



RESEARCH ARTICLE

**UPDATE** **Subcellular protein localisation of *Trypanosoma brucei* bloodstream form-upregulated proteins maps stage-specific adaptations [version 2; peer review: 2 approved, 1 approved with reservations]**

Clare Halliday<sup>1</sup>, Samuel Dean<sup>2</sup>, Jack Daniel Sunter<sup>3</sup>, Richard J. Wheeler<sup>4</sup>

<sup>1</sup>Sir William Dunn School of Pathology, University of Oxford, Oxford, UK

<sup>2</sup>Division of Biomedical Sciences, Warwick Medical School, University of Warwick, Coventry, UK

<sup>3</sup>Department of Biological and Medical Sciences, Oxford Brookes University, Oxford, UK

<sup>4</sup>Nuffield Department of Medicine, University of Oxford, Oxford, UK

**v2** First published: 31 Jan 2023, 8:46  
<https://doi.org/10.12688/wellcomeopenres.18586.1>  
 Latest published: 23 May 2023, 8:46  
<https://doi.org/10.12688/wellcomeopenres.18586.2>

### Abstract

**Background:** Genome-wide subcellular protein localisation in *Trypanosoma brucei*, through our TrypTag project, has comprehensively dissected the molecular organisation of this important pathogen. Powerful as this resource is, *T. brucei* has multiple developmental forms and we previously only analysed the procyclic form. This is an insect life cycle stage, leaving the mammalian bloodstream form unanalysed. The expectation is that between life stages protein localisation would not change dramatically (completely unchanged or shifting to analogous stage-specific structures). However, this has not been specifically tested. Similarly, which organelles tend to contain proteins with stage-specific expression can be predicted from known stage specific adaptations but has not been comprehensively tested.

**Methods:** We used endogenous tagging with mNG to determine the sub-cellular localisation of the majority of proteins encoded by transcripts significantly upregulated in the bloodstream form, and performed comparison to the existing localisation data in procyclic forms.

**Results:** We have confirmed the localisation of known stage-specific proteins and identified the localisation of novel stage-specific proteins. This gave a map of which organelles tend to contain stage specific proteins: the mitochondrion for the procyclic form, and the endoplasmic reticulum, endocytic system and cell surface in the bloodstream form.

**Conclusions:** This represents the first genome-wide map of life cycle stage-specific adaptation of organelle molecular machinery in *T. brucei*.

### Open Peer Review

Approval Status

	1	2	3
<b>version 2</b> (update) 23 May 2023	 <a href="#">view</a>	 <a href="#">view</a>	
<b>version 1</b> 31 Jan 2023	 <a href="#">view</a>	 <a href="#">view</a>	 <a href="#">view</a>

1. **Christine Clayton** , Heidelberg University, Heidelberg, Germany
2. **Alena Zikova** , Czech Academy of Sciences, České Budějovice, Czech Republic
3. **Michael D. Urbaniak** , Lancaster University, Lancaster, UK

Any reports and responses or comments on the article can be found at the end of the article.

## Keywords

parasitology, trypanosoma, trypanosome, microscopy, fluorescent protein, tagging

**Corresponding author:** Richard J. Wheeler ([richard.wheeler@ndm.ox.ac.uk](mailto:richard.wheeler@ndm.ox.ac.uk))

**Author roles:** **Halliday C:** Data Curation, Investigation, Methodology, Validation, Writing – Review & Editing; **Dean S:** Conceptualization, Funding Acquisition, Methodology, Project Administration, Supervision, Validation, Writing – Review & Editing; **Sunter JD:** Conceptualization, Funding Acquisition, Methodology, Project Administration, Supervision, Validation, Writing – Review & Editing; **Wheeler RJ:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Methodology, Project Administration, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work was supported by Wellcome through a Sir Henry Dale Fellowship to RJW [211075, <https://doi.org/10.35802/211075>] and the biomedical resources grant which supported TrypTag [108445, <https://doi.org/10.35802/108445>].

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2023 Halliday C *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Halliday C, Dean S, Sunter JD and Wheeler RJ. **Subcellular protein localisation of *Trypanosoma brucei* bloodstream form-upregulated proteins maps stage-specific adaptations [version 2; peer review: 2 approved, 1 approved with reservations]** Wellcome Open Research 2023, 8:46 <https://doi.org/10.12688/wellcomeopenres.18586.2>

**First published:** 31 Jan 2023, 8:46 <https://doi.org/10.12688/wellcomeopenres.18586.1>

**REVISED Amendments from Version 1**

This version includes three primary changes: Firstly, including specific numbers and percentages for qualitative statements in the text, secondly, including gene set sizes in more figures and, thirdly, including example images of localisations which are potentially suspect - specifically soluble fluorescent protein (mNG) and background autofluorescence localisations. An additional data table has been added to the underlying data (at Zenodo) containing the underlying raw data for the analyses in [Figure 2](#) and [Figure 3](#), and the Zenodo description significantly updated to aid readers in finding and accessing the raw data.

