Churcher C,
799 Fahkro S, Goodhead I, Heath P, Kartvelishvili M, Mungall K, Harris D, Hauser H, Sanders M,
800 Saunders D, Seeger K, Sharp S, Taylor JE, Walker D, White B, Young R, Cross G A M,
801 Rudenko G, Barry J D, Louis E J, Berriman M. Telomeric expression sites are highly conserved
802 in *Trypanosoma brucei*. PLoS one. 2008;3:e3527.

803 40. Müller LSM, Cosentino RO, Förstner KU, Guizetti J, Wedel C, Kaplan N, et al. Genome
804 organization and DNA accessibility control antigenic variation in trypanosomes. Nature.
805 2018;563(7729):121-5.

806 41. Huang J, Van Der Ploeg LH. Maturation of polycistronic pre-mRNA in Trypanosoma brucei:
807 analysis of trans splicing and poly (A) addition at nascent RNA transcripts from the hsp70 locus.
808 Mol Cell Biol. 1991.

809 42. Brandau S, Dresel A, Clos J. High constitutive levels of heat-shock proteins in human-
810 pathogenic parasites of the genus Leishmania. Biochem J. 1995;310 (Pt 1)(Pt 1):225-32.

811 43. Trevor CE, Gonzalez-Munoz AL, Macleod OJS, Woodcock PG, Rust S, Vaughan TJ, et al.
812 Structure of the trypanosome transferrin receptor reveals mechanisms of ligand recognition and
813 immune evasion. Nat Microbiol. 2019;4(12):2074-81.

814 44. Tiengwe C, Muratore KA, Bangs JD. Surface proteins, ERAD and antigenic variation in
815 Trypanosoma brucei. Cell Microbiol. 2016.

816 45. Webb HR, Burns R, Kimblin N, Ellis L, Carrington M. A novel strategy to identify the location
817 of necessary and sufficient *cis*-acting regulatory elements in trypanosomes. RNA.
818 2005;11:1108-16.

819 46. Ferguson MAJ, Low MG, Cross GAM. Glycosyl-sn-1,2-dimyristlyphosphatidly-inositol is
820 covalently linked to *Trypanosoma brucei* variant surface glycoprotein. J Biol Chem.
821 1985;260:14547-55.

822 47. Hanrahan O, Webb H, O'Byrne R, Brabazon E, Treumann A, Sunter JD, et al. The
823 Glycosylphosphatidylinositol-PLC in Trypanosoma brucei Forms a Linear Array on the Exterior
824 of the Flagellar Membrane Before and After Activation. PLoS pathogens. 2009;5(6):e1000468.

825 48. Sunter J, Webb H, Carrington M. Determinants of GPI-PLC localisation to the flagellum and
826 access to GPI-anchored substrates in trypanosomes. PLoS Pathog. 2013;9(8):e1003566.

827 49. Halliday C, Dean S, Sunter JD, Wheeler RJ. Subcellular protein localisation of
828 Trypanosoma brucei bloodstream form-upregulated proteins maps stage-specific adaptations.
829 Wellcome open research. 2023;8:46.

830 50. Alsford S, Turner DJ, Obado SO, Sanchez-Flores A, Glover L, Berriman M, et al. High-
831 throughput phenotyping using parallel sequencing of RNA interference targets in the African
832 trypanosome. Genome Res. 2011;21(6):915-24.

833 51. Carter M, Gomez S, Gritz S, Larson S, Silva-Herzog E, Kim HS, et al. A Trypanosoma
834 brucei ORFeome-Based Gain-of-Function Library Identifies Genes That Promote Survival
835 during Melarsoprol Treatment. mSphere. 2020;5(5).

836 52. Gruszynski AE, DeMaster A, Hooper NM, Bangs JD. Surface coat remodeling during
837 differentiation of *Trypanosoma brucei*. J Biol Chem. 2003;278:24665-72.

838 53. Steverding D, Stierhof Y-D, Chaudri M, Ligtenberg M, Schell D, Beck-Sickinger AG, et al.
839 ESAG 6 and 7 products of *Trypanosoma brucei* form a transferrin binding protein complex. Eur
840 J Cell Biol. 1994;64:78-87.

841 54. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, et al. The
842 genome of the African trypanosome *Trypanosoma brucei*. Science. 2005;309:416-22.

843 55. Rico E, Jeacock L, Kovářová J, Horn D. Inducible high-efficiency CRISPR-Cas9-targeted
844 gene editing and precision base editing in African trypanosomes. Sci Rep. 2018;8(1):7960.

845 56. Tinti M, Güther MLS, Crozier TW, Lamond AI, Ferguson MA. Proteome turnover in the
846 bloodstream and procyclic forms of Trypanosoma brucei measured by quantitative proteomics.
847 Wellcome Open Research. 2019;4:152.

848 57. Manful T, Fadda A, Clayton C. The role of the 5'-3' exoribonuclease XRNA in
849 transcriptome-wide mRNA degradation. Rna. 2011;17(11):2039-47.

850 58. Vasquez J-J, Hon C-C, Vanselow JT, Schlosser A, Siegel TN. Comparative ribosome
851 profiling reveals extensive translational complexity in different *Trypanosoma brucei* life cycle
852 stages. Dev Cell. 2014;5:3623-37.

853 59. Wirtz E, Leal S, Ochatt C, Cross G. A tightly regulated inducible expression system for
854 conditional gene knockouts and dominant-negative genetics in *Trypanosoma brucei*. Mol
855 Biochem Parasitol. 1999;99:89-101.

856 60. Fulwiler AL, Soysa DR, Ullman B, Yates PA. A rapid, efficient and economical method for
857 generating leishmanial gene targeting constructs. Mol Biochem Parasitol. 2011;175(2):209-12.

858 61. Bindels DS, Haarbosch L, van Weeren L, Postma M, Wiese KE, Mastop M, et al. mScarlet:
859 a bright monomeric red fluorescent protein for cellular imaging. Nat Methods. 2017;14(1):53-6.

860 62. McDowell MA, Ransom DA, Bangs JD. Glycosyl phosphatidylinositol-dependent secretory
861 transport in *Trypanosoma brucei.* Biochem J. 1998;335:681-9.

862 63. Tiengwe C, Bush PJ, Bangs JD. Controlling transferrin receptor trafficking with GPI-valence
863 in bloodstream stage African trypanosomes. PLoS Pathog. 2017;13(5):e1006366.

864 64. Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, et al. SOAPnuke: a MapReduce
865 acceleration-supported software for integrated quality control and preprocessing of high-
866 throughput sequencing data. Gigascience. 2018;7(1):1-6.

867 65. Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, et al.
868 TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res.
869 2010;38(Database issue):D457-62.
870
871

872 **Table 1: pRExT2A tagging plasmid templates for N- and C-terminal tagging**

Name	Tag	DrugR	Configurations
N-terminal tagging templates			
pRExT2A-NT-BmNG	3xTy::mNG	BSD::3xTy	BSD::3xTy::T2A::3xTy::mNG
pRExT2A-NT-BmSc	3xTy::mSc	BSD::3xTy	BSD::3xTy::T2A::3xTy::mSc
pRExT2A-NT-BTy	3xTy	BSD::3xTy	BSD::3xTy::T2A::3xTy
pRExT2A-NT-PmSc	3xTy::mSc	PAC::3xTy	PAC::3xTy::T2A::3xTy::mSc
pRExT2A-NT-PmNG	3xTy::mNG	PAC::3xTy	PAC::3xTy::T2A::3xTy::mNG
pRExT2A-NT-PTy	3xTy	PAC::3xTy	PAC::3xTy::T2A::3xTy
pRExT2A-NT-HmNG	3xTy::mNG	HYG::3xTy	HYG::3xTy::T2A::3xTy::mNG
pRExT2A-NT-HTy	3xTy	HYG::3xTy	HYG::3xTy::T2A::3xTy
C-terminal tagging templates			
pRExT2A-CT-BmNG	mNG::3xTy	BSD::3xTy	mNG::3xTy::T2A::BSD::3xTy
pRExT2A-CT-BmSc	mSc::3xTy	BSD::3xTy	mSc::3xTy::T2A::BSD::3xTy
pRExT2A-CT-BTy	3xTy	BSD::3xTy	3xTy::T2A::BSD::3xTy
pRExT2A-CT-PmSc	mSc::3xTy	PAC::3xTy	mSc::3xTy::T2A::PAC::3xTy
pRExT2A-CT-PmNG	mNG::3xTy	PAC::3xTy	mNG::3xTy::T2A::PAC::3xTy
pRExT2A-CT-PTy	3xTy	PAC::3xTy	3xTy::T2A::PAC::3xTy
pRExT2A-CT-BmSc	mSc::3xTy	HYG::3xTy	mSc::3xTy::T2A::HYG::3xTy
pRExT2A-CT-BTy	3xTy	HYG::3xTy	3xTy::T2A::HYG::3xTy

