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 high-throughput CRISPR/Cas9-based tagging system that preserves untranslated regulatory elements and uses a viral 2A peptide from *Thosea asigna* (T2A) to co-express a drug-selectable marker and a tagged protein from the same transcript. This dual-function design enables discrimination between transcriptional and post-translational regulation while maintaining native control elements.

We validate the system by tagging six *Trypanosoma brucei* proteins with fluorescent proteins and epitope tags and demonstrate: (i) high-efficiency positive selection and separation of drug-selectable marker and target protein, (ii) preservation of regulated expression responding 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 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 involve VSG and procyclin, major surface proteins in bloodstream and procyclic forms respectively. The VSG mRNA 3' UTR contains a 16-mer sequence that enhances its stability and translation (12, 13), while the procyclin 3' UTR contains a conserved stem-loop structure that regulates its stage-specific expression (14). A recent genome-wide analysis has predicted numerous UTR motifs associated with stage-specific regulation, identifying thousands of 3' UTRs likely to influence mRNA abundance across different life cycle stages (15). This highlights 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 (16, 17), were efficient, while others required cloning gene fragments for homologous recombination (18). PCR-only tagging in *T. brucei* and similar methods in *Leishmania* enabled scalable tagging but disrupt native UTRs when combined with drug-selectable markers (19). The TrypTag project, which tagged nearly 90% of the *T. brucei* proteome, has been a valuable resource for localisation studies, albeit with the recognised caveat that it replaces UTRs (20). CRISPR/Cas9 improves efficiency and precision but often introduces

78 selection markers that interfere with endogenous regulation (21). Two CRISPR-based marker-
79 free methods have been reported which preserve UTRs but require extensive subcloning by
80 limiting dilution (22, 23). While these methods have advanced functional genomics, most still
81 compromise regulatory elements, highlighting the need for tools that preserve UTR integrity while
82 enabling efficient, high-throughput protein tagging and functional analysis.
83
84 First discovered in the Foot-and-Mouth Disease Virus (FMDV), the viral 2A peptide has become
85 a powerful tool for expressing multiple proteins from a single open reading frame (ORF) (24, 25).
86 Several versions of the 2A peptide have been identified across different viruses, each with similar
87 functionality. These include FMDV (F2A), porcine teschovirus-1 (P2A), and *Thosea asigna* virus
88 (T2A) (26). Interestingly, 2A-like sequences (L2A) have also been found in the L1Tc non-LTR
89 retrotransposons of *T. cruzi*, *T. brucei*, *T. vivax* and *T. congolense,* which represents a rare
90 example of these sequences occurring outside of a non-viral context (27). In general, the T2A
91 peptide, with the sequence -EGRGSLLTCGDVEENPG↓P-, facilitates a mechanism known as
92 "ribosomal skipping," where the glycine-proline bond fails to form during translation, resulting in
93 the production of two distinct proteins from the same transcript (24). This process, often called
94 CHYSEL (cis-acting hydrolase element), “stop-and-go”, “self-processing”, or “co-translational
95 cleavage”, will hereafter be referred to as “cleavage” for simplicity, with instances where the
96 process fails being referred to as “uncleaved” proteins.
97
98 In Kinetoplastids, the T2A peptide has been shown to be highly efficient for co-expressing proteins
99 under endogenous UTRs, enabling the study of post-transcriptional regulatory mechanisms.
100 Some examples include the *L. mexicana* glucose transporter (LmxGT1) (28), the cortical
101 cytoskeletal protein (KHARON1) (29) and the nucleotide transporter NT3 in *L. donovani* (30). The
102 T2A system has also been applied for ectopic expression in *T. cruzi*, improving positive selection
103 of transgenic parasites (31). These applications demonstrate the versatility of the T2A peptide
104 system for studying protein localisation and regulatory mechanisms in kinetoplastids.
105
106 In this study, we combined CRISPR/Cas9 editing with the T2A peptide-based system to test its
107 applicability in *T. brucei* by tagging two secretory proteins (ESAG3, ESAG7) which are under
108 transcriptional and post-transcriptional regulation, and cytosolic Hsp70, part of a multi-gene array
109 whose expression is regulated by its 3' UTR in response to heat shock. We demonstrate efficient
110 T2A peptide cleavage and regulation of tagged mRNA and/or proteins in their native loci. We also
111 constructed new vectors with primer binding sites compatible with the plasmids for PCR-only

112 tagging (pPOT series) (19) and designed an automated primer design tool to facilitate high-
113 throughput tagging while preserving native UTRs. We validated the new vectors on three well-
114 characterised proteins: glycosylphosphatidylinositol phospholipase C (GPI-PLC) (32) and two
115 calpain-related proteins (CAP5.5V and CAP5.5) (7, 8), which show stage-specific expression and
116 confirmed that this differential expression was maintained in the tagged versions. Together, we
117 provide a versatile tool that can be used at scale to study post-transcriptional and post-
118 translational regulatory mechanisms in Kinetoplastids.
119
120 **Results**
121
122 **Development of a CRISPR/Cas9-compatible T2A peptide tagging system for *T. brucei***
123
124 Our primary goal was to exploit the high efficiency of CRISPR/Cas9 genome editing to insert a
125 "T2A peptide cassette" between the open reading frame (ORF) of a target gene and its native 5'
126 or 3' UTR in *T. brucei*. The cassette contains a drug-selectable marker (drug^R), a 2A peptide
127 sequence (T2A), and a fluorescent protein or an epitope tag (Tag) (**Fig. 1**). We initially constructed
128 donor plasmids in two configurations: drug^R::T2A::Tag for insertion at the 5' end (**Fig. 1A**), and
129 Tag::T2A::drug^R for insertion at the 3' end of the target gene (**Fig. 1B**). This design allowed for
130 positive selection of transgenic parasites while preserving native mRNA processing signals within
131 the UTRs, thus maintaining natural regulation. Available plasmids with fluorescent tags:
132 mNeonGreen (mNG) or mScarlet (mSc), and drug-selectable markers: blasticidin (BSD), or
133 puromycin (PAC) are shown in **Fig. S1**.
134
135 For 5'-terminal tagging, the donor repair template drug^R::T2A::Tag was amplified by PCR using
136 primers containing 25-40 bp homology arms, targeting a protospacer adjacent motif (PAM) site
137 near the start of the ORF. The ATG start codon was removed, but the 5' UTR remained intact.
138 For 3'-terminal tagging, the Tag::T2A::drug^R donor template was amplified using primers with
139 homology arms to integrate the cassette before the stop codon, targeting a PAM site near the end
140 of the ORF but ensuring the 3' UTR remained intact. In both cases the single-guide RNA (sgRNA)
141 was generated by combining a gene-specific primer encoding a T7 promoter upstream, a 20-bp
142 targeting sequence specifying the target site, and a 3'-end complementary to the sgRNA scaffold,
143 as described in (33).
144
145 Both sgRNA and donor T2A cassettes were amplified by PCR using well-established protocols

146 described by Beneke et al. (21). The PCR products were pooled for transfection into bloodstream
147 form (BF) *T. brucei*, which constitutively express T7 RNA polymerase and Cas9. Cas9 introduces
148 a double strand break at the target site, guided by the sgRNA designed near the PAM site at
149 either the 5' or 3' end of the target gene. We incorporated mutated PAM sites into the repair
150 template to prevent further Cas9 edits post-integration. Once integrated, the T2A peptide
151 sequence mediates co-translational cleavage, allowing for independent expression of the drug
152 resistance marker and the tagged protein of interest in equal amounts (34).
153
154 The donor plasmids described here were originally constructed for *Leishmania*, where we
155 determined that T2A peptide conferred the highest cleavage efficiency among the various 2A
156 peptides tested (**Fig S2**).
157
158 **Test case I: ESAG3 endogenous tagging shows efficient T2A cleavage and robust positive**
159 **selection**
160
161 To assess the efficiency of T2A peptide cleavage in *T. brucei* and demonstrate its utility for protein
162 localisation studies, we targeted the expression site associated gene 3 (ESAG3, gene IDs:
163 Tb427.BES40.10, Tb427.BES40.16). ESAG3 encodes an N-terminal putative ER targeting signal
164 sequence with a predicted cleavage site between Ala22/Leu23 (SignalP v5.0). A two-component
165 PCR amplicon mix was transfected into *T. brucei* T7/Cas9-expressing cells, consisting of a 20-bp
166 sgRNA targeting adjacent to the PAM site near the stop codon and a donor mNG::T2A::PAC
167 cassette with 40/30 bp homology arms. The stop codon was excluded from repair primers to allow
168 in-frame fusion of mNG at the C-terminus, while synonymous mutations in the PAM site prevented
169 further Cas9 edits.
170
171 Seven days post-transfection, genomic DNA PCR confirmed in-frame fusion of *ESAG3* ORF with
172 mNG (expected size 594 bp), *PAC* with *ESAG3* 3'UTR (211 bp), *ESAG3* ORF with its native UTR
173 (2.5 Kb) (**Fig 2A**). No PCR product was generated using genomic DNA from parental cells
174 confirming specificity. To compare the relative mobility of the ESAG3::mT2A protein, we
175 generated independent cell lines with C-terminally mNG-tagged ESAG3 fused with 6xTy
176 (ESAG3::mTy) using the PCR-only tagging method (33) (**Fig 2B**). Western blot analysis revealed
177 that the ESAG3::mT2A protein (69 kDa) migrated faster than ESAG3::mTy (76 kDa), indicative of
178 efficient T2A cleavage. The expected MW of uncleaved ESAG3::mNG::T2A::PAC is 90 kDa.
179 Since there are two copies of ESAG3 in the active expression site (BES1, Tb427_telo40_v2) in

180 the Lister427 cell line (35), native anti-ESAG3 antibody immunoblots confirmed that tagging was
181 single copy for most clones, as the untagged copy was detectable at the expected MW of ~43
182 kDa (**Fig 2B)**. Flow cytometry showed that 93-95% of cells in the drug-selected clonal population
183 were mNG-positive, confirming robust positive selection of correctly tagged cells (**Fig 2C**).
184 Immunofluorescence microscopy showed a typical ER localisation pattern for ESAG3::mT2A,
185 confirming proper targeting to the secretory pathway (**Fig 2D**).
186
187 These results demonstrate precise successful insertion of the mNG::T2A::PAC cassette at the 3'
188 end of *ESAG3*, with highly efficient positive selection of tagged cells, and successful co-
189 translational T2A cleavage in *T. brucei*. Furthermore, this tagging system enabled us to localise
190 ESAG3 in BF trypanosomes, for the first time.
191
192 **Test case II: Endogenous T2A tagging of Hsp70 retains mRNA-level heat shock regulation**
193
194 To further demonstrate the robustness of the T2A system, we targeted the cytosolic heat shock
195 protein Hsp70. In *T. brucei*, Hsp70 is encoded by a tandem array of eight nearly identical genes
196 (IDs: Tb427_110128800, Tb427_110128900 … Tb427_110129500) on chromosome 11 (36). In
197 the Tb927 genome assembly, this array is collapsed into a single gene (Tb927.11.11330) (4, 37).
198 Hsp70 mRNA stability is regulated by heat response elements in the 3′-UTR during heat shock
199 (4-6), but corresponding changes at the protein level have not been demonstrated. Our goals
200 were to investigate the behaviour of endogenously T2A-tagged Hsp70 following heat shock and
201 to assess whether our system facilitated single-copy integration within a multi-gene array.
202
203 We designed primers based on the Tb927 genome copy to tag Hsp70 at both the N- and C-termini
204 with mScarlet (mSc) and mNeonGreen (mNG), respectively, under puromycin selection (**Fig. 3A**).
205 The design of sgRNA and donor repair templates followed the strategy used for ESAG3. A GSGS
206 linker was added between the mSc tag and the *Hsp70* ORF for N-terminal tagging, and 25 bp
207 homology arms were used in both repair templates. Post-transfection, locus-specific PCR
208 genotyping of two clones produced products of sizes consistent with the predicted fusions: *Hsp70*
209 ORF with mNG (409 bp), *PAC* with the *Hsp70* 3' UTR (304 bp), and *Hsp70* 5' UTR with *PAC* (339
210 bp) (**Fig. 3A**). No products were generated from genomic DNA purified from parental cells,
211 confirming locus-specific modification.
212
213 Western blot analysis with anti-Hsp70 antibodies showed two bands in all selected clones: one at