**Any further responses from the reviewers can be found at the end of the article**

**Introduction**

*Trypanosoma brucei* is a unicellular eukaryotic parasite and, like any unicellular organism, adjusts its gene expression profile to adapt to different environments. As an obligate parasite, the environments it encounters are exclusively within the host and vector and gene expression profile changes give rise to the appropriate protein machinery to adapt the parasite to these niches. *T. brucei* has three main replicative life cycle stages: the procyclic form (PCF, fly midgut), the epimastigote form (EMF, fly salivary glands) and the bloodstream form (BSF, mammalian host bloodstream), although within these stages there is also additional specialisation<sup>1,2</sup>. The PCF and BSF are readily grown in culture.

The PCF and BSF have many well characterised differences, including the BSF VSG surface coat and associated expression machinery<sup>3</sup>, metabolic differences and associated remodelling of the mitochondrion<sup>4</sup>, morphology, and morphogenesis adaptations<sup>5,6</sup>, along with many more. However, genome-wide mapping of the global changes are broadly limited to gene expression level, most extensively determined at the mRNA level<sup>7-11</sup> which does not correlate fully with protein abundance<sup>12</sup>. Fewer studies consider later steps in protein production: translation (mRNA ribosome footprinting)<sup>7,11</sup> and protein abundance (quantitative proteomics)<sup>13-15</sup>. Despite the comparative ease of culturing PCFs and BSFs and the powerful reverse genetic tools available, a huge number of genes with evidence for BSF upregulation are not characterised.

Here, we aim to address this using subcellular protein localisations. We have demonstrated the power of this approach in PCFs with the TrypTag genome-wide protein localisation project<sup>16</sup>. This showed how informative localisation can be for holistic mapping of potential protein function, although naturally localisation does not determine specific molecular function. We also previously used high throughput tagging of BSF-upregulated genes to identify ESB1, necessary for transcription of the expression site containing the VSG gene along with expression site associated genes<sup>17</sup>. However, our previous analysis of these BSF localisations was minimal, aiming only to identify expression site body components. Here, we present analysis of an extended version of this BSF localisation dataset as both evidence for how BSFs are adapted relative to PCFs and as a resource for the research community.

**Methods****Cell culture**

Bloodstream form *Trypanosoma brucei brucei* strain Lister 427 pJ1339 was grown in HMI-9 at 37°C with 5% CO<sub>2</sub><sup>18</sup>, maintained in log phase growth and at less than ~2×10<sup>6</sup> cells/ml by regular subculture. To enable CRISPR/Cas9 genome modifications, this cell line expresses T7 RNA polymerase, Tet repressor, Cas9 nuclease and puromycin drug selectable marker<sup>17</sup> and were maintained with periodic drug selection using 0.2 µg/ml Puromycin Dihydrochloride. Culture density was measured with a CASY model TT cell counter (Roche Diagnostics) with a 60 µm capillary and exclusion of particles with a pseudo diameter below 2.0 µm.

**Electroporation and drug selection**

For endogenous tagging of a protein, electroporation was used to transfect *T. brucei* with two linear DNA constructs; one from which a CRISPR sgRNA is transiently expressed and one carrying the fluorescent protein and drug selectable marker which has homology arms allowing homologous recombination into the target locus. Constructs for endogenous N or C terminal tagging constructs were generated using long primer PCR from a pPOTv7 mNeonGreen (mNG) / blasticidin deaminase template, and PCR was used to generate DNA encoding sgRNA with a T7 promoter, both as previously described<sup>19,20</sup> (for primer sequences see *Underlying data*<sup>21</sup>).

For DNA encoding the drug selectable marker and fluorescent protein, 0.2 mM dNTPs, 30 ng pPOT plasmid, 2 µM gene-specific forward and reverse primer and 1 unit HiFi Polymerase (Roche) were mixed in 1× HiFi reaction buffer with MgCl<sub>2</sub> and 3% v/v DMSO, in 50 µl total volume. PCR cycling conditions were 5 min at 94°C followed by 40 cycles of 30 s at 94°C, 30 s at 65°C, 2 min 15 s at 72°C followed by a final elongation step for 7 min at 72°C on a SimpliAmp Thermal Cycler (ThermoFisher).

For DNA encoding sgRNAs, 0.2 mM dNTPs, 2 µM of sgRNA scaffold primer (aaaagcaccgactcgtgccacttttcaagtgataacg-gactagccttatttaactgctatttctagctctaaac) and gene-specific primer and 1 unit HiFi Polymerase were mixed in 1× HiFi reaction buffer with MgCl<sub>2</sub>, 50 µl total volume. PCR cycling conditions were 30 s at 98°C followed by 35 cycles of 10 s at 98°C, 30 s at 60°C, 15 s at 72°C on a SimpliAmp Thermal Cycler. 2 µl of each reaction were run on a 2% agarose gel to check for the presence of a product of the expected size. For gel images, please see the associated Zenodo deposition<sup>21</sup>.