873

874

875 **Table 2:** Primer sequences used for PCR amplification of sgRNA and DNA repair templates. ReF;
876 forward repair and ReR; reverse repair. Primer binding sites corresponding to regions in the
877 plasmid are shown in italics. Bold indicate synonymous mutations to prevent further Cas9 edits.
878

Primer description	Sequence
G00 sgRNA	aaaagcaccgactcggtgccactttttcaagttgataacggactagccttattttaacttgctatttctagctctaaaac
ESAG3	
3'-sgRNA E3::m2A	gaaattaatacgactcactataggTTAAAGCAAAGAAACTGTAAgttttagagctagaaatagc
E3::m2A-ReF	CGCGGTCAAGGAGAATGACTTGTTTAAAGCAAAGAA**G**CT**C***ggcggtGTGTCCAAGGGC*
E3::m2A-ReR	GCCGCAAGAAAATCTTTAGTTTTGTATC**G**C*gccttcacacaggaaacagctatgac*
E3::mTy-ReF	GAGAATGACTTGTTTAAAGCAAAGAAACTG*ggttctggtagtggttccgg*
E3::mTy-ReR	TTTTCATGTCAACATAGACTTGTGCGGCCA*ccaatttgagagacctgtgc*
3'-sgRNA E3::mTy	gaaattaatacgactcactataggACCAACACGCATGGCCAAGAgttttagagctagaaatagc
Hsp70	
5'-sgRNA 2AmSc::H	gaaattaatacgactcactataggCTTTGAAGGATGACATACGAgttttagagctagaaatagc
2AmSc::H-ReF	GGAATATCTGCCTCTTTGAAGGATG**TCTAGA**gccATGACGGAGTACAAACC
2AmSc::H-ReR	CGATACCAATGGCGCCTTCGTATGTgctacccgagccCTTGTACAGCTCGTCCATGC
3'-sgRNA H::m2A	gaaattaatacgactcactataggATCGTCCGGGCCTAAAGTCGgttttagagctagaaatagc
H::m2A-ReF	CGCTGCGGCATCGTCCGGGCCTAAAGTCGA**A**GA**A**GTTGAC*ggcggtGTGTCCAAGGGC*
H::m2A-ReR	GCAGCACCGGTCCGAAATACACCTGGGAAC*gccttcacacaggaaacagctatgac*
ESAG7	
3'-sgRNA E7::m2A	gaaattaatacgactcactataggAGGCAATACAAAAGTACCTA*gttttagagctagaaatagc*
E7::mTy-ReF	ACGAAGAATTCGAGGCAATACAAAAGTACCTAGGGTCTCTG*ggttctggtagtggttccgg*
E7::mTy-ReR	CAGTATAGTTGAATTCGCCTTTACTCCATT*ccaatttgagagacctgtgc*
E7::m2A-ReF	ACGAAGAATTCGAGGCAATACAAAAGTACCTAGGGTCTCTG*ggcggtGTGTCCAAGGGC*
E7::m2A-ReR	CAGTATAGTTGAATTCGCCTTTACTCCATT*gccttcacacaggaaacagctatgac*

879 **Fig. 1 Strategy for endogenous T2A peptide-based protein tagging**
880 The T2A peptide-based tagging system relies on CRISPR/Cas9-mediated insertion of a repair
881 template either on the 5' end (**A**) or the 3'end (**B**) of the open reading frame of a target gene. The
882 repair template consists of a drug-selectable marker ($Drug^R$) linked by a *Thosea asigna* T2A
883 peptide sequence to a fluorescent protein or an epitope tag (TAG), placed at the N- or C-terminus
884 of a protein of interest (POI). PCR amplification of donor repair templates uses primers containing
885 ~25 - 40 bp of homology arms (orange and purple lines) flanking the $Drug^R$::T2A::TAG cassette.
886 The single guide RNA (sgRNA) directs Cas9 to introduce a double-stranded break adjacent to
887 the protospacer-adjacent motif (PAM), and mutated PAM sites from the repair template prevent
888 further edits. Following correct integration, the T2A peptide causes co-translational cleavage,
889 resulting in expression of both the $Drug^R$ and POI in equal amounts. In both configurations, the
890 sgRNA and repair template primers are designed so that the full-length endogenous 5' and 3'
891 UTR sequences remain intact and drive expression of the $Drug^R$ protein and POI.
892
893 **Fig. 2: High-efficiency T2A cleavage in *T. brucei* and localisation of ESAG3::mNG::T2A**
894 **A.** PCR validation of *ESAG3::mT2A* tagged genomic locus. Diagram shows the *ESAG3* gene
895 locus after insertion of a repair cassette containing mNeonGreen (mNG), T2A peptide sequence
896 (T2A), and puromycin (PAC) resistance gene for selection. PCR was performed using genomic
897 DNA from parental cells (Par) and two transfected ESAG3::mT2A clones (1, 2). Coloured lines
898 indicate positions of PCR primers to verify *ESAG3* locus-specific integration, with corresponding
899 size fragments shown on the right of the agarose gel image. Molecular weight markers (M) are
900 shown on the left. All diagrams are not drawn to scale. Note: The weaker intensity of the 2.5 Kb
901 band reflects the expected reduced PCR efficiency for longer products.
902 **B.** Validation of tagged ESAG3::mT2A protein. Top panel: Diagram shows predicted molecular
903 weights (MW) without (90 kDa) and with (69 kDa) T2A cleavage during translation of
904 *ESAG3::mT2A::PAC* (mT2A) mRNA. Independent cell lines were generated by CRISPR/Cas9
905 mediated fusion of ESAG3 with mNG with 6xTy (ESAG3::mTy, mTy, predicted MW 76 kDa)
906 without incorporating a T2A peptide sequence. The ESAG3::mTy protein served as an internal
907 control for SDS-PAGE migration analysis. Note that ESAG3::mTy was generated using the pPOT
908 method which replaces the endogenous 3' UTR with *PFR2* intergenic region.
909 Bottom panel: Western blot analyses using anti-mNG (αmNG, top) and anti-ESAG3 peptide
910 (αESAG3, bottom) antibodies. Asterisks indicate αESAG3 non-specific cross-reacting
911 polypeptide.

912 **C**. Evaluation of mNG-positive cells by flow cytometry. Parental cells (Par, gray histogram) and
913 ESAG3::mNG::T2A-tagged cells (ESAG3::mT2A, clones 1 and 2, green histograms) are shown.
914 Inset: percentage of mNG-positive cells.
915 **D.** Localisation of ESAG3::mT2A protein visualised by live-cell epifluorescence microscopy
916 (green), co-stained with Hoechst 33342 (magenta; DNA). Scale bar = 5 μM.
917
918 **Fig. 3: Effect of heat shock on T2A-tagged Hsp70 mRNA and protein following heat shock.**
919 **A.** PCR validation of *Hsp70* tagged loci. Top panel shows schematic of the locus after insertion
920 of a mNG::2A::drug repair cassette. Bottom panel shows agarose gel image of PCR products
921 using genomic DNA extracted from parental (Par) cells or clonal cells transfected with repair
922 templates for Hsp70 tagged either at the C-terminus with mNeonGreen (mNG, Hsp70::mT2A) or
923 at the N-terminus with mScarlet (mSc, T2A-mS::Hsp70). Coloured lines indicate positions of PCR
924 primers with corresponding size fragments shown above each lane on the agarose gel image. M:
925 molecular markers, two clones for mNG (1, 2) and mSc (3, 4).
926 **B.** Top panel: Diagram shows the predicted MWs of tagged Hsp70 protein before (122 kDa) and
927 after (101 kDa) T2A-mediated co-translational intra-ribosome skipping. Bottom panel: Western
928 blot analyses of tagged and untagged (WT) Hsp70 protein with native anti-Hsp70 (αHsp70)
929 antibody.
930 **C.** Flow cytometry analysis of tagged Hsp70 cells. Parental cells (Par, gray) and cells in which
931 Hsp70 was either mNG-tagged (left) or mSc-tagged (right). Inset shows percentage of positive
932 fluorescent cells.
933 **D.** Localisation of Hsp70::mT2A protein visualised by live-cell epifluorescence microscopy
934 (green), co-stained with Hoechst 33342 (magenta; DNA). Scale bar = 5 μM.
935 **E.** Regulation of tagged Hsp70 protein expression after heat shock (HS). Cell cultures were
936 maintained at 37 °C (-) or transferred to 42 °C (+) for one hour. Total protein extracts were
937 analysed by western blot with anti-mSc (αmScar), anti-mNG (αmNG) and anti-Hsp70 (αHsp70)
938 antibodies. Two clones are shown for parental (Par) and each tagged cell line. Note: The same
939 membrane was sequentially probed without stripping, starting with αmScar, followed by αmNG,
940 and finally αHsp70.
941 **F.** Regulation of *Hsp70* mRNA expression after heat shock. Cell cultures were maintained at 37
942 °C (Control, Con) or transferred to 42 °C (Heat shock, HS) for one hour and mRNA levels
943 quantified by qRT-PCR using primers specific to endogenous Hsp70. Fold change is calculated
944 relative to 37 °C controls with *ZFP3* as endogenous control. Error bars represent standard
945 deviation from two clonal cell populations with three technical duplicates for each clone.