214 ~70 kDa (untagged endogenous Hsp70) and another at ~100 kDa, (tagged Hsp70: 101 kDa for
215 Hsp70::mT2A and 98 kDa for T2A-mSc::H). The higher MW bands were absent in parental cells
216 (**Fig. 3B**), indicating successful tagging. Flow cytometry revealed that ~99% of cells were positive
217 for either mNG or mSc (**Fig 3C**), and live-cell microscopy confirmed the characteristic cytosolic
218 localisation of Hsp70::mT2A (**Fig. 3D**), indicating that T2A tagging did not cause Hsp70 mis-
219 localisation.
220
221 To assess whether expression of mT2A-tagged Hsp70 tracked with the untagged allele following
222 heat shock, we performed immunoblotting with anti-mScarlet, anti-mNG, and anti-Hsp70
223 antibodies. We observed a slight reduction in both tagged and endogenous Hsp70 protein levels
224 after heat shock (**Fig. 3E**). However, qRT-PCR analysis showed a 2-fold upregulation of Hsp70
225 mRNA (**Fig. 3F**), suggesting a typical heat shock response at the mRNA level (6). We attribute
226 the reduction in protein levels to either global translational arrest or cell stress resulting from the
227 heat shock treatment (38).
228
229 Overall, our results confirmed UTR-dependent *Hsp70* mRNA regulation after heat shock, but not
230 at the protein level - a finding that suggests complex post-transcriptional regulation consistent
231 with observations in *Leishmania* (38). We successfully achieved efficient single-copy tagging of
232 cytosolic Hsp70 at both the N- and C-termini within a complex multi-gene array. The apparent
233 lower expression of tagged Hsp70 compared to the wild type likely indicates that only one of the
234 eight nearly identical gene copies was tagged, accounting for approximately 1/16th of the total
235 signal in Western blots of a diploid genome. This was further confirmed by whole genome
236 sequencing (see below).
237
238 **Test case III: T2A-tagged ESAG7 preserves iron-dependent regulation and function**
239
240 In functional studies, it is essential that tagging does not disrupt protein function or its regulatory
241 elements at the RNA level. In *T. brucei,* ESAG7 (E7) exemplifies this dual requirement: at the
242 protein level, it forms a 1:1 stoichiometric complex with ESAG6 to create a functional transferrin
243 receptor (TfR) for iron uptake (39). At the RNA level, its expression is regulated by iron-responsive
244 elements in its 3' UTR that control mRNA stability in response to iron availability (40). Tagging
245 approaches that replace the native 3' UTR can disrupt this iron-dependent regulation, potentially
246 altering the expression levels and stoichiometry of ESAG6 and ESAG7, thereby interfering with
247 TfR function (10, 11).

248

249 To investigate regulation and function of ESAG7, we generated mNG::T2A- and mNG::6xTy-
250 tagged ESAG7 cell lines (E7::mT2A and E7::mTy, respectively) (**Fig 4A**). E7::mT2A tagging
251 retained the native *E7* 3' UTR, while the E7::mTy tagging replaced the 3' UTR with the *PFR2*
252 intergenic region. Hsp70 (H::mT2A) was also tagged as a control. All analyses were performed
253 after treatment with the iron chelator Deferoxamine (DFO), which induces iron starvation. qRT-
254 PCR analysis showed 2-fold increased mNG transcript levels in E7::mT2A cells following DFO
255 treatment, while E7::mTy showed a reduction, and H::mT2A cells remained unchanged (**Fig 4B**).
256 This pattern was reflected at the protein level, as Western blotting confirmed increased E7::mT2A
257 protein abundance, with E7::mTy and H::mT2A remaining unaffected (**Fig 4C**). Flow cytometry
258 further supported these findings, showing a 3-fold increase in mNG expression in E7::mT2A cells,
259 while no significant changes were observed in E7::mTy or H::mT2A (**Fig 4D**).
260

261 The functional implications of these findings were assessed by analysing native TfR levels and
262 transferrin (Tf) uptake. Both qRT-PCR and Western blot analyses revealed increased TfR levels
263 in parental, H::mT2A, and E7::mT2A cells after DFO treatment, but showed a significant reduction
264 in E7::mTy cells (**Fig 4E and 4F**). Tf uptake assays demonstrated significant increased uptake in
265 E7::mT2A cells, confirming functional TfR activity, whereas E7::mTy cells showed no Tf uptake,
266 indicating a disruption of iron-dependent regulation and receptor function (**Fig 4G**).
267

268 Collectively, these data demonstrate that mNG::T2A-tagged ESAG7, when retaining its native 3'
269 UTR, maintains iron-dependent regulation and function. In contrast, replacing the 3' UTR with the
270 *PFR2* intergenic region disrupts both iron-responsive regulation and TfR function. This validates
271 our T2A-tagging approach as an effective tool for studying post-transcriptional iron regulation and
272 hence likely other regulatory systems in *T. brucei*.
273

274 **Confirmation of site-specific integration and in-frame fusion of the T2A gene tagging by**
275 **whole genome sequencing**
276

277 Finally, to confirm site-specific integration and in-frame fusion of mNG::T2A with the target ORFs,
278 we performed paired-end whole genome sequencing (WGS) of the five tagged cell lines at 30X
279 genome coverage. The reads were mapped to either the wild-type genomic locus or a manually
280 reconstructed locus containing the mNG::T2A or mNG::6xTy tagging cassettes fused to the target
281 genes (**Fig. S3**). In all cases, WGS confirmed that the donor repair cassettes were inserted in-

282 frame with the target gene in tagged cells, with no frameshifts or premature stop codons observed,
283 confirming site-specific in-frame fusion. Collectively, combining data from WGS, integration
284 PCRs, and Western blot analyses, we conclude that mNG::T2A fused to target ORFs
285 predominantly as single-copy integrations, and that T2A peptide-mediated cleavage occurs with
286 high efficiency. Collectively these results show that this approach is highly suitable for epitope-
287 tagging secretory and cytosolic proteins at their endogenous loci in *T. brucei*.
288
289 **Dedicated modular plasmid templates for high-throughput CRISPR/Cas9 T2A tagging: The**
290 **pRExT2A system**
291
292 To enhance the efficiency and versatility of T2A tagging in high-throughput applications, we
293 developed the pRExT2A system - plasmids for *R*egulated co-*Ex*pression mediated by *T2A*
294 peptide) - a set of dedicated plasmid templates optimised for *T. brucei*. This new design addresses
295 two key limitations of the initial T2A tagging system where: (i) the coding sequences for *mNG*,
296 *mSc*, and *drug*^{R} were optimised for *Leishmania* codon usage, which differs from *T. brucei* and
297 could lead to inefficient translation and reduced expression levels; and (ii) the addition of 57 bp to
298 the 3' UTR during C-terminal tagging, which could potentially disrupt native mRNA processing
299 signals and affect gene regulation.
300
301 To overcome these issues, we designed pRExT2A to be fully compatible with the pPOT (plasmids
302 for PCR-only tagging) platform by incorporating the coding sequences for drug resistance markers
303 and fluorescent proteins from pPOT into our new plasmids (19). Our goal was to integrate the
304 efficiency, modularity, and cloning-free advantages of pPOT while ensuring that primer usage is
305 compatible across both systems.
306
307 **Features and primer design of the pRExT2A system:** The pRExT2A system features modular
308 N- and C-terminal tagging donor plasmids encoding either a 3xTy-tagged fluorescent protein or a
309 3xTy epitope tag, linked via the T2A peptide to a 3xTy-tagged drug^{R} (**Fig. 5**). For N-terminal
310 tagging, the constructs follow the format: drug^{R}::3xTy::T2A::3xTy::FP or drug^{R}::3xTy::T2A::3xTy
311 (**Fig. 5A**), while for C-terminal tagging the orientation was as follows: FP::3xTy::T2A::drugR::3xTy
312 or 3xTy::T2A::drug^{R}::3xTy (**Fig. 5B**).
313
314 The primer design workflow remains consistent with our previous protocol (**Fig. 1**), but we
315 modified the primer binding sites to match the pPOT system. This allows users to use identical

316 primer pairs and sgRNAs across both systems. Although an improvement over the original T2A
317 tagging cassettes, this design introduces a short 20 bp scar into the 5' or 3' UTR after integration
318 - depending on whether N- or C-terminal tagging is used (**Fig. 5A** and **5B**). To facilitate high-
319 throughput applications, we developed an automated primer design tool
320 [https://github.com/zephyris/scarlesstagging] which generates primers compatible with both
321 pRExT2A and pPOT systems. This tool significantly simplifies experimental setup and minimises
322 design errors. To facilitate adoption, we provide pre-designed primers (including scarless options)
323 for all *T. brucei* genes in TREU927 and Lister 427 strains, and related Kinetoplastid species.
324
325 **Dual functionality of the 3xTy-tagged drug^R:** A key feature of the pRExT2A system is the
326 inclusion of a 3xTy-tagged drug^R which serves two functions: as a selection marker controlled by
327 the endogenous UTRs of the target gene and a reporter for studying mRNA/translational
328 regulation and post-translational regulation (**Fig. 5C)**. Since the drug^R and the target protein are
329 translated from the same mRNA but yield distinct proteins, any changes at the mRNA or
330 translational level affect both proteins equally. However, post-translational regulation only impacts
331 the target protein (**Fig. 5C** and **5D**). This design enables researchers to determine the level of
332 regulation by comparing the relative abundance of the mRNA (using qRT-PCR) and both proteins
333 by Western blotting, or flow cytometry, all performed within the framework of a single experiment.
334 A protein/drug^R ratio of 1 indicates regulation at the mRNA or translational level, while ratios
335 deviating from 1 suggest post-translational regulation of the target protein (**Fig 5D**).
336
337 **Modularity of the pRExT2A System**: We developed plasmid versions with various drug-
338 selectable markers, including blasticidin (BSD), hygromycin (HYG), and puromycin (PAC), as well
339 as two fluorescent proteins, mNeonGreen (mNG) and mScarlet (mSc) (**Table S1**). This modular
340 design allows for easy swapping of tags and drug^R markers, providing flexibility to accommodate
341 different experimental needs.
342
343 **The pRExT2A system retains life cycle stage-specific gene regulation and confirms UTR-**
344 **mediated control**
345
346 To evaluate the ability of the pRExT2A system to preserve critical regulatory elements, we tested
347 three developmentally regulated genes in *T. brucei*: glycosylphosphatidylinositol phospholipase
348 C (GPI-PLC), which is expressed exclusively in bloodstream forms (BF) via 3' UTR regulatory
349 elements, and two paralogous calpain-related proteins, CAP5.5V (BF-specific) and CAP5.5