~5 µg of DNA from the PCRs was purified by phenol chloroform extraction, resuspended in 10 µl water, then mixed with approximately 3×10<sup>7</sup> cells resuspended in 100 µl of Roditi Tb-BSF buffer<sup>22</sup>. Transfection was carried out using program X-001 of the Amaxa Nucleofector IIB (Lonza) electroporator in 2 mm gap cuvettes. Following electroporation, cells were transferred to 10 ml pre-warmed HMI-9 for 6 h then 5.0 µg/ml Blasticidin S Hydrochloride added to select for cells with successful construct integration. Healthy resulting populations were maintained with periodic drug selection using 0.2 µg/ml Puromycin Dihydrochloride and 5.0 µg/ml Blasticidin S Hydrochloride.

## Selection of genes for tagging

BSF tagging was carried out in the *T. brucei* Lister 427 cell line, and we considered genes for tagging if they had a syntenic ortholog in *T. brucei* TREU927. Genes were selected for tagging as described in the main text using TrypTag PCF protein localisation data available up to 12<sup>th</sup> March 2018 and TriTrypDB version 36, with the following specific exclusion criteria to avoid tagging of large well-known gene families and genes encoding GPI-anchored proteins known to be refractory to N and C terminal tagging. VSG, the major BSF surface coat protein was excluded by removing known (named) VSG genes and pseudogenes. In the interest of unbiased analysis, we ensured surface coat proteins characteristic of other life cycle stages were also excluded: EP procyclins, also called procyclic acidic repetitive proteins (PARPs), and *brucei* alanine rich proteins (BARPs). Known (named) invariant surface glycoproteins (ISGs) were excluded, with the exception of tagging controls ISG65 and GPI-PLC, and VSG expression site associated genes and related genes (ESAGs and GRESAGs) were excluded. Finally ribosomal proteins, which we deemed unlikely to be of interest, were excluded.

## Light microscopy

Cells were prepared for light microscopy by centrifugation to remove medium, followed by resuspension in FCS-free HMI-9 containing 1 µg/mL Hoechst 33342 before a second centrifugation and resuspension in a small volume (~20 µl) of FCS-free HMI-9. An equal volume of 0.04% (v/v) formaldehyde in FCS free HMI-9 was added to lightly fix the cells<sup>17,23</sup>. Images were captured on a DM5500 B (Leica) upright wide-field epifluorescence microscope using a plan apo NA/1.4 63× phase contrast oil-immersion objective (Leica, 15506351) and a Neo v5.5 (Andor) sCMOS camera using MicroManager (version 1.4.18)<sup>24</sup>.

## Statistics

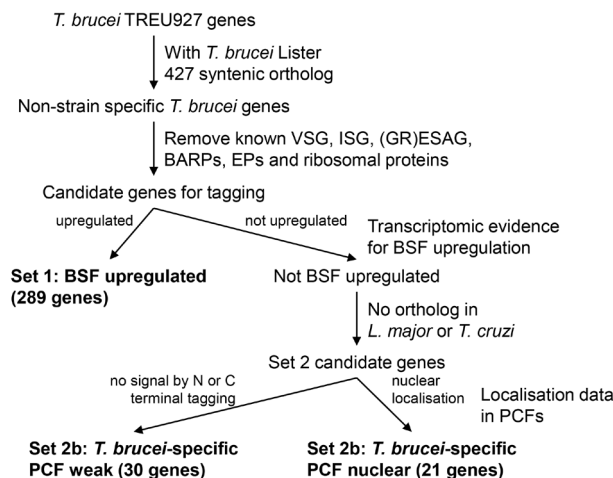
Statistical significance of change in localisation annotation terms usage for the PCF and BSF upregulated gene sets was

evaluated using the Chi squared test (using Excel version 2210, Microsoft), taking the annotation term usage in the genome-wide PCF set as the null hypothesis. Fold change in individual term usage was calculated as the ratio of term count in the PCF or BSF upregulated set to the term count in the genome-wide PCF set, eg. count of axoneme annotation terms in the BSF upregulated gene set divided by count of axoneme annotation terms genome-wide in PCFs. This is an approximation for BSFs, as we do not know the genome-wide term usage in BSFs. Error was estimated using the standard error of proportion (SEP) for each annotation term (using Excel). Fold change in term usage was normalised (and SEP scaled appropriately) to the total number of annotation terms in each set, such that no bias in usage between sets is unity.