946

947 **Fig. 4: Endogenous T2A tagging of ESAG7 maintains iron-dependent regulation.**

948 **A.** Schematic shows the modified *ESAG7* (*E7)* and *Hsp70* genomic DNA locus after insertion of
949 the mNG::2A::PAC repair cassette between the ORF and their native 3' UTRs (E7::mT2A and
950 H::mT2A) or insertion of mNG::6xTy with *paraflagellar rod protein 2* intergenic region *(PFR2* IGR;
951 E7::mTy). E7::mT2A and E7::mTy tagging preserve endogenous 5' UTRs. Diagrams not drawn
952 to scale.

953 **B.** qRT-PCR analysis of *mNG* transcript abundance in cells expressing E7::mT2A, E7::mTy, or
954 H::mT2A following treatment with the iron chelator Deferoxamine (DFO). Transcript levels were
955 normalised to *ZFP3* signal and expressed as a ratio to control treatment (iron-replete) values.
956 Data are presented as means ± SD, n = 3 independent experiments from a single clone, with 2
957 technical replicates for each n.

958 **C.** Western blot analysis of E7::mT2A protein levels using α-mNG antibody in different cell lines
959 under iron-replete (DFO, -) and iron-starved (DFO, +) conditions. Ponceau staining indicates
960 protein loading.

961 **D.** Flow cytometry analysis of mNG expression in different cell lines under iron-replete (DFO, -)
962 and iron-starved (DFO, +) conditions. Histograms show fluorescence intensity indicating mNG
963 expression levels.

964 **E.** qRT-PCR analysis of *E7* transcript levels in parental (Par) cells or cells expressing E7::mT2A,
965 E7::mTy, or H::mT2A in iron-starved (DFO, +) conditions expressed as a ratio to control treatment
966 (iron-replete) values and normalised to *ZFP3* signal.

967 **F.** Western blot analysis of total protein using α-TfR antibody to detect total TfR expression in
968 whole cell extracts. TfR is a heterodimer of ESAG6 (E6) and ESAG7 (E7). In parental (Par) and
969 H::mT2A lanes, the predicted MW of endogenous E7 (E7::WT) is 37 kDa, while E6 (E6::WT)
970 migrates as two species: ~44 kDa (un-glycosylated) and >50 kDa (fully glycosylated) – see lanes
971 1 and 2 (par) and Lanes 7 and 8 (H::mT2A). The predicted MW of "cleaved" E7::mT2A, E7::mTy,
972 H::mT2A are 67 kDa, 74 kDa, and 110 kDa respectively (lanes 3, 4, 5, and 6). Par and H::mT2A
973 tagged cells serve as a control for DFO treatment. Ponceau-stained membrane shows loading
974 control. The blot is representative of n = 3 independent experiments from one clone.

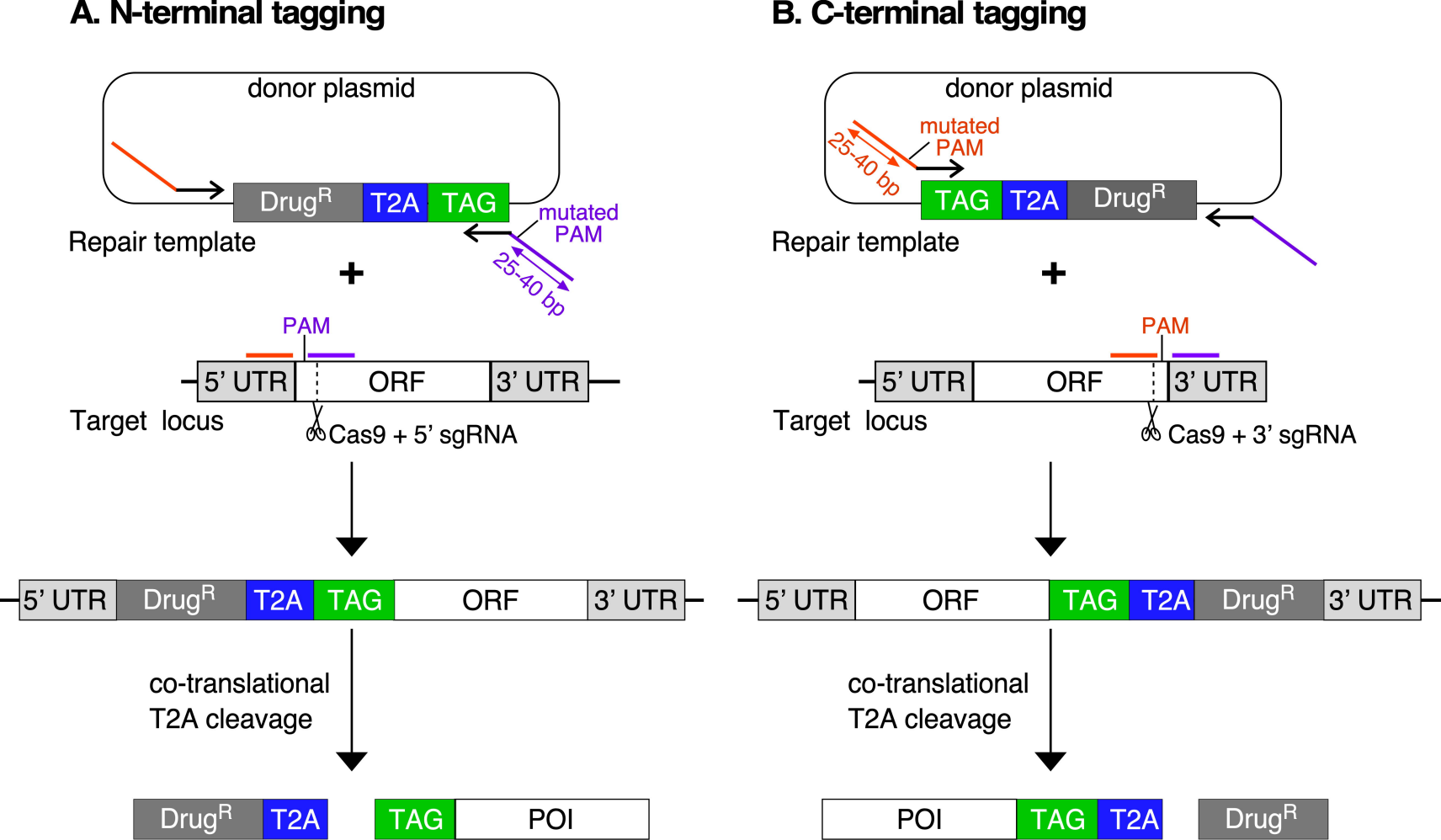
975 **G.** Flow cytometry analysis of transferrin uptake in different cell lines under iron-replete (DFO, -)
976 and iron-starved (DFO, +) conditions. Histograms show fluorescence intensity indicating
977 transferrin uptake levels.