350 (procyclic form, PF-specific) (7, 8, 41). These genes provide ideal test cases for assessing the
351 system's ability to maintain post-transcriptional regulatory mechanisms during *in vitro*
352 differentiation.
353
354 **GPI-PLC example:** GPI-PLC (Tb927.2.6000) is expressed exclusively in BF via regulatory
355 elements in its 3' UTR, and it is enzymatic activity is essential for VSG shedding during the early
356 stages of differentiation to PF (41, 42). To track its expression during BF to PF differentiation, we
357 generated C-terminal mNG::3xTy fusions (GPI-PLC::mT2A::3xTy) using the pRExT2A system,
358 under hygromycin selection (**Fig. 6A)**. Live imaging showed GPI-PLC::mT2A::3xTy predominantly
359 localised to the flagella membrane (**Fig. 6B**), consistent with native GPI-PLC antibody staining
360 (43) and GPI-PLC-eYFP localisation (44). This pattern differs from the dual flagella/cell body
361 distribution seen with pPOT tagging where the 3'-UTR was replaced (21). Western blot analysis
362 using validated anti-GPI-PLC antibodies (43), showed expression of both native GPI-PLC (~41
363 kDa) and GPI-PLC::mT2A::3xTy (G::mT2A; ~73 kDa) specifically in BF cells (**Fig 6C, left**). The
364 stoichiometric ratio between tagged and wild-type alleles indicated single-copy integration in the
365 diploid genome, and efficient T2A cleavage was confirmed by detection of separate 3xTy::HYG
366 product (42 kDa) in anti-Ty Western blots (**Fig 6C, right**). Importantly, flow cytometry showed that
367 GPI-PLC::mT2A expression dropped to near-background levels after differentiation to PF (**Fig**
368 **6D, 6E**), indicating that the pRExT2A system preserves the native UTR-mediated regulation of
369 GPI-PLC expression and localisation. The distinct localisation patterns between our UTR-
370 preserved and UTR-replaced constructs (21) highlight the importance of maintaining endogenous
371 UTRs for proper protein expression and localisation.
372
373 **CAP5.5V and CAP5.5 examples:** CAP5.5V (Tb927.8.8330) and CAP5.5 (Tb927.4.3950) play
374 crucial roles in maintaining cell morphology by organising the subpellicular cytoskeleton in BF and
375 PF, respectively (8). In the TrypTag project, CAP5.5V showed no expression in PF when tagged
376 at the N-terminus, but C-terminal tagging, which replaced the native 3' UTR, led to inappropriate
377 expression in PF forms, demonstrating the critical role of UTR-mediated regulation (45). To
378 prevent such dysregulation, we preserved the native UTRs while tagging CAP5.5V and CAP5.5
379 at their C-termini using the pRExT2A system (**Fig. 7A)**.
380
381 Live-cell microscopy confirmed the characteristic cytoskeletal localisation of CAP5.5V across the
382 cell body (excluding the flagellum) in BF cells (**Fig. 7B**), while CAP5.5 expression was not
383 detectable above background in BF, consistent with its known stage-specific expression. Anti-Ty

384 immunoblots revealed stoichiometric expression of the tagged proteins: CAP5.5V::3xTy::T2A
385 (104 kDa), CAP5.5V::mNG::3xTy::T2A (131 kDa), CAP5.5::3xTy::T2A (100 kDa), and
386 CAP5.5::mNG::3xTy::T2A (127 kDa), as well as the cleaved 3xTy::PAC product (~25 kDa),
387 indicating efficient T2A processing (**Fig. 7C**). After differentiation to PF, flow cytometry showed
388 the expected stage-specific regulation: CAP5.5V fluorescence decreased, while CAP5.5 signal
389 increased (**Fig. 7D** and **7E**), corresponding to the previously reported 19-fold upregulation of
390 CAP5.5V mRNA in BF and 5-fold enrichment of CAP5.5 mRNA in PF (8). This differential
391 expression of the two paralogs during differentiation validates the ability of the pRExT2A system
392 to preserve complex stage-specific gene expression patterns mediated by endogenous UTRs.
393
394 **Discussion**
395
396 Kinetoplastids, such as *T. brucei,* serve as powerful models for understanding fundamental
397 biological processes, particularly the diversity and evolution of eukaryotic gene control (1, 3). The
398 genetic toolbox for *T. brucei* is extensive, with existing tools for genome-wide conditional
399 expression and localisation screens (20, 46, 47). However, efficiently tagging proteins in their
400 native regulatory context, with positive selection for high throughput studies, remains challenging.
401
402 Our new CRISPR/Cas9-T2A system using the pRExT2A plasmids adds to this toolbox, enabling
403 precise, efficient, scalable, stage-specific expression while retaining native mRNA processing
404 signals, specifically native spliced leader addition sites (SLAS), native polyadenylation sites
405 (PAS), native untranslated regions (UTRs), and any regulatory elements contained within these
406 regions (1, 15). Importantly, the tagged proteins tested in this system retain correct localisation
407 and differential expression in the tested life cycle stages. We demonstrate the versatility of this
408 system by tagging proteins involved in key pathways, including GPI-PLC - which is critical for
409 membrane lipid remodelling (48), CAP5.5V and CAP5.5, which are involved in cytoskeletal
410 organisation (7, 8), and ESAG7, a subunit of the transferrin receptor responsible for iron uptake
411 (49). The high efficiency allows for tagging up to 10 proteins per week in BF parasites, as the
412 protocols for generating repair templates are well-established (21). A beginner's guide for using
413 this toolkit is outlined in **Fig. 8**.
414
415 A key innovation of our system is the dual functionality of the drug resistance marker as both a
416 selection tool and an expression control (**Fig. 5**). This feature allows users to distinguish between
417 transcriptional/translational regulation and post-translational effects by comparing the levels of

418 the resistance marker to the tagged protein of interest. This approach has proven valuable in
419 related kinetoplastids, such as *L. mexicana*, where T2A-based tagging showed that life cycle-
420 dependent expression of LmxGT1 is controlled by changes in protein stability rather than mRNA
421 levels (28). In contrast, in *L. donovani*, T2A applications showed that upregulation of LdNT3 is
422 driven by increased transcript stability and translational efficiency, not changes in protein turnover
423 (30).
424
425 Using this system in *T. brucei*, we can systematically study proteins involved in nutrient transport
426 and their regulation across different stages of the life cycle. For example, this system can enable
427 investigation into the multi-layered control of iron response genes by examining multiple proteins
428 identified from transcriptomics and proteomics screens (10). This approach can help distinguish
429 between transcriptional, post-transcriptional, and post-translational regulatory events (**Fig. 5D**),
430 while characterising iron response elements (IREs) in a high-throughput format. Furthermore,
431 combining our system with single-molecule RNA FISH enables simultaneous analysis of RNA
432 processing dynamics, offering insights into gene regulation at all levels in a single experiment.
433
434 Another advantage of this system is its compatibility with the pPOT platform, enabling seamless
435 integration into existing workflows and allowing for direct comparisons with protein expression
436 levels where exogenous UTRs from the pPOT system serve as controls (21). For instance, we
437 show that tagging ESAG7 by swapping UTRs diminished native regulation and function (**Fig. 4).**
438 The modular architecture of the pRExT2A offers flexibility for easy swapping of tags, drug
439 resistance markers, or regulatory elements without the need for extensive re-cloning. This
440 adaptability facilitates rapid development of new constructs for diverse experimental applications.
441
442 A key strength of this system is its ability to track stage-specific protein expression during
443 differentiation, as demonstrated for the dynamic regulation of cytoskeletal proteins CAP5.5V, and
444 CAP5.5 during life cycle transitions. By tagging multiple components with different fluorophores,
445 we can visualise complex assembly and disassembly in real-time during differentiation, revealing
446 both the timing of protein interactions and factors affecting complex formation. For example,
447 tagging components of mitochondrial respiratory complexes alongside iron-sulfur cluster
448 assembly proteins could allow tracking of respiratory complex assembly relative to transferrin
449 receptor (TfR) downregulation. Such experiments could reveal how iron is mobilised from storage
450 proteins to newly synthesised mitochondrial complexes and how iron availability affects these
451 metabolic transitions.

452

453 Our T2A system has several technical considerations. First, CRISPR-based integration leaves a
454 small (20 bp) scar sequence that could theoretically affect gene expression, though we observed
455 no impacts during validation experiments. To address this potential limitation, we provide pre-
456 designed primers for scarless integration, although these remain to be tested. Second, the system
457 predominantly tags one gene copy, which can be challenging when targeting genes within
458 multigene arrays that share identical UTRs, such as *Hsp70*. For such cases, specific targeting
459 can be achieved by manually designing primers to regions that are not identical within the UTRs.
460 For protein detection, the shared 3xTy epitope tag between the drug resistance marker and target
461 protein creates specific constraints. While Western blotting can distinguish these proteins by size,
462 immunofluorescence cannot separate their signals. We recommend either live imaging or using
463 anti-mNeonGreen antibodies for localisation studies and choosing drug resistance markers of a
464 different molecular to your target protein. Additionally, the limited dynamic range of mNeonGreen
465 may not detect subtle expression changes, such as responses to nutrient starvation. To overcome
466 this, we are developing luciferase-based vectors that offer enhanced sensitivity for monitoring
467 dynamic responses to environmental signals.
468

469 Our T2A system has the potential to offer new insights into the complex biology of these parasites
470 and contribute to our broader understanding of eukaryotic gene regulation, particularly in systems
471 where post-transcriptional control plays a dominant role. By preserving native regulatory contexts,
472 this approach could accelerate our understanding of molecular mechanisms underlying parasite
473 adaptation and survival.
474

475 **Materials and methods**

476

477 ***T. brucei* cell culture and generation of constitutive Cas9 expressing cells**

478 A tetracycline inducible single-marker bloodstream form Lister 427 strain *T. brucei brucei* was
479 used for all experiments (50). The cells were transfected with a humanised *Streptococcus*
480 *pyogenes* Cas9 plasmid (pTB011) that integrates upstream of the first β-tubulin gene in the tubulin
481 array and is under Pol II transcription (33). Cas9 expression was confirmed by Western blot using
482 anti-Cas9 (Abcam). The cells were cultured at 37°C, 5% CO_2 in HMI-9 medium supplemented
483 with 20% heat inactivated Foetal Bovine Serum, 100 U/ml penicillin and 100 µg/ml streptomycin
484 (Gibco). Cell lines were propagated with the appropriate selection drugs where necessary: 2.5