## Results and discussion

To enrich for proteins likely to have BSF-specific functions, we devised three gene tagging sets based on data available at the time (Figure 1). Set 1) 289 genes with mRNAs upregulated in BSFs, based primarily on mRNAseq from 7 but manually incorporating some genes identified as strongly upregulated in 8–11 not in 7. Set 2) *T. brucei*-specific genes (defined as those which lack both an *L. major* Friedlin and *T. cruzi* Brener non-Esmarello ortholog) not already included in Set 1, which met one of two criteria based on TrypTag PCF tagging data available at the time: Set 2a) the 30 genes that had failed to give a convincing signal above background by both N and C terminal tagging, and Set 2b) the 21 genes which had a nucleoplasm or nucleolar localisation. The former were selected to test whether lack of PCF signal correlated with BSF stage-specific expression, and the latter as candidates for *T. brucei*-specific BSF nuclear structure adaptation potentially associated with antigenic variation/variant surface glycoprotein expression.

We prioritised N terminal tagging because this preserves the 3' untranslated region (UTR), suspected to confer most gene regulation in trypanosomes<sup>25</sup>. However, when a protein had a



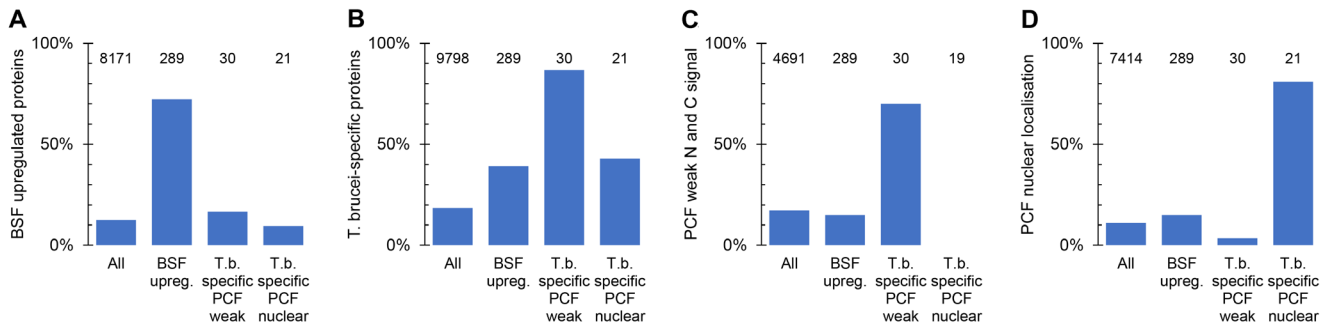
**Figure 1.** Flowchart for selection of genes for tagging in BSFs.

predicted N terminal signal peptide C terminal tagging was instead necessary. If we failed to generate a drug resistant population, we repeated construct generation and transfection at least once. The final success rate generating cell lines (for full listing see *Underlying data*<sup>21</sup>) was 72.9% (Figure 3A), of which 76.6% had signal we manually classified as unlike background fluorescent signal (Figure 3B) – i.e. a convincing subcellular localisation.

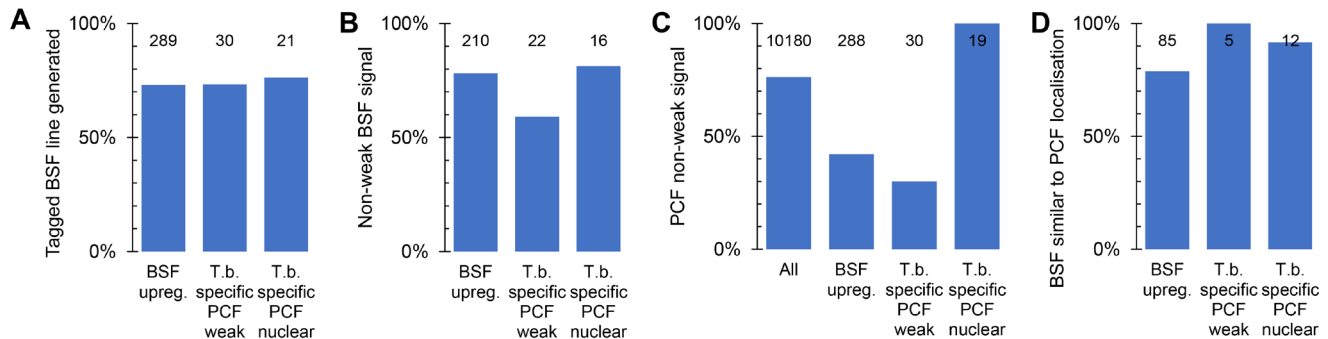
For the final analysis of these localisations, we re-analysed the gene sets based on the entire TrypTag PCF localisation dataset<sup>16</sup> and TriTrypDB version 59<sup>26</sup> (Figure 2. There were some changes; altered OrthoMCL sensitivity due to addition of new genomes (Figure 2B), additional PCF tagging repeats providing

a strong convincing localisation where only weak signal was previously observed (Figure 2C), and changed PCF localisation annotation (e.g. from nucleoplasm to nuclear envelope, Figure 2D). We also defined a final criterion for upregulation in the BSF: transcripts significantly upregulated ( $p < 0.05$ , Student's T test) by mRNAseq in the BSF relative to the PCF (data from 7). However, overall, the gene sets well reflect their original purpose.