978

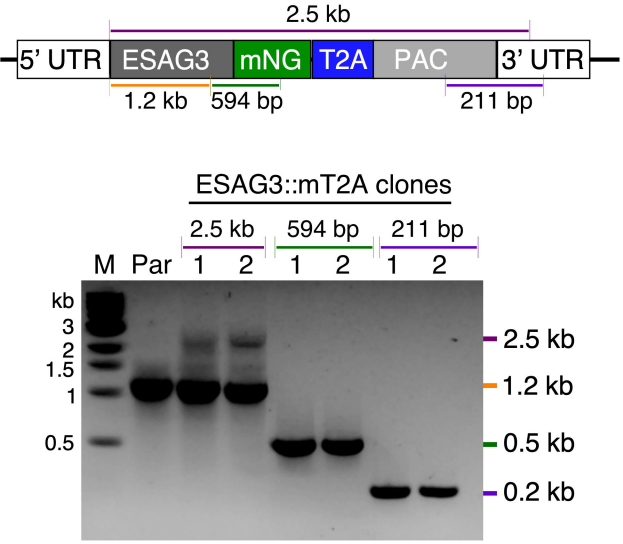
979 **Fig. 5. Map of pRExT2A modular template plasmids and strategy for studying post-**
980 **transcriptional and post translational gene regulation**
981 The schematic illustrates the design of CRISPR/Cas9-mediated modular donor vectors for N-
982 terminal (**A**) and C-terminal (**B**) T2A tagging of target genes in *T. brucei*. The repair cassette
983 includes a drug resistance marker (drug^R, BSD, PAC, HYG), a 3xTy epitope, a T2A peptide, and
984 a fluorescent protein tag (FP), such as mNeonGreen (mNG) or mScarlet (mSc). 3xTy epitope
985 tags, drug^R, and FP are flanked by unique restriction sites (EcoRI, NcoI, HindIII, etc.) to facilitate
986 cloning and customisation. Arrows indicate the positions of forward and reverse primers used for
987 the amplification of repair templates. The forward and reverse primers contain a 20 bp binding
988 site to donor plasmid and 30 bp homology arms matching the target sequence containing a
989 mutated PAM site, as described in Fig.1. The plasmid backbone is pUC57-simple (GenScript),
990 and the full list of sequences is deposited in Addgene.
991 **C.** The process of N-terminal tagging is depicted, where CRISPR/Cas9-mediated insertion of the
992 repair cassette replaces one wild-type (WT) allele with a tagged allele at the 5' end of the target
993 gene. During translation, the T2A sequence facilitates co-translational cleavage, producing two
994 distinct polypeptides: one containing the 3xTy-tagged drug^R marker and another containing the
995 3xTy-tagged FP-tagged protein. This strategy ensures that both the drug^R and FP-tagged proteins
996 are independently stable, allowing for accurate analysis of gene regulatory mechanisms.
997 **D.** The table outlines the different levels of regulation that can be studied using this system. The
998 mRNA level can be analysed using qRT-PCR. At the translation level, protein expression can be
999 monitored using flow cytometry or Western blot to detect 3xTy-tagged FP-tagged protein and the
1000 3xTy-tagged drug^R marker. By measuring relative protein abundance of 3xTy-tagged protein to
1001 the 3x-tagged drug^R marker before and after stimulus exposure using the drugR as an internal
1002 reference, one can distinguish between regulation at the transcript level (where both proteins
1003 change similarly) and differential protein stability (where the target protein changes
1004 independently).
1005
1006 **Fig. 6: Tagging of GPI-PLC preserves life cycle-dependent expression patterns during**
1007 **bloodstream (BF) to procyclic form (PF) differentiation.**
1008 **A.** Schematic of the GPI-PLC protein tagged with mNG::3xTy, fused via the T2A peptide to a
1009 Hygromycin resistance protein (3xTy::HYG) at its C-terminus. The approximate MW (in kDa) of
1010 the tagged and untagged proteins are indicated above the schematics.
1011 **B.** Localisation of GPI-PLC::mT2A protein visualised by live-cell epifluorescence microscopy
1012 (green), co-stained with Hoechst 33342 (magenta; DNA). Scale bar = 5 μM.

1013 **C.** Western blot analysis of GPI-PLC expression in parental (Par) and GPI-PLC::mT2A-tagged
1014 cells in BF and PF cells. The α-GPI-PLC antibody confirms BF-specific GPI-PLC expression
1015 showing stoichiometric expression of endogenous (GPI-PLC) and tagged GPI-PLC (G:mT2A).
1016 Ponceau staining showing a band corresponding to the MW of variant surface glycoprotein (VSG)
1017 in bloodstream but not in procyclic forms. α-Ty antibody (right) detected GPI-PLC::mT2A
1018 (G::mT2A) and 3xTy::HYG (HYG) proteins in BF cells, with signals at the expected size (~73
1019 kDa). No signal was observed in PF cells.
1020 **D.** Flow cytometry analysis of mNG fluorescence in GPI-PLC::mT2A-tagged cells during *in vitro*
1021 differentiation of BF to PF stages. BF cells display high fluorescence, confirming GPI-PLC
1022 expression, whereas PF cells show minimal fluorescence. Parental cells (Par) serve as a negative
1023 control.
1024 **E.** Quantification of flow cytometry data in D, showing median fluorescence intensity for mNG in
1025 BF and PF stages.
1026
1027 **Fig. 7: Tagging of CAP5.5V and CAP5.5 confirms stage-specific expression during**
1028 **bloodstream to procyclic form differentiation.**
1029 **A.** Schematic of CAP5.5V and CAP5.5 proteins tagged with mNG::3xTy::2A::3xTy::PAC or
1030 3xTy::2A::3xTy::PAC at their C-termini. The predicted MW of the tagged protein variants are
1031 indicated.
1032 **B.** Western blot analysis using α-Ty antibody to detect tagged CAP5.5V (left) and CAP5.5 (right)
1033 proteins in bloodstream form (BF) cells. The α-Ty blot shows expression of both mT2A::3xTy and
1034 3xTy – tagged variants and T2A-mediated release of 3xTy-tagged puromycin (3xTy::PAC)
1035 protein. Parental cells (Par) serve as negative control.
1036 **C.** Localisation of CAP5.5V::mT2A (CAP5.5V::mNG) and CAP5.5::mT2A (CAP5.5::mNG) in BF
1037 stages visualised by live-cell epifluorescence microscopy (green), co-stained with Hoechst 33342
1038 (magenta; DNA). CAP5.5 is below detection limit in BF stage cells. Arrowhead indicates no
1039 staining of the flagellum. Scale bar = 5 μM.
1040 **D.** Flow cytometry analysis of mNG fluorescence in CAP5.5V and CAP5.5-tagged cells before
1041 and after differentiation from BF to PF life cycle stages.
1042 **E.** Quantification of flow cytometry results in D, showing median fluorescence intensity for mNG
1043 in BF and PF stages.
1044
1045 **Fig. 8: Beginner's guide to using the pRExT2A tagging system**: Start by selecting your target
1046 gene and deciding whether to tag it at the N- or C-terminus, or both. Next, choose the appropriate

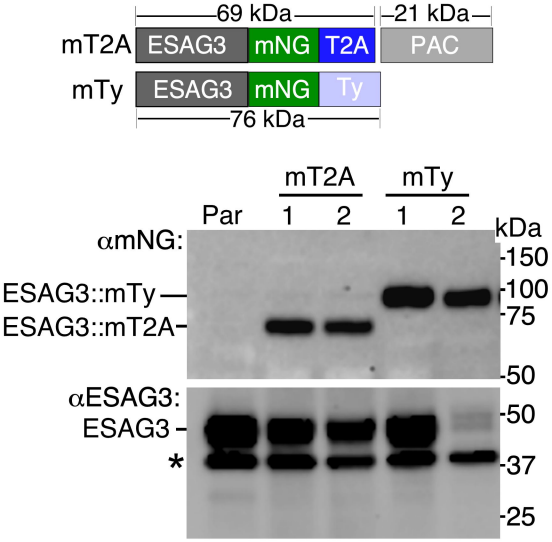
1047 plasmid containing the desired fluorescent or epitope tag, along with a compatible drug-resistance
1048 marker (e.g., Blasticidin or Puromycin). Design primers using the automated tool to generate PCR
1049 amplicons for both the donor repair template and the sgRNA. Once the donor template and
1050 sgRNAs are ready, mix and transfect into *T. brucei* cells expressing Cas9. After transfection, apply
1051 drug selection to isolate successfully tagged cells. Validate the transgenic cells by performing
1052 PCR and Sanger sequencing to confirm correct integration. Finally, verify protein expression and
1053 localisation using Western blot, flow cytometry, or microscopy. Created with BioRender.com.