485 µg/ml G418 (parental) and 10 μg/ml Blasticidin for Cas-expressing cells and 0.2 µg/ml Puromycin
486 for reporter or epitope-tagged cell lines.
487
488 **Construction of T2A peptide-based plasmid templates for endogenous tagging**
489 The "donor vectors" for T2A peptide-based tagging were originally designed as "SfiI cassettes" to
490 fit into our multi-fragment assembly method for generating targeting constructs for integration via
491 homologous recombination (51). The PAC::T2A::mNG and mNG::T2A::PAC tagging cassettes
492 were synthesised in full by GenScript. The PAC and mNeon genes were encoded with a
493 *Leishmania donovani* codon bias (i.e., predominantly G or C in the wobble position), and a nine
494 amino acid flexible linker was included upstream of the T2A peptide sequence to facilitate efficient
495 co-translational cleavage (34) The mScarlet (52) gene was PCR amplified from
496 pCytERM_mScarlet_N1 (Addgene_85066) and inserted into N-terminal and C-terminal "Empty
497 T2A::Drug^R cassettes as described in the **Fig. S1**. All pRExT2A plasmids were synthesised by
498 Genscript using sequences from the pPOT series which have been optimised for *T. brucei* (21).
499
500 **sgRNA and repair template design and assembly**
501 All oligonucleotides for generation of donor DNA repair templates and for analyses of edited
502 genomic loci were synthesised by Life Technologies (Invitrogen, UK) and presented below or in
503 the supplementary material. sgRNA primers targeting specific genomic loci were selected by
504 choosing a 20 bp sequence adjacent to a Protospacer Adjacent Motif (PAM, 5'-NGG) site that is
505 either close to the start codon (for 3' targeting) or close to the STOP codon (for 5' targeting).
506 SgRNA primers were cross-checked for any off-target binding. PCR amplification of all sgRNAs
507 followed the protocol published Beneke et al using 2 μM of primer G00 (sgRNA scaffold) as
508 template and a gene-specific sgRNA primer (33). The resulting sgRNA PCR product and
509 matching PCR-amplified repair template were pooled and transfected into cells that express T7
510 polymerase for T7 RNAP-driven transcription (50).
511
512 The repair template primers were designed to amplify the fluorescent TAG::T2A::drug^R with 25 -
513 40 bp homology arms targeting the gene of interest and the adjacent UTR (Fig 1). The homology
514 arms were selected such that full length endogenous 5' and 3' UTR sequences remained intact
515 following repair. The STOP and START codons were omitted for either C-terminus or N-terminus
516 tagging, respectively. To disrupt the PAM sequence or sgRNA binding regions within the repair
517 templates synonymous mutations or an *Xba*I site was introduced in the repair template primer,

518 highlighted in bold in the primer sequences (22). For PCR amplification of repair templates, 30 ng
519 of plasmid was used with gene-specific forward and reverse primers (**Table 1**).
520
521 **PCR of repair templates and analyses of genomic integrations**
522 All PCR assays for amplification of repair templates were performed in a 50 μl reaction volume
523 using Q5 High-Fidelity 2X MasterMix (New England Biolabs NEB Cat #M0515) according to
524 manufacturer's instructions. The reagent components were assembled on ice and transferred to
525 a preheated thermal cycler at 98.0°C. The annealing temperature for each primer pair was
526 calculated using the NEB Tm calculator online tool. To validate correct integration of tags at the
527 target loci, DNA was extracted from wild type and transfected cells using a DNeasy blood and
528 tissue Kit (Qiagen). All PCR assays for genomic DNA analysis of edited loci were performed in a
529 25 μl reaction volume using Hot Start Taq 2X Master Mix (NEB, Cat #M0496). Five microlitres of
530 all PCR reactions were resolved on a 1% agarose gel to verify the presence of correct size PCR
531 products. For high-throughput applications, we recommend well-established workflows published
532 by Beneke et al 2017 (21).
533
534

535 **Table 1:** Primer sequences used for PCR amplification of sgRNA and DNA repair templates. ReF;
536 forward repair and ReR; reverse repair. Primer binding sites corresponding to regions in the
537 plasmid are shown in italics. Bold indicate synonymous mutations to prevent further Cas9 edits.
538

Primer description	Sequence
G00 sgRNA	aaaagcaccgactcggtgccactttttcaagttgataacggactagccttattttaacttgctatttctagctctaaaac
E3	
3'-sgRNA E3::m2A	gaaattaatacgactcactataggTTAAAGCAAAGAAACTGTAAgttttagagctagaaatagc
E3::m2A-ReF	CGCGGTCAAGGAGAATGACTTGTTTAAAGCAAAGAA**G**CT**C***ggcggtGTGTCCAAGGGC*
E3::m2A-ReR	GCCGCAAGAAAATCTTTAGTTTTGTATO**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*
E7	
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*

539
540
541 **Transfection of pooled PCR-amplified sgRNA and donor repair templates.** 1-3 x 10^7 Lister
542 427 bloodstream cells were harvested and resuspended in 1x TbBSF Roditi buffer premixed with
543 50 µL of pooled PCR reactions (25 µl repair template fragment + 25 µl sgRNA; ~25 µg total DNA
544 each) without purification. The cell suspension in a final volume of 200 µl was electroporated
545 using an Amaxa Nucleofector IIb (Lonza), program X-001. Cells were recovered for 6h at 37
546 ^{0}C/5% CO_2 before selection with appropriate antibiotics and distributed in 24 well plates (2 ml in
547 each well). Positive clones were selected after 5 – 7 days and analysed by flow cytometry and

548 Western blots. Correct integration of repair templates at target loci were confirmed by locus-
549 specific PCR and genotyped by performing whole genome sequencing at
550 BGI technologies (China). Locus-specific PCR primers are detailed in **Table S2**.
551
552 **Western blotting and antibodies.** Total protein was extracted from wild type and transgenic
553 cells, washed in 1x Phosphate buffered saline, lysed in 1x SDS loading buffer, separated by SDS
554 PAGE, transferred to PVDF membranes and immunoblotted with appropriate antibodies for 1 hr
555 at room temperature: anti-ESAG3 (rabbit, 1:2500), anti-TfR (rabbit, 1:2500), anti-Hsp70 (rabbit,
556 1:5000) (53), anti-mNG (Mouse monoclonal antibody [32F6], ChromoTek 1:0000), anti-RFP
557 (Mouse monoclonal antibody [6G6], ChromoTek 1:1000) and anti-2A (monoclonal clone 3H4,
558 Merck, 1:1000,). The membranes were blocked in 5% milk, washed 3x in 1xPSBT after antibody
559 incubation and incubated for 1 hr at room temperature with the appropriate horseradish
560 peroxidase (HRP)-conjugated secondary antibodies (anti-mouse IgG, 1:5000; and anti-rabbit IgG,
561 1:5000 dilution), developed after 1 min incubation with Super Signal ECL reagent (Thermo Fisher
562 Scientific), and imaged using a ChemiDoc Gel Imaging MP System (Bio-Rad).
563
564 **Quantitative real time PCR (qRT-PCR).** All qRT-PCR analyses used ~1–5 x 10^7 cells total, and
565 RNA was extracted from cell pellets using Qiagen RNeasy Extraction Kit. qRT-PCR from total
566 RNA was performed using Luna Universal One-step kit (NEB) per manufacturer's instructions.
567 All reactions were performed in technical duplicates and specificity of each reaction was checked
568 by ensuring single peaks from melt curves. Fold change was calculated by comparative C_T
569 method using the Applied Biosciences software. Relative transcript levels were compared to
570 endogenous control, *ZFP* (Tb927.3.720, nts 241-301) relative to untreated controls. Primer
571 sequences are available in supplementary material.
572
573 **Heat shock and iron starvation inductions**.Iron starvation was induced for 5 hr at cell densities
574 £ $5x10^5$ cells/ml with 25 mM Deferoxamine (Sigma Aldrich) in HMI-9 medium. For heat shock,
575 Hsp70 tagged cells were maintained at 37 °C or transferred to 42 °C for one hour in HMI-9 medium
576 at same cell densities £ $5x10^5$ cells/ml. In both cases, cells were harvested and analysed by qRT-
577 PCR, flow cytometry or Western blots.
578
579 **Immunofluorescence microscopy and DAPI staining.** Immunofluorescence assays were
580 performed as previously described in (54). αBiP (Rabbit) antibody dilution was 1:2000, anti-mNG
581 (Mouse monoclonal antibody [32F6], ChromoTek 1:100). Slides were washed with 1x PBS then

582 incubated with secondary antibody for 1 hr at room-temperature: Alexa-488 conjugated a-mouse
583 (1:500) and Alexa-594 conjugated a-rabbit (1:500) (Molecular Probes, USA). Slides were washed
584 in 1x PBS and mounted in VectaShield mounting medium (Vector Laboratories) containing 4′,6-
585 diamidino-2-phenylindole (DAPI), sealed with coverslips and imaged on a Zeiss AxioImager M2
586 with an ORCA Flash 4 camera.
587
588 **Flow cytometry.** Live cells were harvested, washed in 1x PBSG, resuspended in 1 mL 1x PBSG,
589 and analysed with an LSR Fortessa cell analyser using BD FACS Diva software (BD Biosciences).
590 For uptake of Alexa-647 transferrin (Tf-647), the cells were pre-incubated with serum-free HMI9
591 supplemented with 0.5 μg/mL BSA at 37 ^{0}C for 10 minutes. Tf-647 was added at 50 μg/mL and
592 incubated for 30 minutes. The cells were washed twice in 1x PBSG and analysed by flow
593 cytometry. Post-acquisition analysis of flow cytometry data was performed with FlowJo v10
594 software (BD Biosciences).
595
596 **Development of an automated primer design tool.** To facilitate use of this tool, we developed
597 code for automatic design of primers. This was implemented as a Python module, using sequence
598 and metadata retrieval from TriTrypDB (55) to allow design of primers from a gene ID or list of
599 gene IDs as found in TriTrypDB. For ease of use, we also wrote an iPython/Jupyter Notebook
600 that runs in Google Collaboratory. Anyone can visit this notebook (available at:
601 https://github.com/zephyris/scarlesstagging) and design primers without any need for coding
602 ability or installing software on a computer.
603
604 The automated primer design attempts to design primers which make absolutely no change to
605 the 5` (for N terminal tagging) or 3` (for C terminal tagging) UTRs. Without a CRISPR- mediated
606 double strand break, this could be done by providing homology arms for the homologous repair
607 construct which are, for N terminal tagging, the end of the 5` UTR and start of the ORF or, for C
608 terminal tagging, the end of the ORF (without the stop codon) and the start of the 3` UTR. For
609 CRISPR-assisted homologous recombination, a double strand cut site defined by a protospacer
610 adjacent motif (PAM) site of GG must be selected near the start (for N) or end (for C) of the ORF.
611 However, any such site would also cut the repair construct. Therefore, we search for protospacer
612 adjacent motif (PAM) sites within the ORF which can be modified to remove the PAM without
613 changing the encoded amino acid. This search range is defined by the number of bases typically
614 available for primer synthesis (100 bases) minus the length of the template plasmid annealing
615 site (template dependent, around 18 bases) and the length of homology (25 bases), searching

616 within the start or end of the ORF for N or C terminal tagging respectively. The PAM site closest
617 to the start (for N terminal) or end (for C) of the ORF is selected. The final primer sequences for
618 the generation of the sgDNA are then constructed from for the automatically selected PAM site,
619 and the forward and reverse primer sequences for the homologous repair construct constructed
620 for the user-selected template plasmid and automatically selected PAM site. The reverse (for N
621 terminal) and forward (for C) primers contain a single point mutation in the ORF which removes
622 the PAM site and so prevents CRISPR-mediated cutting.
623
624 **Code availability:** The automated primer design tool for our tagging system is freely available at
625 https://github.com/zephyris/scarlesstagging, along with pre-designed primers for *T. brucei*
626 (TREU927, Lister 427).
627
628 **Acknowledgements:** pCytERM_mScarlet_N1 was a gift from Dorus Gadella (Addgene plasmid
629 # 85066; http://n2t.net/addgene:85066 ; RRID:Addgene_85066). The authors are grateful to
630 Professor Jay Bangs (University at Buffalo, USA) for generous anti-TfR and anti-Hsp70
631 antibodies, and Professor Keith Gull, FRS for his insightful discussions and advice on gene choice
632 for developmental regulation and experimental design.
633
634 **Funding:** This work was funded by a Wellcome Trust and Royal Society Sir Henry Dale
635 Fellowship (208780/Z/17/Z) to CT, and a NIAID, National Institutes of Health Grant R03AI137636
636 to PAY. Richard Wheeler is supported by a Wellcome Trust Sir Henry Dale Fellowship
637 (211075/Z/18/Z)
638

References

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

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

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

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

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

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

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

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

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

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

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

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

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

679 14. Hehl A, Vassella E, Braun R, Roditi I. A conserved stem-loop structure in the 3'
680 untranslated region of procyclin mRNAs regulates expression in Trypanosoma brucei. Proc Natl
681 Acad Sci U S A. 1994;91(1):370-4.