We observed convincing fluorescent signal in BSFs for many (164/289, 56.7%) tagged proteins in Set 1 (upregulated in BSFs at the mRNA level, Figure 3B). In this gene set, disproportionately many genes (39.1% vs. 18.4% genome-wide) were also *T. brucei*-specific (Figure 2B), and dispro-



**Figure 2. Post hoc analysis of the target gene sets for BSF tagging.** Bar charts showing the proportion of genes in a gene set (x axis) which meet a particular criterion. Total number of genes in each set is shown above each bar. **A.** Proportion of genes with at least 2.5-fold upregulated mRNA and  $p < 0.05$  (two-tailed T test) from 7, for each target gene set; BSF upregulated, *T. brucei*-specific with PCF weak signal by N and C terminal tagging and *T. brucei*-specific proteins which localise to the nucleus in PCFs, in comparison to all *T. brucei* genes. **B.** Proportion of genes with no *L. major* and no *T. cruzi* ortholog in each target gene set. **C.** Proportion of genes with N and C terminal tagging data in PCFs from the TrypTag project for which both termini had weak, i.e. no strong localisation to an identifiable organelle. **D.** Proportion of genes annotated as localising to the nucleus, nucleoplasm or nucleolus by either N or C terminal tagging in PCFs from the TrypTag project. In each graph, the number of genes for which data is available in each group is shown at the top of each column.



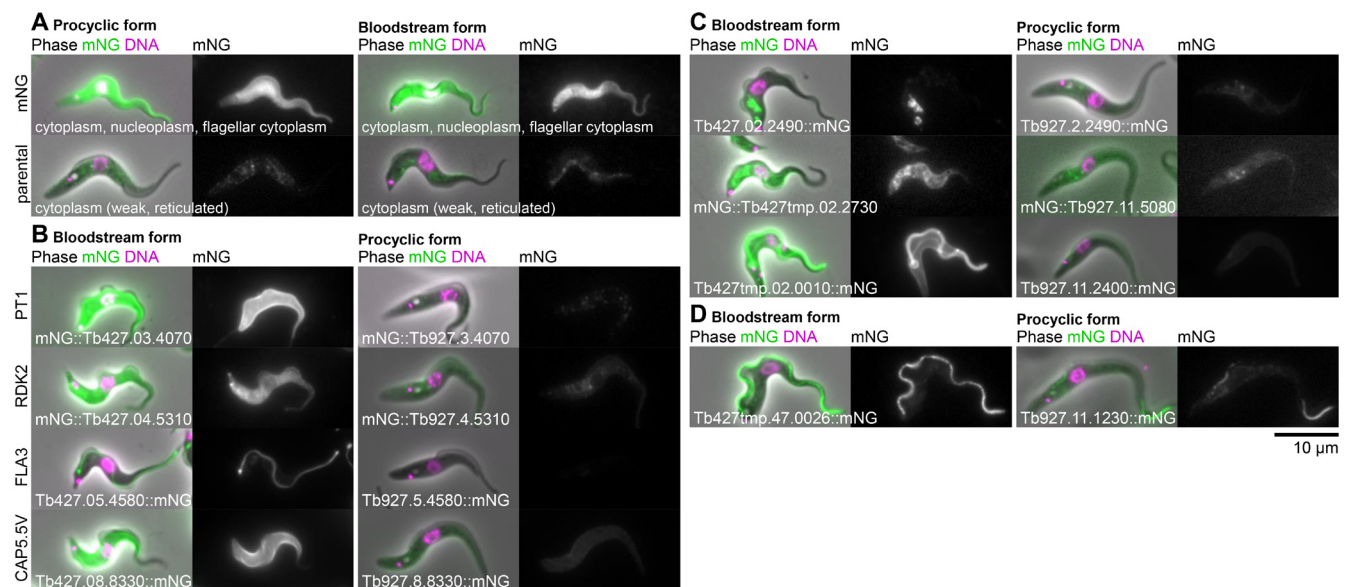
**Figure 3. Success rates generating BSF localisations.** Bar charts showing the proportion of genes in a gene set (x axis) which met a tagging success criterion. Total number of genes in each set is shown above each bar. **A.** Proportion of cell lines generated for each target gene set: Set 1, BSF upregulated; set 2a, *T. brucei*-specific with PCF weak signal by N and C terminal tagging; set 2b, *T. brucei*-specific proteins which localise to the nucleus in PCFs. **B.** Proportion of BSF cell lines generated for each target gene set which had a weak signal, i.e. no strong localisation to an identifiable organelle in the BSF. **C.** Proportion of each target gene set for which strong localisation to an identifiable organelle was observed for either N or C terminal tagging in PCFs, in comparison to all genes with PCF data. Data from the TrypTag project. **D.** The proportion of BSF cell lines for each target gene set with strong localisation to an identifiable organelle which gave a similar localisation to either N or C terminal tagging in PCFs. In each graph, the number of genes for which data is available in each group is shown at the top of each column.