A. N-terminal tagging
donor plasmid
DrugR
T2A
TAG
mutated PAM
25-40 bp
Repair template
+
PAM
5' UTR
ORF
3' UTR
Target locus
Cas9 + 5' sgRNA
5' UTR
DrugR
T2A
TAG
ORF
3' UTR
co-translational T2A cleavage
DrugR
T2A
TAG
POI
B. C-terminal tagging
donor plasmid
mutated PAM
25-40 bp
TAG
T2A
DrugR
Repair template
+
PAM
5' UTR
ORF
3' UTR
Target locus
Cas9 + 3' sgRNA
5' UTR
ORF
TAG
T2A
DrugR
3' UTR
co-translational T2A cleavage
POI
TAG
T2A
DrugR

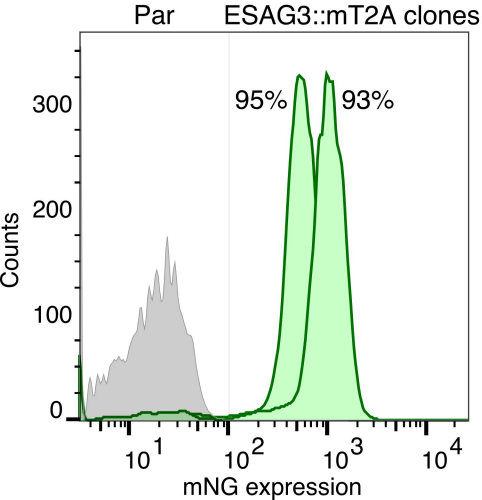
A. gDNA PCR



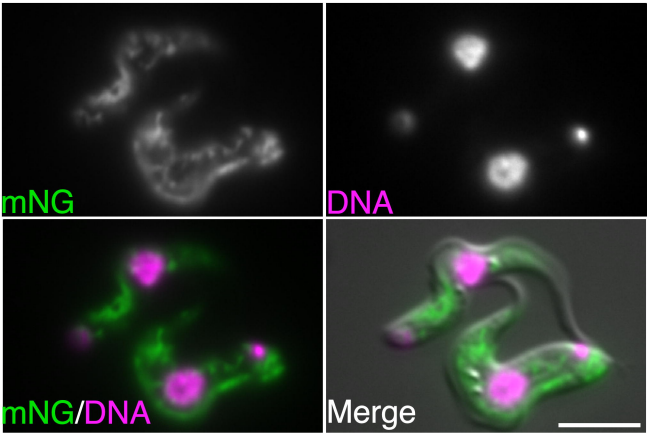
B. Western blot



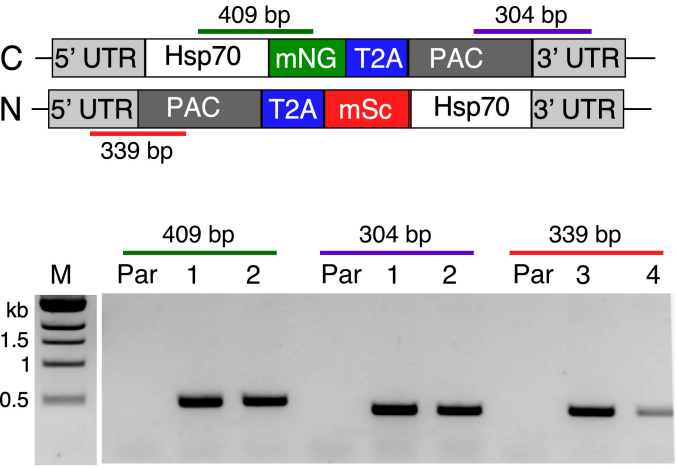
C. Flow cytometry



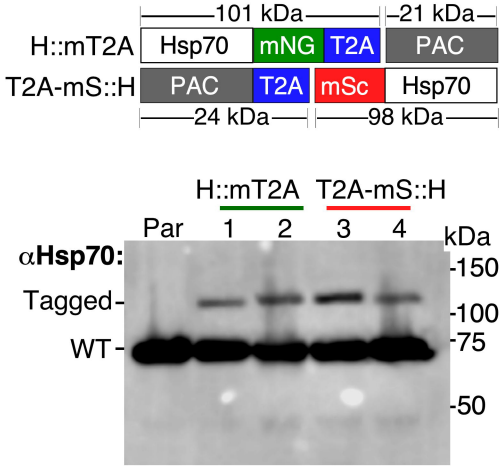
D. Microscopy