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

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

686 17. Oberholzer M, Morand S, Kunz S, Seebeck T. A vector series for rapid PCR-mediated C-
687 terminal in situ tagging of Trypanosoma brucei genes. Mol Biochem Parasitol. 2006;145(1):117-
688 20.

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

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

694 20. Billington K, Halliday C, Madden R, Dyer P, Barker AR, Moreira-Leite FF, et al. Genome-
695 wide subcellular protein map for the flagellate parasite Trypanosoma brucei. Nat Microbiol.
696 2023;8(3):533-47.

697 21. Beneke T, Madden R, Makin L, Valli J, Sunter J, Gluenz E. A CRISPR Cas9 high-
698 throughput genome editing toolkit for kinetoplastids. Royal Society open science.
699 2017;4(5):170095.

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

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

705 24. Donnelly ML, Luke G, Mehrotra A, Li X, Hughes LE, Gani D, et al. Analysis of the
706 aphthovirus 2A/2B polyprotein 'cleavage'mechanism indicates not a proteolytic reaction, but a
707 novel translational effect: a putative ribosomal 'skip'. Journal of General Virology.
708 2001;82(5):1013-25.

709 25. Ryan MD, King AM, Thomas GP. Cleavage of foot-and-mouth disease virus polyprotein is
710 mediated by residues located within a 19 amino acid sequence. Journal of General Virology.
711 1991;72(11):2727-32.

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

714 27. Heras S, Thomas M, Garcia-Canadas M, Felipe Pd, Garcia-Perez J, Ryan MD, et al. L1Tc
715 non-LTR retrotransposons from Trypanosoma cruzi contain a functional viral-like self-cleaving
716 2A sequence in frame with the active proteins they encode. Cellular and Molecular Life
717 Sciences CMLS. 2006;63:1449-60.

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

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

725 30. Licon MH, Yates PA. Purine-responsive expression of the Leishmania donovani NT3 purine
726 nucleobase transporter is mediated by a conserved RNA stem–loop. J Biol Chem.
727 2020;295(25):8449-59.

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

731 32. Webb H, Burns R, Ellis L, Kimblin N, Carrington M. Developmentally regulated instability of
732 the *GPI-PLC* mRNA is dependent on a short-lived protein factor. Nucleic Acids Res.
733 2005;33:1503-12.

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

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

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

743 36. Müller LSM, Cosentino RO, Förstner KU, Guizetti J, Wedel C, Kaplan N, et al. Genome
744 organization and DNA accessibility control antigenic variation in trypanosomes. Nature.
745 2018;563(7729):121-5.

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

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

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

754 40. Tiengwe C, Muratore KA, Bangs JD. Surface proteins, ERAD and antigenic variation in
755 Trypanosoma brucei. Cell Microbiol. 2016.

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

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

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

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

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

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

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

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

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

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

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

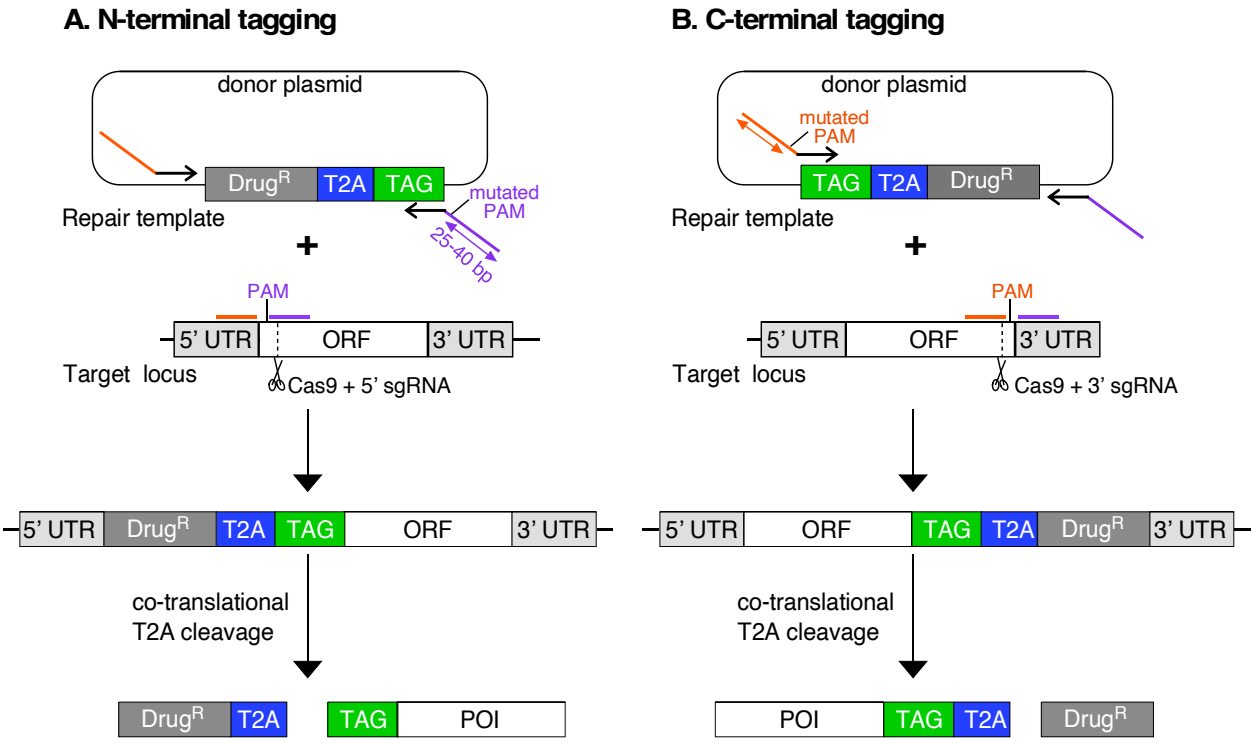
52. Bindels DS, Haarbosch L, van Weeren L, Postma M, Wiese KE, Mastop M, et al. mScarlet: a bright monomeric red fluorescent protein for cellular imaging. Nat Methods. 2017;14(1):53-6.

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

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

792 55. Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, et al.
793 TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res.
794 2010;38(Database issue):D457-62.
795
796

797 **Figures**
798
799
800



801
802
803
804 **Fig. 1 Strategy for endogenous T2A peptide-based protein tagging**
805 The T2A peptide-based tagging system relies on CRISPR/Cas9-mediated insertion of a repair
806 template either on the 5' end (**A**) or the 3'end (**B**) of the open reading frame of a target gene. The
807 repair template consists of a drug-selectable marker (Drug^R) linked by a *Thosea asigna* T2A
808 peptide sequence to a fluorescent protein or an epitope tag (TAG), placed at the N- or C-terminus
809 of a protein of interest (POI). PCR amplification of donor repair templates uses primers containing
810 ~25 - 40 bp of homology arms (orange and purple lines) flanking the Drug^R::T2A::TAG cassette.
811 The single guide RNA (sgRNA) directs Cas9 to introduce a double-stranded break adjacent to
812 the protospacer-adjacent motif (PAM), and mutated PAM sites from the repair template prevent
813 further edits. Following correct integration, the T2A peptide causes co-translational cleavage,
814 resulting in expression of both the Drug^R and POI in equal amounts. In both configurations, the
815 sgRNA and repair template primers are designed so that the full-length endogenous 5' and 3'
816 UTR sequences remain intact and drive expression of the Drug^R protein and POI.
817
818