portionately few (42.0% vs. 76.3% genome-wide) had no convincing above-background localisation observed in the PCF (Figure 2C). We also observed a convincing fluorescent signal in BSFs for many (13/30, 43.3%) in Set 2a (*T. brucei*-specific genes with no detectable PCF signal, Figure 3B). Lack of fluorescent signal in the PCF tagging previously raised our suspicions that these genes may not be expressed in this life cycle stage, never expressed, or encode a non-functional, and therefore degraded, protein product. Similarly, failure to generate a PCF tagged cell line may indicate inaccurate sequence data for that locus or that the drug selectable marker cannot be expressed from that locus. This was an acute concern when the gene was *T. brucei* specific and therefore had no evidence from evolutionary conservation for being functional. Our BSF localisation provides evidence that many of these genes (67/161, 41.6%) encode an expressed and likely functional protein (on the basis that the proteins often targeted to a specific organelle), supporting proteomic analyses<sup>15</sup>. As would be expected, fluorescent signal in a tagged cell line therefore broadly correlates with mRNA abundance across life cycle stages and failure to observe a convincing localisation in PCFs is, as we previously proposed<sup>16</sup>, at least partially predictive of a stage specific protein expression.

As described above, with the exception of Set 2b, the set of *T. brucei* specific nuclear genes which were selected based on

a specific PCF localisation, our BSF tagging was of proteins disproportionately more likely to have no detected signal from PCF tagging (Figure 3C). However, when a PCF localisation was available it was likely to be similar to the BSF localisation we observed, overall ~85% were manually classified as similar (Figure 3D). When dissimilar, the localisation observed in either the PCF or BSF was typically either a weak cytoplasmic signal or a cytoplasm, nuclear lumen and flagellar cytoplasm localisation (examples shown in Figure 4A). The former is simply background autofluorescence signal. The latter is the localisation we observed in PCFs for mNG when not fused to a protein. As we previously described for PCF tagging<sup>27</sup>, these can arise from frame shifts, likely originating from stochastic errors in synthesis of the primers for tagging. Alternatively, they may be poorly tolerated fusion proteins – truncated or partially degraded leading to expression of effectively mNG alone. Overall, we therefore conclude that the vast majority of proteins differ only in expression level and not localisation. One, however, featured a clear change; see below.

For Set 1, the set of BSF upregulated genes, whether or not a PCF localisation was visible the BSF localisation gave a much stronger signal – detectable as we used the same microscope, camera and image processing settings for PCFs and BSFs, making signal intensity in the images approximately