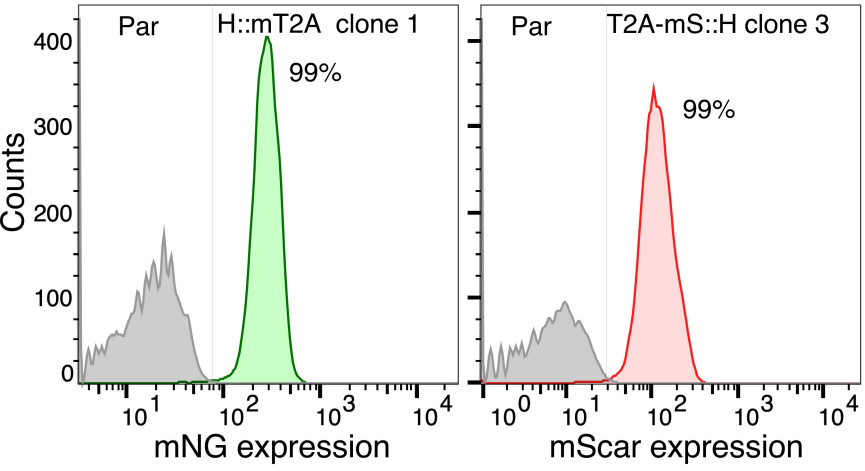
A. gDNA PCR



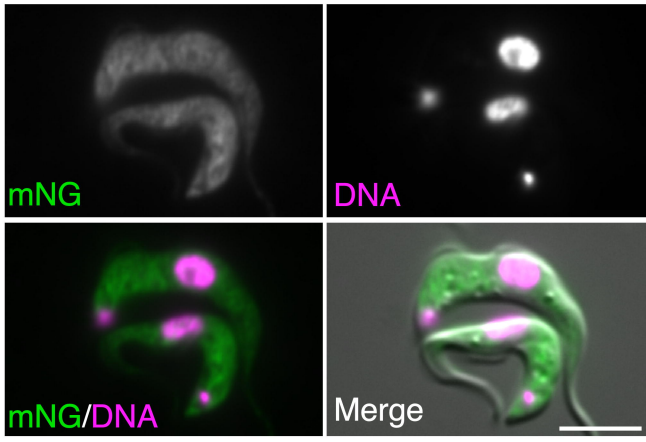
B. Western blot



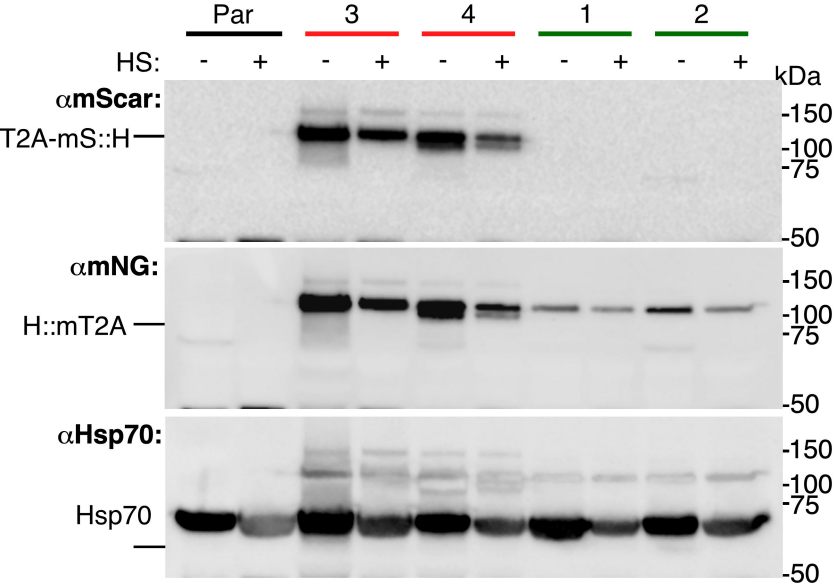
C. Flow cytometry



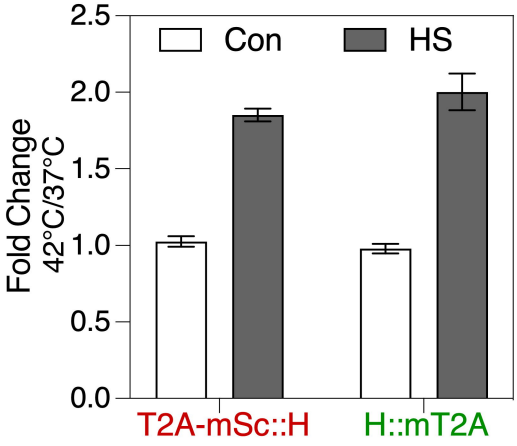
D. Microscopy



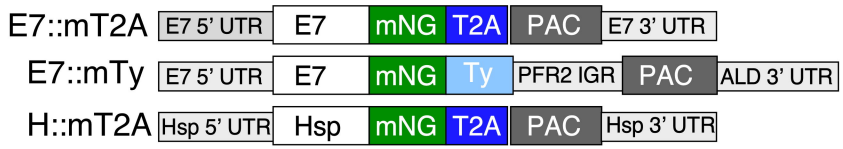
E. Western blot



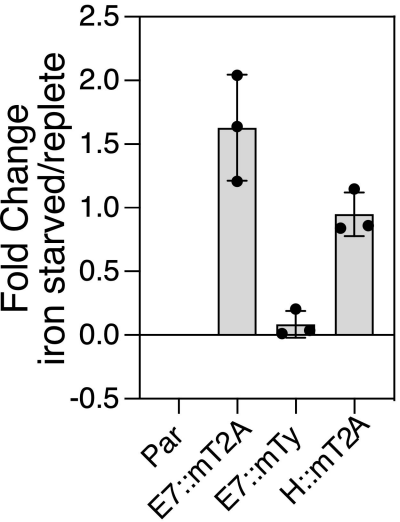
F. qRT-PCR (mRNA)



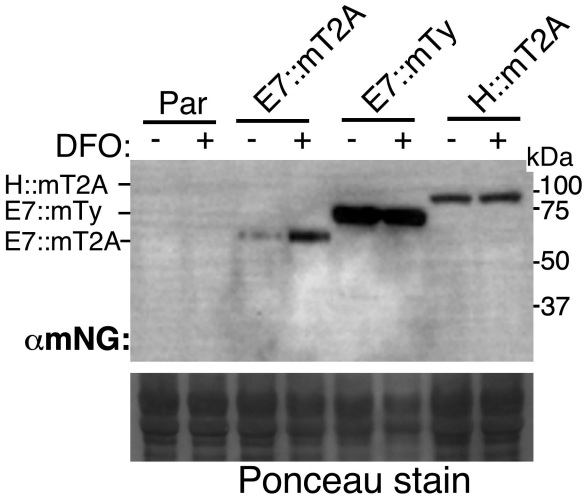
A. Tagged gDNA loci: ESAG7 (E7) and Hsp70 (H)



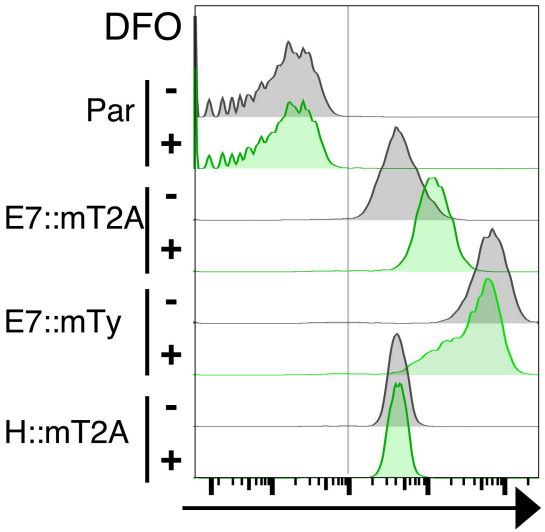
B. qRT-PCR (mNG)



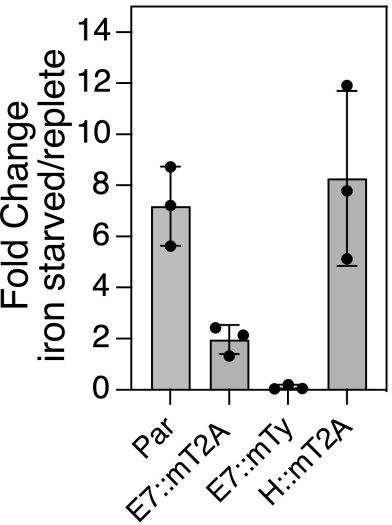
C. α-mNG Western blot



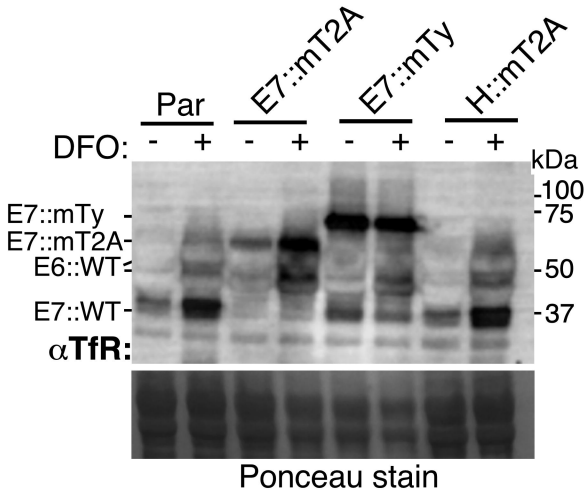
D. mNG expression



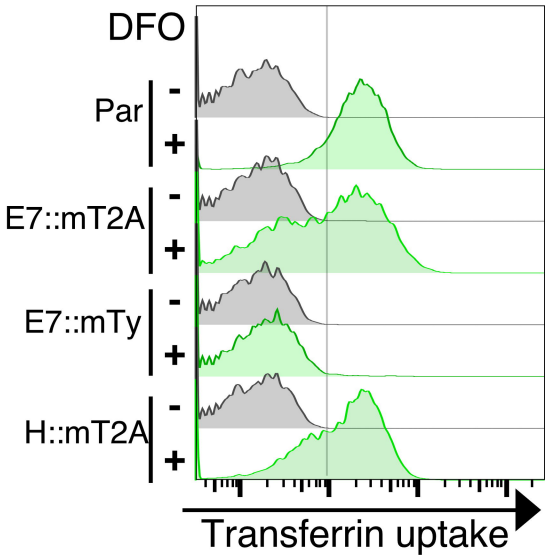
E. qRT-PCR (E7)