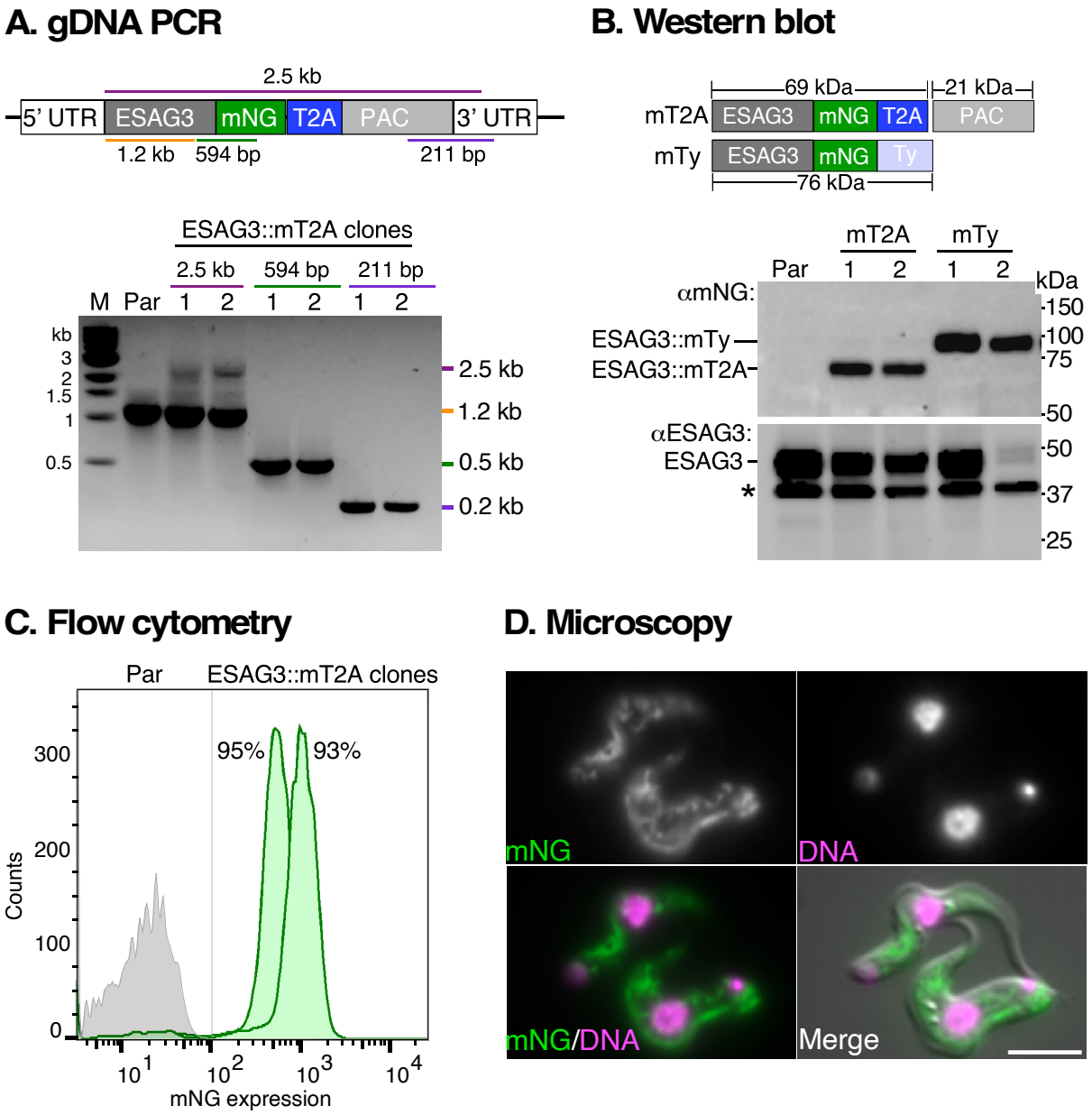


Fig. 2: High-efficiency T2A cleavage in *T. brucei* and localisation of ESAG3::mNG::T2A.
A. PCR validation of *ESAG3::mT2A* tagged genomic locus. Diagram shows the *ESAG3* gene locus after insertion of a repair cassette containing mNeonGreen (mNG), T2A peptide sequence (T2A), and puromycin (PAC) resistance gene for selection. PCR was performed using genomic DNA from parental cells (Par) and two transfected ESAG3::mT2A clones (1, 2). Coloured lines and arrows indicate positions of PCR primers to verify *ESAG3* locus-specific integration, with corresponding size fragments shown on the right of the agarose gel image. Molecular weight markers (M) are shown on the left. All diagrams are not drawn to scale.
B. Validation of tagged ESAG3::mT2A protein. Top panel: Diagram shows predicted molecular weights (MW) without (90 kDa) and with (69 kDa) T2A cleavage during translation of *ESAG3::mT2A::PAC* (mT2A) mRNA. Independent cell lines were generated by CRISPR/Cas9

833 mediated fusion of ESAG3 with mNG with 6xTy (ESAG3::mTy, mTy, predicted MW 76 kDa)
834 without incorporating a T2A peptide sequence. The ESAG3::mTy protein served as an internal
835 control for SDS-PAGE migration analysis. Bottom panel: Western blot analyses using anti-mNG
836 (αmNG, top) and anti-ESAG3 peptide (αESAG3, bottom) antibodies. Asterisks indicate αESAG3
837 non-specific cross-reacting polypeptide.
838 **C**. Evaluation of mNG-positive cells by flow cytometry. Parental cells (Par, gray histogram) and
839 ESAG3::mNG::T2A-tagged cells (ESAG3::mT2A, clones 1 and 2, green histograms) are shown.
840 Inset: percentage of mNG-positive cells.
841 **D.** Localisation of ESAG3::mT2A protein visualised by live-cell epifluorescence microscopy
842 (green), co-stained with Hoechst 33342 (magenta; DNA). Scale bar = 5 μM.
843

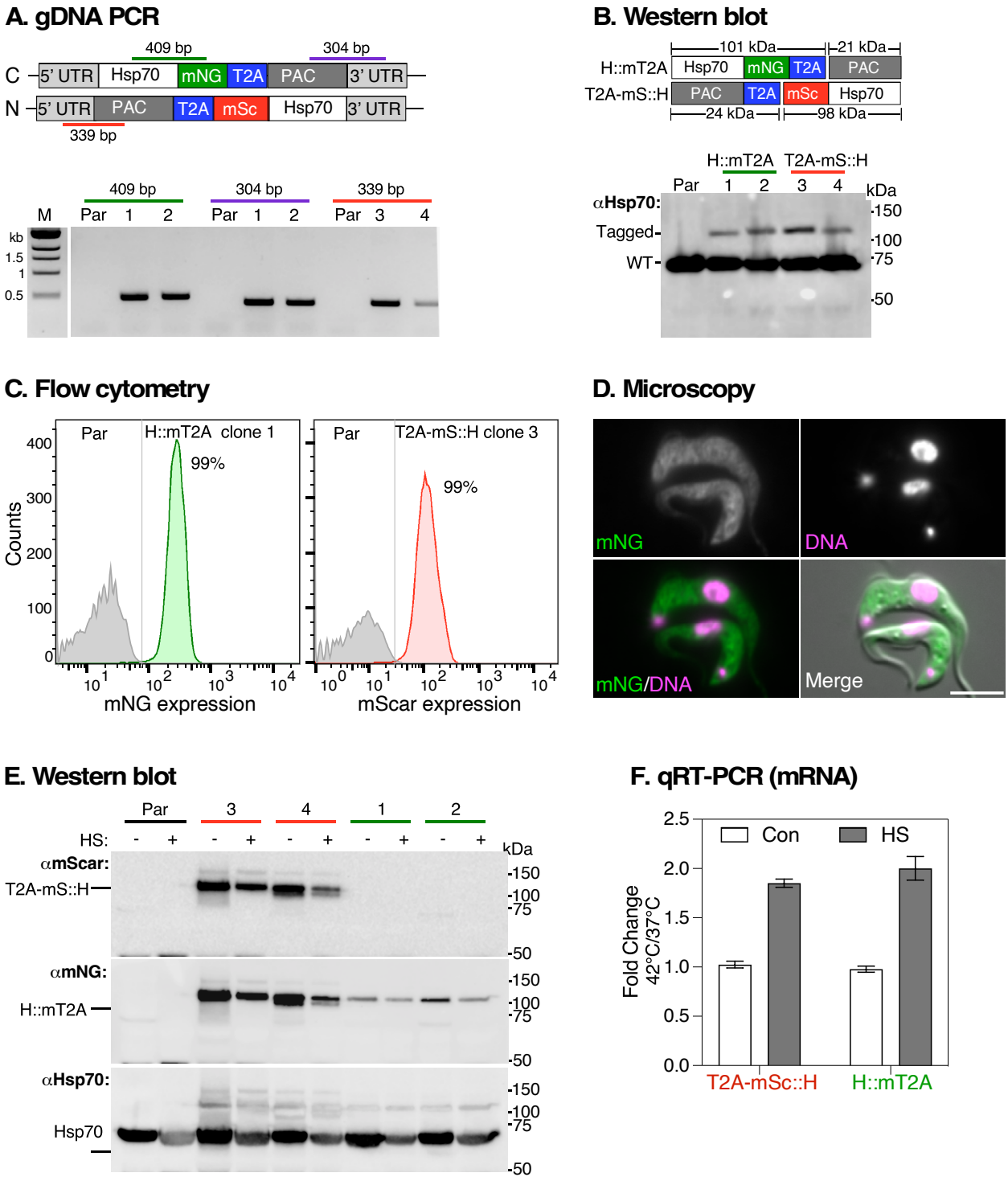


Fig. 3: Effect of heat shock on T2A-tagged Hsp70 mRNA and protein following heat shock. **A.** PCR validation of *Hsp70* tagged loci. Top panel shows schematic of the locus after insertion of a mNG::2A::drug repair cassette. Bottom panel shows agarose gel image of PCR products using genomic DNA extracted from parental (Par) cells or clonal cells transfected with repair templates for Hsp70 tagged either at the C-terminus with mNeonGreen (mNG, Hsp70::mT2A) or

851 at the N-terminus with mScarlet (mSc, T2A-mS::Hsp70). Coloured lines indicate positions of PCR
852 primers with corresponding size fragments shown above each lane on the agarose gel image. M:
853 molecular markers, two clones for mNG (1, 2) and mSc (3, 4).
854 **B.** Top panel: Diagram shows the predicted MWs of tagged Hsp70 protein before (122 kDa) or
855 after (101 kDa) 2A-mediated co-translational intra-ribosome skipping. Bottom panel: Western blot
856 analyses of tagged and untagged (WT) Hsp70 protein with native anti-Hsp70 (αHsp70) antibody.
857 **C.** Flow cytometry analysis of tagged Hsp70 cells. Parental cells (Par, gray) and cells in which
858 Hsp70 was either mNG-tagged (left) or mSc-tagged (right). Inset shows percentage of positive
859 fluorescent cells.
860 **D.** Localisation of Hsp70::mT2A protein visualised by live-cell epifluorescence microscopy
861 (green), co-stained with Hoechst 33342 (magenta; DNA). Scale bar = 5 μM.
862 **E.** Regulation of tagged Hsp70 protein expression after heat shock. Cell cultures were maintained
863 at 37 °C (-) or transferred to 42 °C (+) for one hour. Total protein extracts were analysed by
864 western blot with anti-mSc (αmSc), anti-mNG (αmNG) and anti-Hsp70 (αHsp70) antibodies. Two
865 clones are shown for parental (Par) and each tagged cell line. Note: The same membrane was
866 sequentially probed without stripping, starting with αmSc, followed by αmNG, and finally αHsp70.
867 **F.** Regulation of *Hsp70* mRNA expression after heat shock. Cell cultures were maintained at 37
868 °C (Control, Con) or transferred to 42 °C (Heat shock, HS) for one hour and mRNA levels
869 quantified by qRT-PCR using primers specific to endogenous Hsp70. Fold change is calculated
870 relative to 37 °C controls with *ZFP3* as endogenous control. Error bars represent standard
871 deviation from two clonal cell populations with three technical duplicates for each clone.
872
873
874