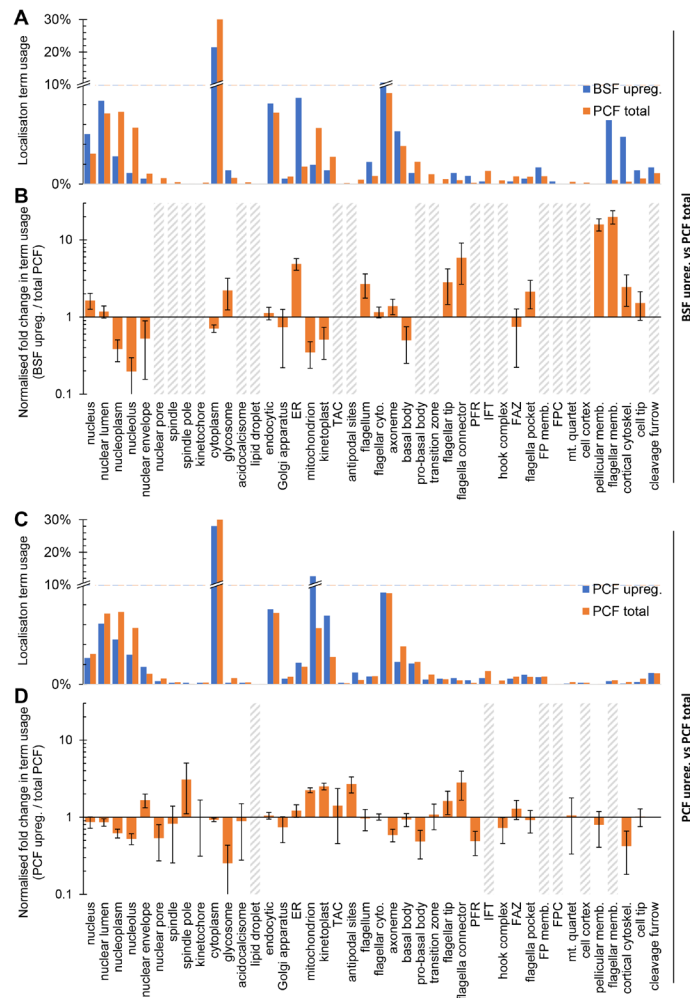


**Figure 4. Example subcellular localisations of BSF-specific or strongly upregulated proteins. A.** Examples of two potentially artefactual/spurious proteins localisations: cytoplasm, nucleoplasm and flagellar cytoplasm (similar to mNG alone) and weak reticulated cytoplasm (similar to background autofluorescence). On the left, PCFs expressing either mNG or no fluorescent protein (parental) and equivalent suspect localisations in BSF on the right. **B.** Known or expected BSF-specific proteins, showing the BSF localisation in *T. brucei* Lister 427 on the left and the localisation of the *T. brucei* TREU927 ortholog in PCFs from the TrypTag project on the right. For each cell line, an overlay of the phase contrast, mNG fluorescence and the Hoechst 33342 DNA stain is shown on the left and the mNG fluorescence alone in greyscale on the right. The gene ID and mNG fusion is shown in the bottom left. BSF and PCF mNG fluorescence are shown at approximately equal contrast levels to enable comparison of protein levels. **C.** Examples of previously uncharacterised BSF-specific proteins localising to (from top to bottom) the endocytic system, the endoplasmic reticulum and the pellicular and flagellar membranes. **D.** The only identified example of a protein whose subcellular localisation differs between BSFs and PCFs and was not a cytoplasm, nucleoplasm and flagellar cytoplasm or weak reticulated in either BSFs or PCFs. This protein localised to the whole axoneme in BSFs and concentrated in the distal axoneme in PCFs.

quantitative. This includes proteins known or expected to be BSF-upregulated: pyruvate transporter 1, PT1<sup>28</sup>; repressor of differentiation kinase 2, RDK2<sup>29</sup>; flagellum adhesion protein 3, FLA3<sup>30</sup>; and cytoskeleton associated protein CAP5.5V<sup>31</sup> (Figure 4B). However, it also includes novel or uncharacterised proteins localising to a range of different organelles (examples in Figure 4C). We also noted one clear example where protein localisation differed between the PCF and BSF. Tb927.11.1230 and its syntenic ortholog Tb427tmp.47.0026 localised to the distal axoneme (occasionally with weak proximal signal) in PCFs and the entire axoneme in BSFs (Figure 4D). PCF to BSF localisation differences have been previously observed, for example MCP6 and  $\alpha$ -KDE1<sup>32,33</sup>, but

most notably the Tb927.11.1230/Tb427tmp.47.0026 localisation change is comparable to that of the flagellar protein FLAM8 (flagellar member 8)<sup>34</sup>.

We noted that BSF-upregulated proteins often localised to membranous structures - the pellicular or flagellar membrane, the endoplasmic reticulum or the endocytic system (Figure 4B,C). We therefore tested for a bias in localisation annotation term usage relative to genome-wide usage in PCFs. Taking only the target genes for BSF tagging not selected based on a nuclear PCF localisation, i.e. excluding Set 2b, there was indeed a significant bias in term usage ( $p < 10^{-30}$ , chi-squared test). Normalised fold-change in usage of annotation terms revealed a strongly