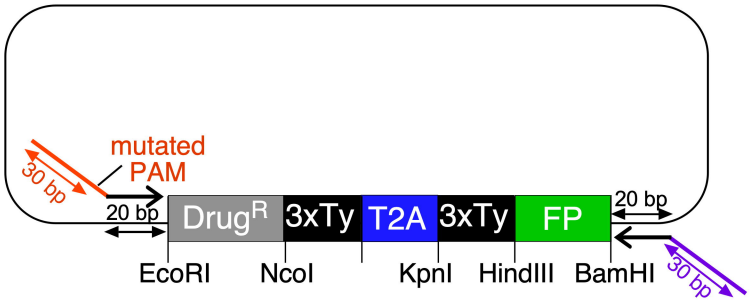
F. α-TfR Western blot



G. Transferrin uptake

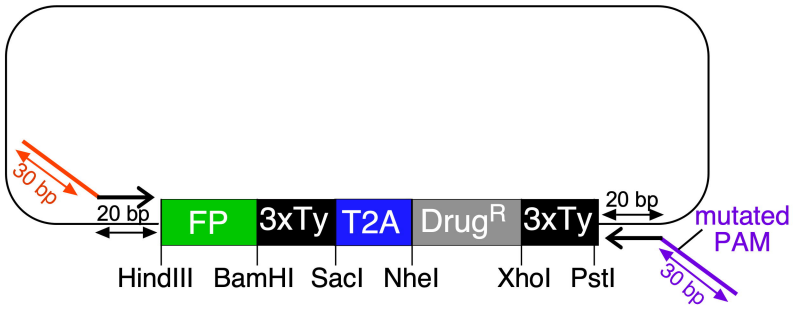


A. N-terminal plasmid templates



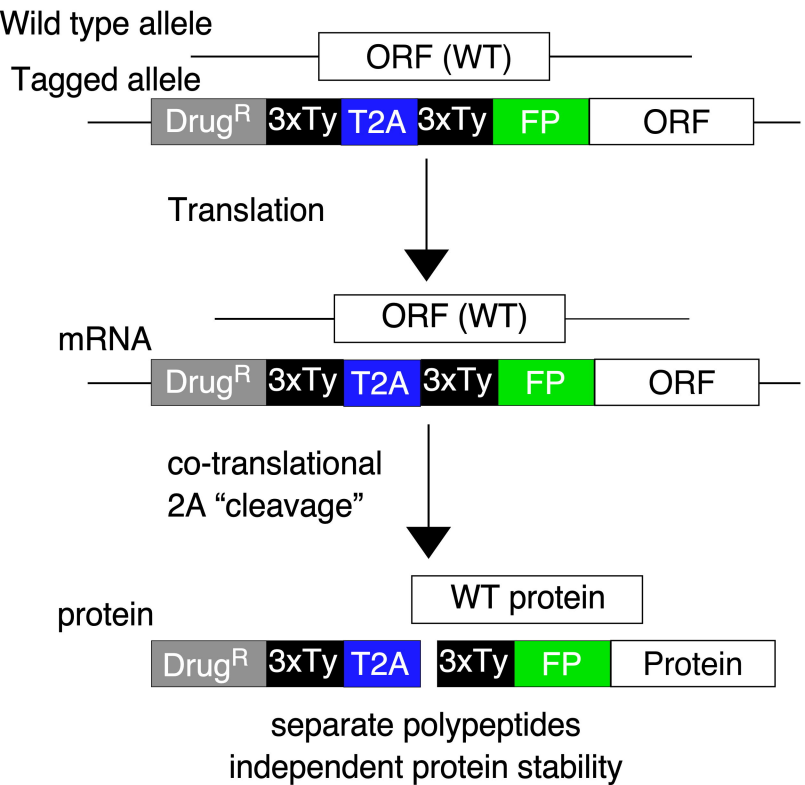
drugR markers	FP
blasticidin (BSD)	mNeonGreen (mNG)
puromycin (PAC)	mScarlet (mSc)
hygromycin (HYG)	3xTy (without FP)

B. C-terminal plasmid templates



FP	drugR markers
mNeonGreen (mNG)	blasticidin (BSD)
mScarlet (mSc)	puromycin (PAC)
3xTy (without FP)	hygromycin (HYG)

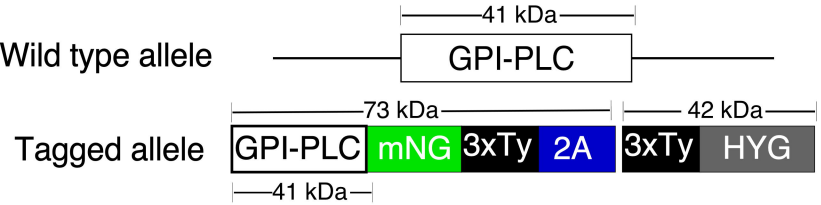
C. N-terminal tagging



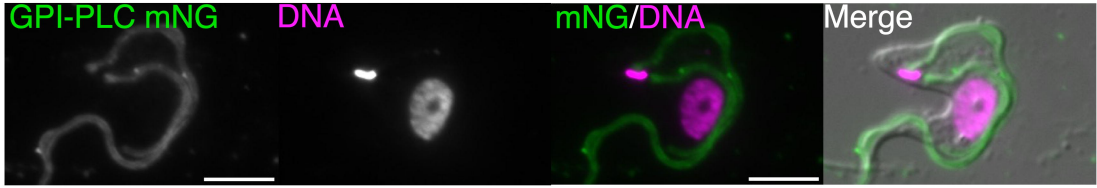
D. mRNA/translation/protein abundance reporter analysis

Level of regulation	Output	Interpretation
mRNA	qRT-PCR	changes affect both proteins equally
Translation	Flow cytometry Western blot	changes affect both proteins equally
Protein abundance	ratio: $\frac{\text{3xTy:protein}}{\text{3xTy::drug}^R}$	Baseline ratio ≈ 1 changes indicate differential stability