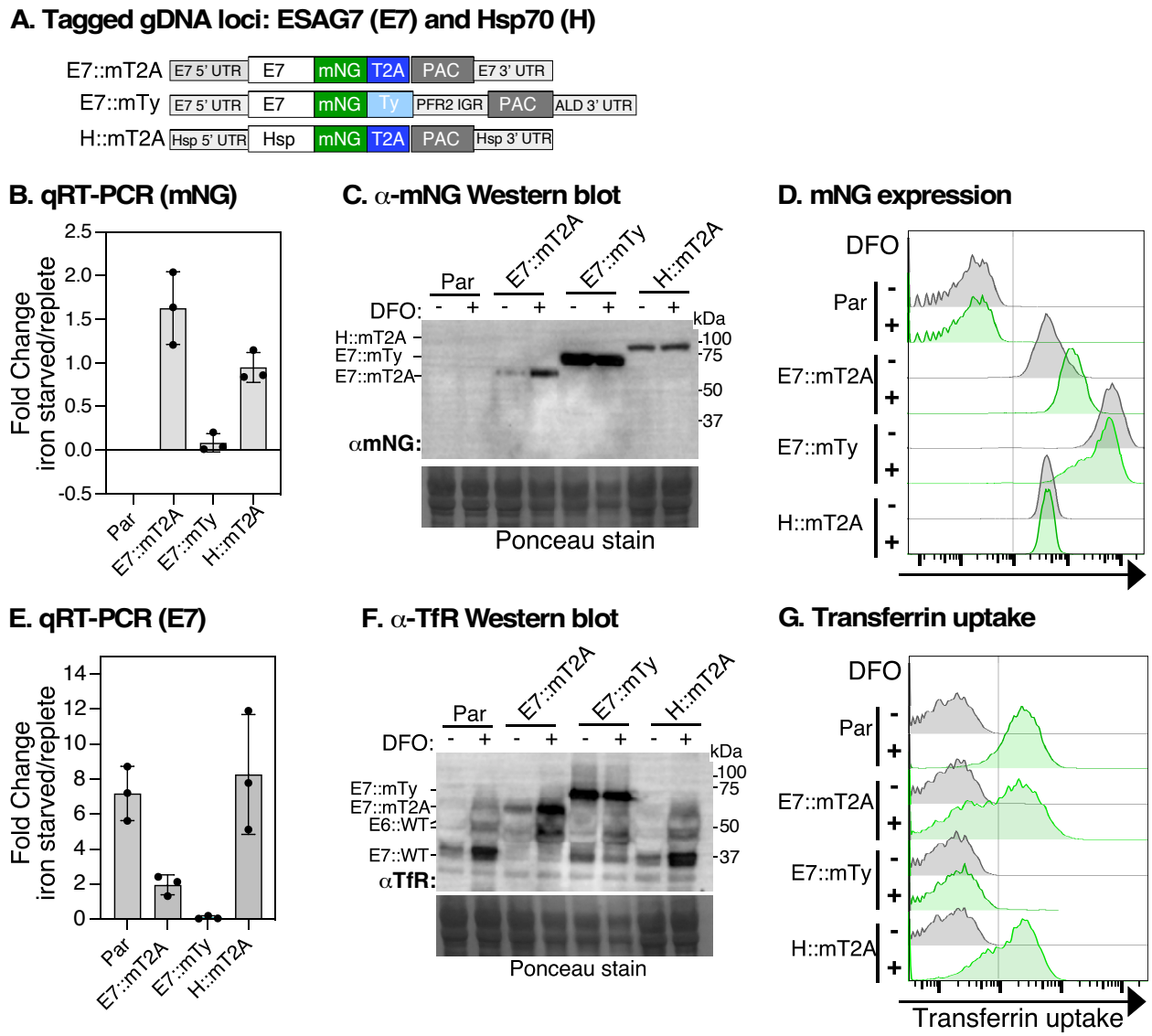


Fig. 4: Endogenous T2A tagging of ESAG7 maintains iron-dependent regulation.

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

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

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

890 **D.** Flow cytometry analysis of mNG expression in different cell lines under iron-replete (DFO, -)
891 and iron-starved (DFO, +) conditions. Histograms show fluorescence intensity indicating mNG
892 expression levels.
893 **E.** qRT-PCR analysis of *E7* transcript levels in parental (Par) cells or cells expressing E7::mT2A,
894 E7::mTy, or H::mT2A in iron-starved (DFO, +) conditions expressed as a ratio to control treatment
895 (iron-replete) values and normalised to *ZFP3* signal.
896 **F.** Western blot analysis of total protein using α-TfR antibody to detect total TfR expression in
897 whole cell extracts. TfR is a heterodimer of ESAG6 (E6) and ESAG7 (E7). In parental (Par) and
898 H::mT2A lanes, the predicted MW of endogenous E7 (E7::WT) is 37 kDa, while E6 (E6::WT)
899 migrates as two species: ~44 kDa (un-glycosylated) and >50 kDa (fully glycosylated) – see lanes
900 1 and 2 (par) and Lanes 7 and 8 (H::mT2A). The predicted MW of "cleaved" E7::mT2A, E7::mTy,
901 H::mT2A are 67 kDa, 74 kDa, and 110 kDa respectively (lanes 3, 4, 5, and 6). Par and H::mT2A
902 tagged cells serve as a control for DFO treatment. Ponceau-stained membrane shows loading
903 control. The blot is representative of n = 3 independent experiments from one clone.
904 **G.** Flow cytometry analysis of transferrin uptake in different cell lines under iron-replete (DFO, -)
905 and iron-starved (DFO, +) conditions. Histograms show fluorescence intensity indicating
906 transferrin uptake levels.
907

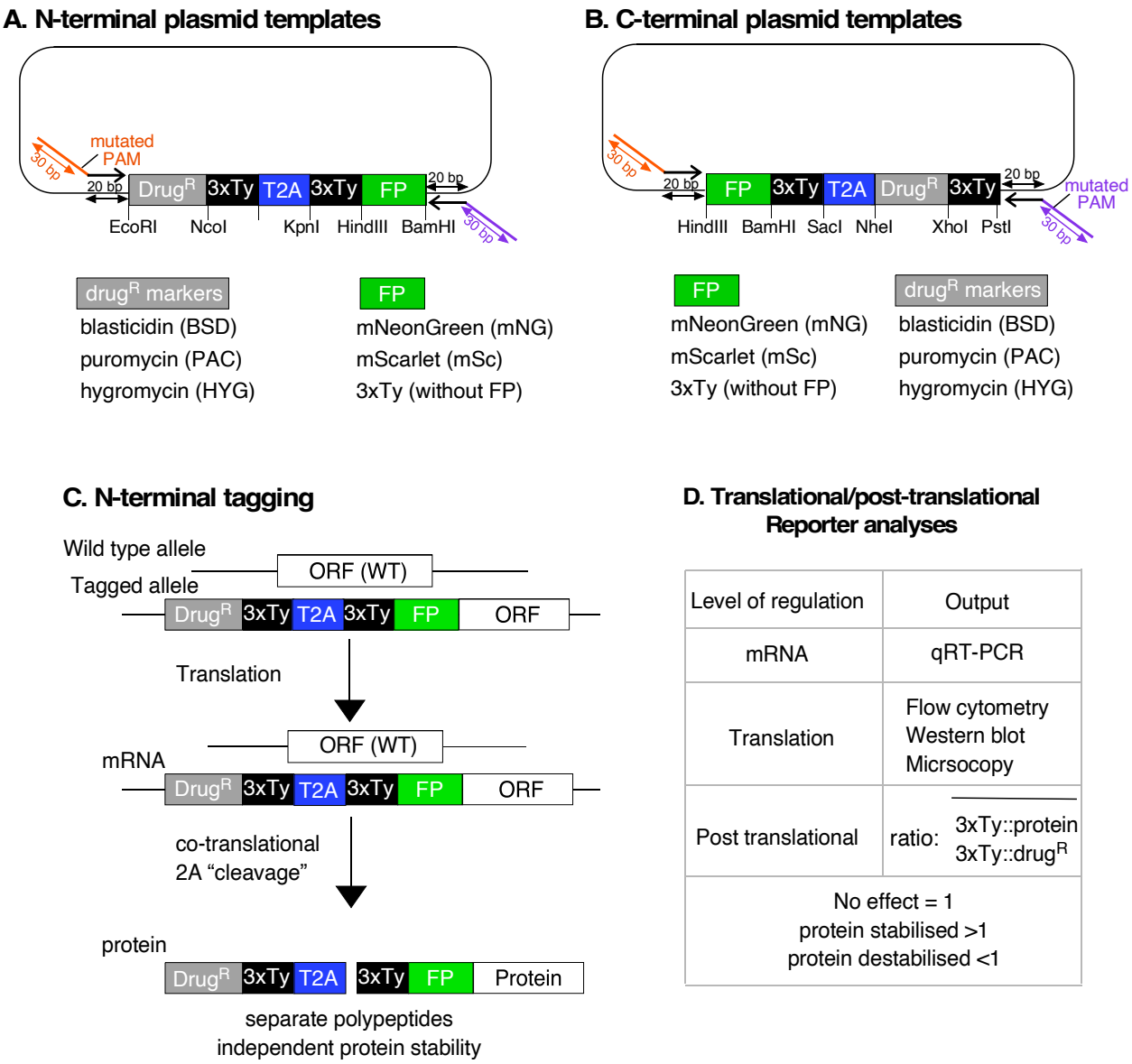


Fig. 5. Map of pRExT2A modular template plasmids and strategy for studying post-transcriptional and post translational gene regulation

The schematic illustrates the design of CRISPR/Cas9-mediated modular donor vectors for N-terminal (**A**) and C-terminal (**B**) T2A tagging of target genes in *T. brucei*. The repair cassette includes a drug resistance marker (drug^R, BSD, PAC, HYG), a 3xTy epitope, a T2A peptide, and a fluorescent protein tag (FP), such as mNeonGreen (mNG) or mScarlet (mSc). 3xTy epitope tags, drug^R, and FP are flanked by unique restriction sites (EcoRI, NcoI, HindIII, etc.) to facilitate cloning and customisation. Arrows indicate the positions of forward and reverse primers used for the amplification of repair templates. The forward and reverse primers contain a 20 bp binding site to donor plasmid and 30 bp homology arms matching the target sequence containing a mutated PAM site, as described in Fig.1. The plasmid backbone is pUC57-simple (GenScript), and the full list of sequences is deposited in Addgene.

C. The process of N-terminal tagging is depicted, where CRISPR/Cas9-mediated insertion of the repair cassette replaces one wild-type (WT) allele with a tagged allele at the 5' end of the target

924 gene. During translation, the T2A sequence facilitates co-translational cleavage, producing two
925 distinct polypeptides: one containing the 3xTy-tagged drug^R marker and another containing the
926 3xTy-tagged FP-tagged protein. This strategy ensures that both the drug^R and FP-tagged proteins
927 are independently stable, allowing for accurate analysis of gene regulatory mechanisms.
928 **D.** The table outlines the different levels of regulation that can be studied using this system. The
929 mRNA level can be analysed using qRT-PCR, allowing users to assess changes in mRNA
930 stability. At the translation level, protein expression can be monitored using flow cytometry or
931 Western blot to detect 3xTy-tagged FP-tagged protein and the 3xTy-tagged drug^R marker. Post-
932 translational regulation is assessed by determining the ratio of 3xTy-tagged protein to the 3x-
933 tagged drug^R marker. A ratio of 1 indicates no effect on stability, a ratio greater than 1 suggests
934 protein stabilisation, and a ratio less than 1 suggests protein destabilisation.
935

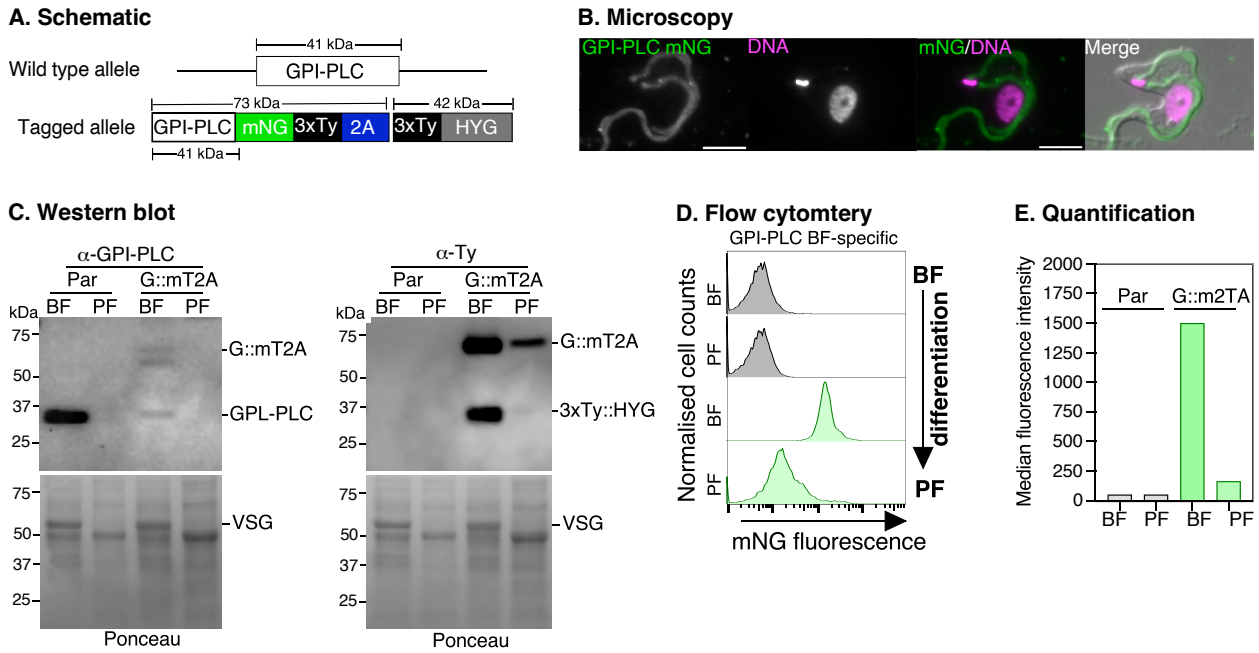


Fig. 6: Tagging of GPI-PLC preserves life cycle-dependent expression patterns during BF to PF differentiation.

A. Schematic of the GPI-PLC protein tagged with mNG::3xTy, fused via the T2A peptide to a puromycin resistance protein (3xTy::PAC) at its C-terminus. The approximate MW (in kDa) of the tagged components are indicated above the schematic.

B. Western blot analysis of GPI-PLC expression in parental (Par) and GPI-PLC::mT2A-tagged cells in BF and PF cells. α-Ty antibody (left) detected GPI-PLC::mT2A (G::mT2A) and 3xTy::PAC (PAC) proteins in BF cells, with signals at the expected size (~73 kDa). No signal was observed in PF cells. The α-GPI-PLC antibody confirms BF-specific GPI-PLC expression showing stoichiometric expression of endogenous (GPI-PLC) and tagged GPI-PLC (G:mT2A). Ponceau staining showing a band corresponding to the MW of variant surface glycoprotein (VSG) in BF but not in PF.

C. Localisation of GPI-PLC::mT2A protein visualised by live-cell epifluorescence microscopy (green), co-stained with Hoechst 33342 (magenta; DNA). Scale bar = 5 μM.

D. Flow cytometry analysis of mNG fluorescence in GPI-PLC::mT2A-tagged cells during in vitro differentiation of BF to PF stages. BF cells display high fluorescence, confirming GPI-PLC expression, whereas PF cells show minimal fluorescence. Parental cells (Par) serve as a negative control.

E. Quantification of flow cytometry data in D, showing median fluorescence intensity for mNG in BF and PF stages.

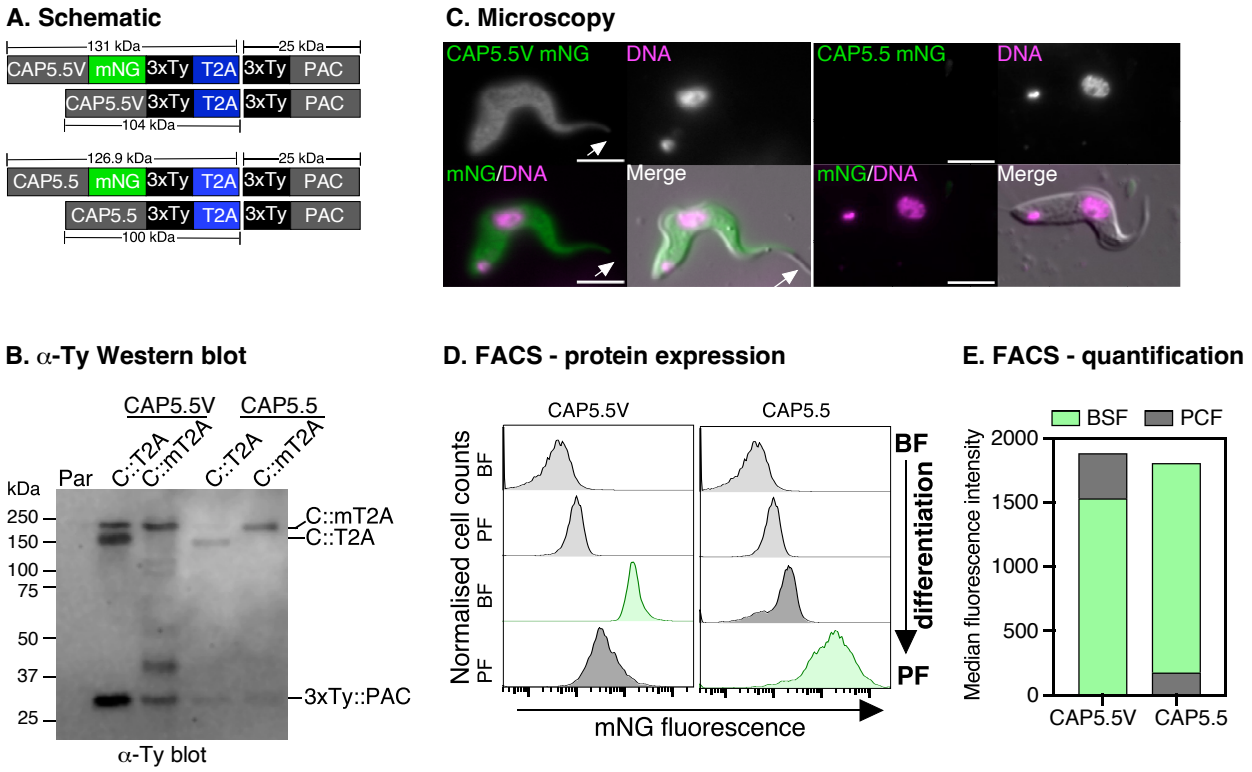


Fig. 7: Tagging of CAP5.5V and CAP5.5 confirms stage-specific expression during BF to PF differentiation.

A. Schematic of CAP5.5V and CAP5.5 proteins tagged with mNG::3xTy::2A::3xTy::PAC or 3xTy::2A::3xTy::PAC at their C-termini. The predicted MW of the tagged protein variants are indicated.

B. Western blot analysis using α-Ty antibody to detect tagged CAP5.5V (left) and CAP5.5 (right) proteins in bloodstream form (BF) cells. The α-Ty blot shows expression of both mT2A::3xTy and 3xTy – tagged variants and T2A-mediated release of 3xTy-tagged puromycin (3xTy::PAC) protein. Parental cells (Par) serve as negative control.

C. Localisation of CAP5.5V::mT2A (CAP5.5V::mNG) and CAP5.5::mT2A (CAP5.5::mNG) in BF stages visualised by live-cell epifluorescence microscopy (green), co-stained with Hoechst 33342 (magenta; DNA). CAP5.5 is below detection limit in BF stage cells. Arrowhead indicates no staining of the flagellum. Scale bar = 5 μM.

D. Flow cytometry analysis of mNG fluorescence in CAP5.5V and CAP5.5-tagged cells during differentiation from BF to PF life cycle stages.

E. Quantification of flow cytometry results in D, showing median fluorescence intensity for mNG in BF and PF stages.

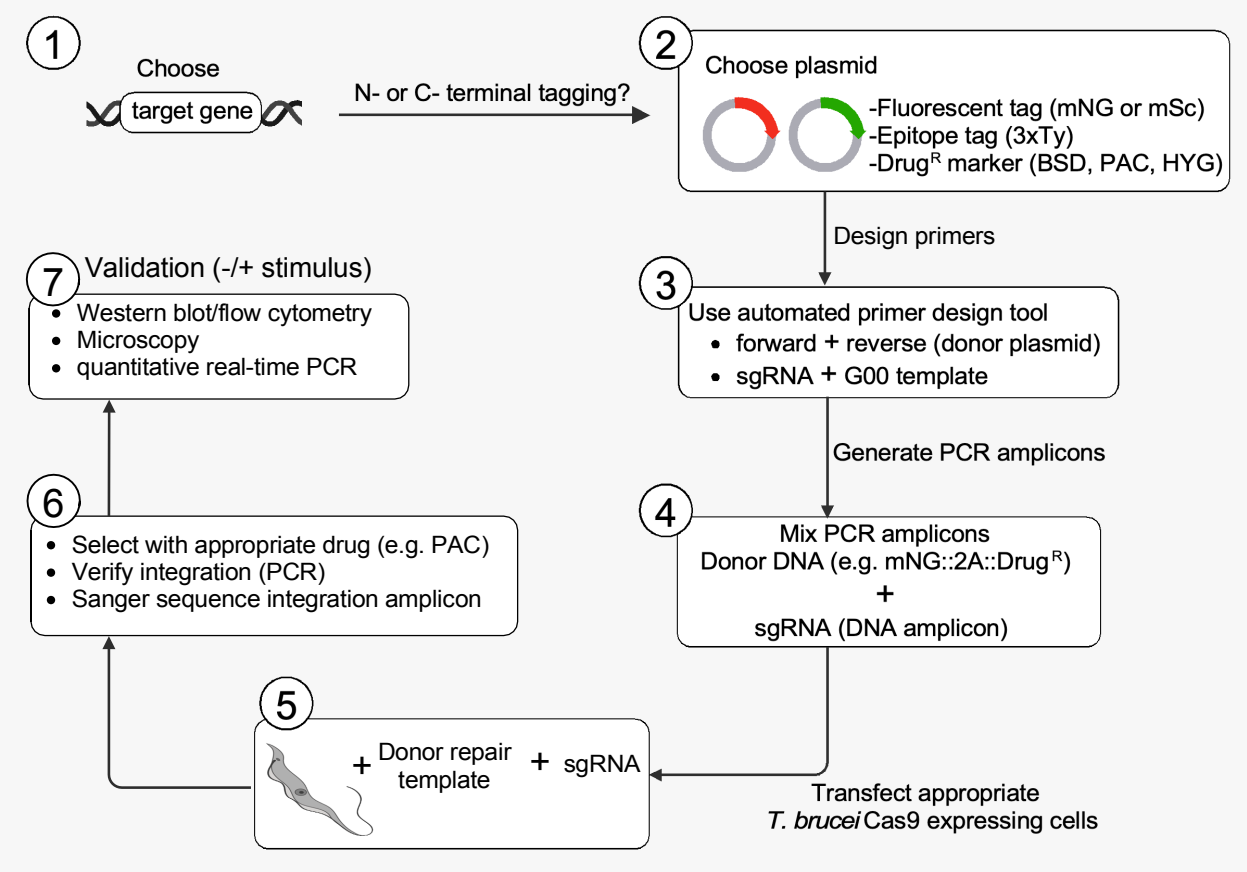


Fig. 8: Beginner's guide to using the pRExT2A tagging system: Start by selecting your target gene and deciding whether to tag it at the N- or C-terminus, or both. Next, choose the appropriate plasmid containing the desired fluorescent or epitope tag, along with a compatible drug-resistance marker (e.g., Blasticidin or Puromycin). Design primers using the automated tool to generate PCR amplicons for both the donor repair template and the sgRNA. Once the donor template and sgRNAs are ready, mix and transfect into *T. brucei* cells expressing Cas9. After transfection, apply drug selection to isolate successfully tagged cells. Validate the transgenic cells by performing PCR and Sanger sequencing to confirm correct integration. Finally, verify protein expression and localisation using Western blot, flow cytometry, or microscopy.