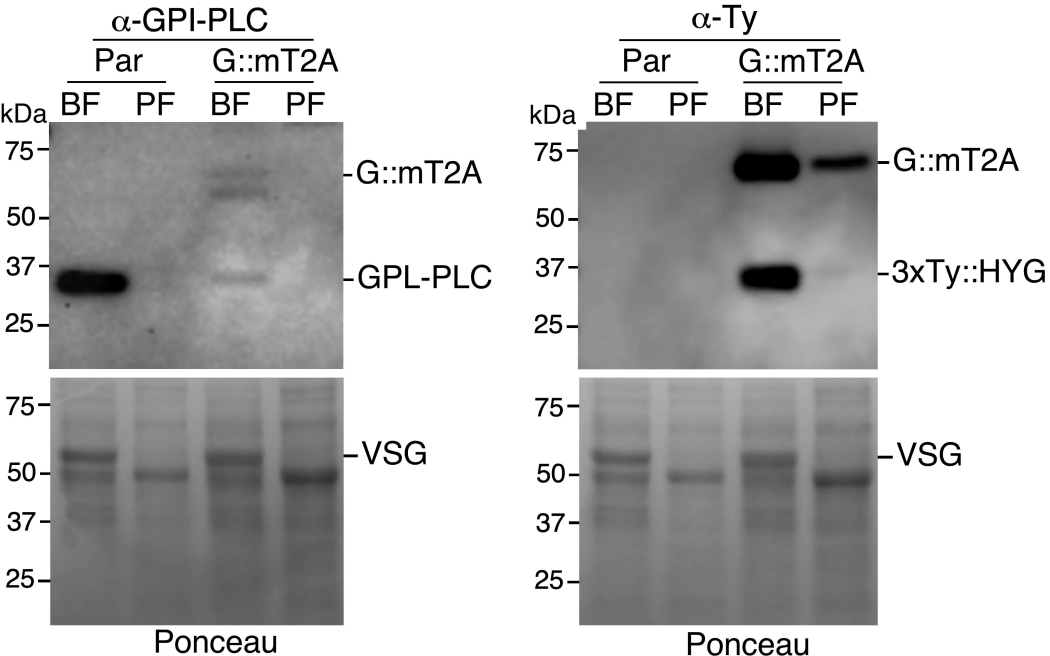
A. Schematic



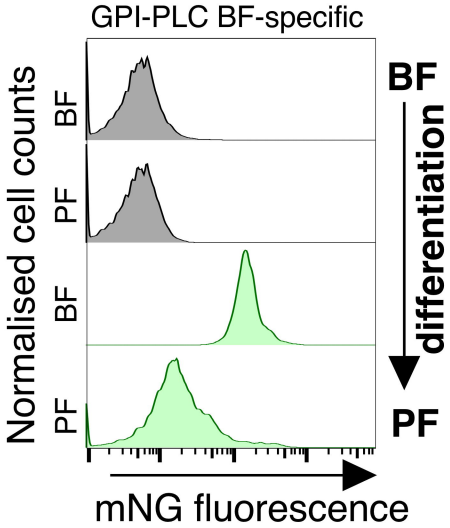
B. Microscopy



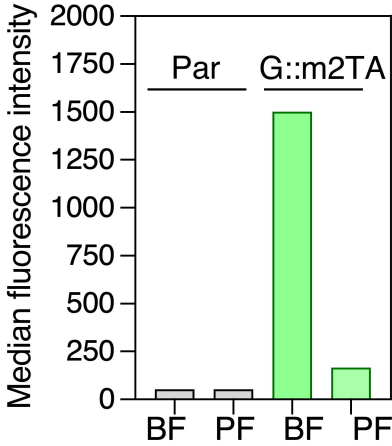
C. Western blot



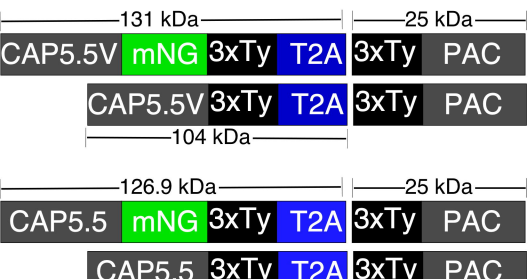
D. Flow cytometry



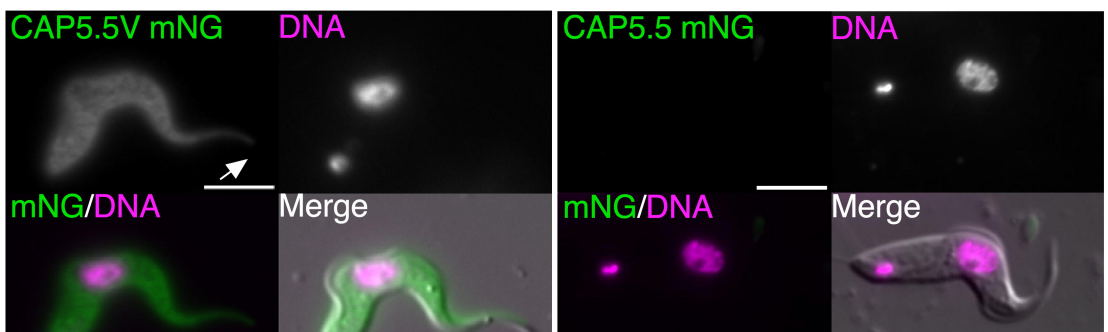
E. Quantification



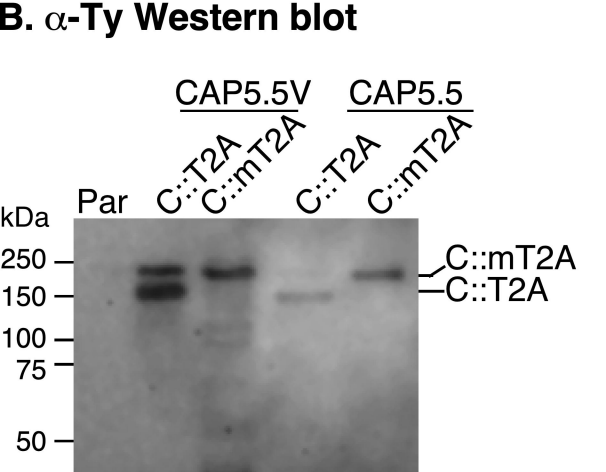
A. Schematic



C. Microscopy

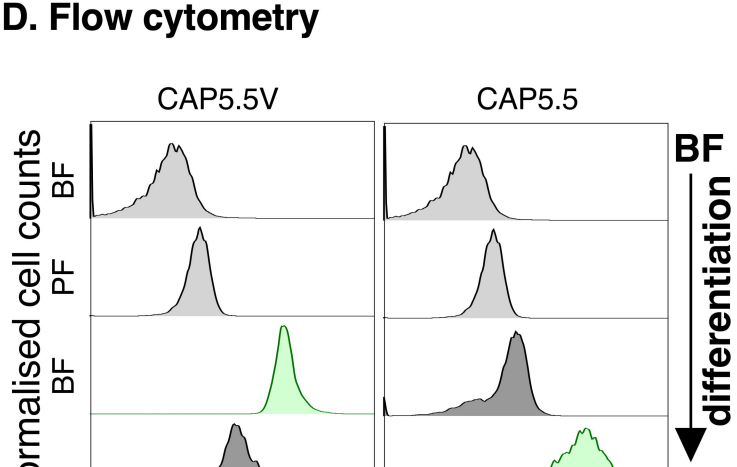


B. α-Ty Western blot

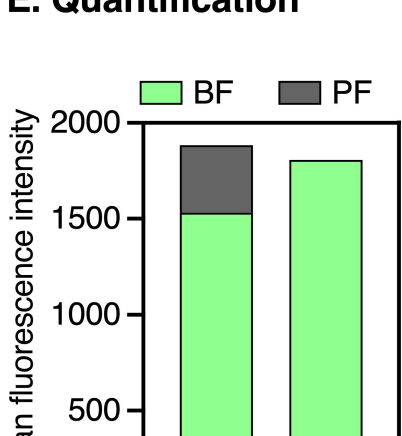


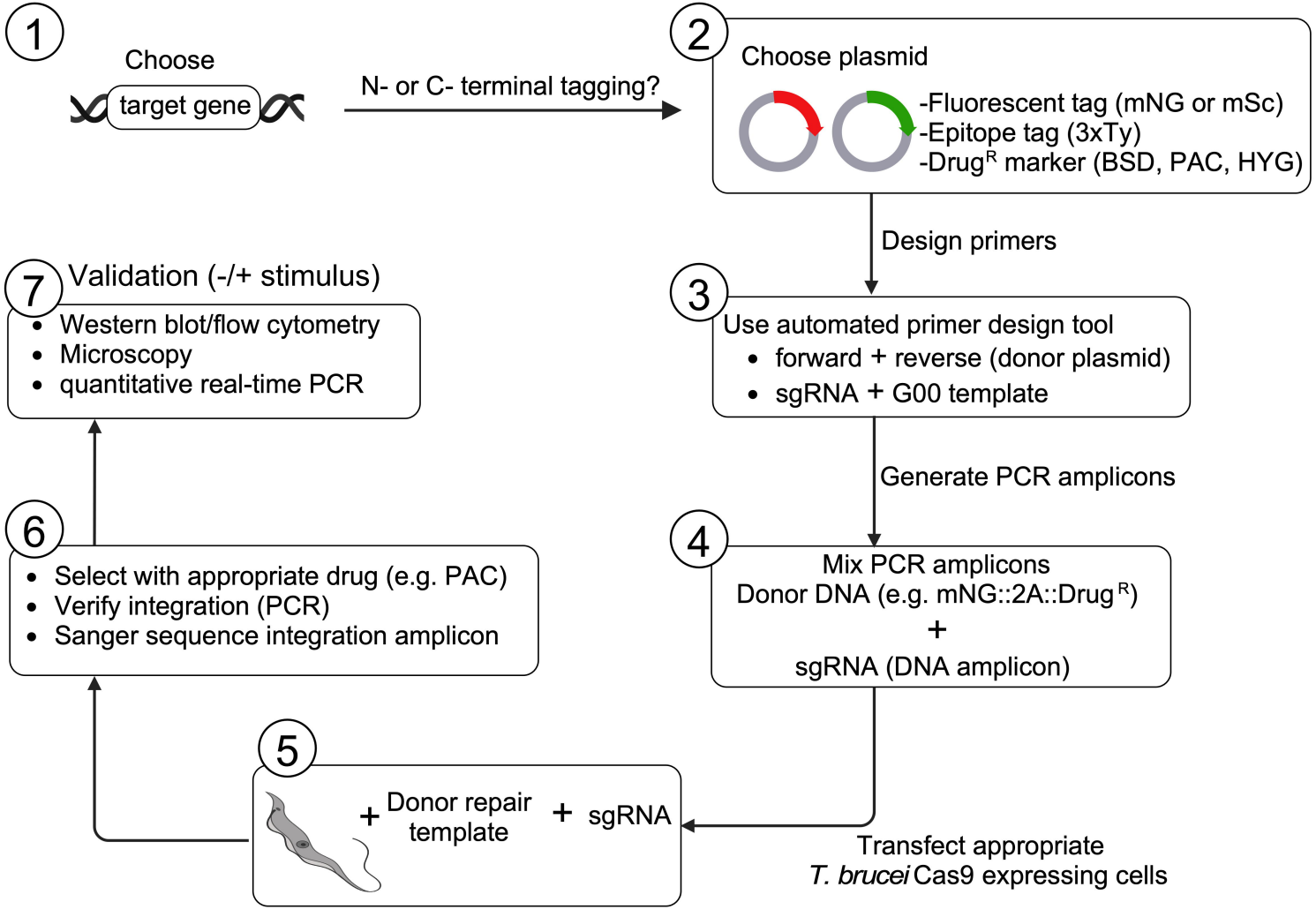
α-Ty blot

D. Flow cytometry



E. Quantification




1
Choose
target gene
N- or C- terminal tagging?
2
Choose plasmid
-Fluorescent tag (mNG or mSc)
-Epitope tag (3xTy)
-Drug^R marker (BSD, PAC, HYG)
Design primers
3
Use automated primer design tool
• forward + reverse (donor plasmid)
• sgRNA + G00 template
Generate PCR amplicons
4
Mix PCR amplicons
Donor DNA (e.g. mNG::2A::Drug^R)
+
sgRNA (DNA amplicon)
Transfect appropriate
T. brucei Cas9 expressing cells
5
+ Donor repair template + sgRNA
6
• Select with appropriate drug (e.g. PAC)
• Verify integration (PCR)
• Sanger sequence integration amplicon
7
Validation (-/+ stimulus)
• Western blot/flow cytometry
• Microscopy
• quantitative real-time PCR