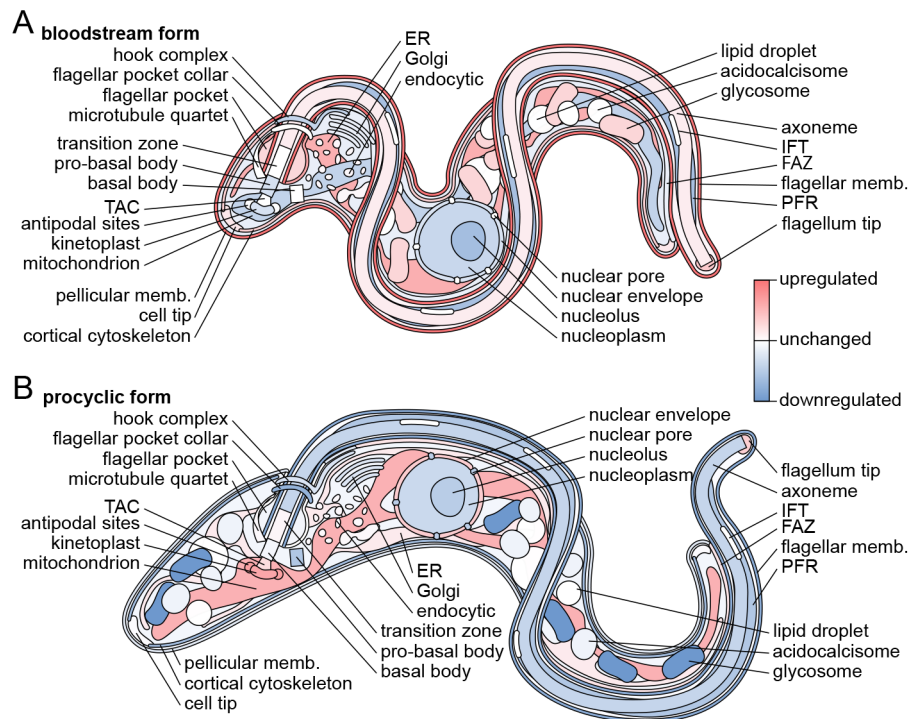
**Figure 5. Stage-specific organelle adaptation mapped using localisation term usage.** **A.** Localisation annotation term usage, as the proportion of all annotation terms used localisations, comparing all PCF (N and C terminal tagging) localisation terms to all BSF localisations described here, excluding the target gene set 2b; *T. brucei*-specific nuclear localising proteins. All localisation annotations for N and/or C terminal tagging, whichever are available, so long as they did not have the 'weak' or '<10%' modifiers. **B.** The data in **A**, except plotted as the ratio of term usage in BSF upregulated vs. total PCF, normalised to number of annotation terms in the BSF set. Error bars represent standard error of proportion. Grey hatched bars indicate too few (<3) BSF upregulated protein localisations for accurate fold change calculation. **C.** Analogous analysis of PCF upregulated genes from TrypTag data: Localisation annotation term usage, as the proportion of all annotation terms used for non-weak localisation, comparing all PCF localisation terms with those for proteins encoded by genes significantly upregulated at the mRNA level in PCFs. **D.** The data in **C**, except plotted as the ratio of term usage in PCF upregulated vs. PCF total term usage, normalised to number of terms in the PCF set. Error bars represent standard error of proportion. Grey hatched bars indicate too few (<3) PCF upregulated protein localisations for accurate fold change calculation.

disproportionately high usage of terms associated with the surface membrane and the endo/exocytic system (pellicular and flagellar membrane, ER and endocytic). There were also weaker biases in BSFs for 1) general (nucleus, nuclear lumen) rather than specific (nucleoplasm, nucleolus) nuclear localisation annotations, 2) fewer mitochondrion and kinetoplast annotations, 3) more glycosome terms, and 4) more flagellum tip and flagellar connector-like<sup>5,35</sup> annotation terms (Figure 5A,B). The BSF cell surface therefore has the greatest adaptation between BSFs and PCFs, with this change plausibly supported and/or maintained by changes in the ER and endocytic system.

The converse analysis, taking genes upregulated in the procyclic form ( $p < 0.05$ , Student's T test, by mRNAseq in the PCF relative to BSF, data from 7) and analysing localisation annotation term usage relative to genome-wide usage in PCFs also revealed a significant change ( $p < 10^{-30}$ , chi-squared test) in term usage, reflecting adaptation in the PCF. We identified 1) disproportionately high usage of mitochondrion and kinetoplast terms, 2) high usage of flagellar tip and flagellar connector terms, and 3) few glycosome terms. This speaks to the known upregulation of oxidative phosphorylation

(mitochondrial) relative to glycolysis (glycosomal) as the major ATP source in procyclic form and adaptation of the flagellum tip likely linked with new flagellum outgrowth<sup>5</sup>, but limited other changes (Figure 5C,D).

In conclusion, we have mapped which organelles contain proteins upregulated in the *T. brucei* BSF and PCF life cycle stages (summarised in Figure 6), thus mapping where the molecular machinery responsible for their stage-specific adaptations likely act in the cell. This includes uncharacterised proteins with little or no bioinformatic insight into likely function. Lack of fluorescent signal by endogenous tagging in the PCF was often predictive of BSF expression, confirming the power of the TrypTag genome-wide protein localisation resource as a protein expression level resource. We also showed that it is likely that a large majority of *T. brucei* proteins, when expressed, have similar localisations in BSFs and PCFs – the dominant adaptive process therefore appears to be change in expression level rather than change in localisation. We suggest that this also likely applies to other life cycle stages and the different life cycle stages of other trypanosomatid parasites.



**Figure 6. Diagrammatic summary of stage-specific organellar adaptation.** Diagrammatic representation of G1 *T. brucei* cell structure with organelles colour-coded by whether they contain disproportionately many genes with stage-specific expression level. **A.** Summary of organelles tending to contain proteins with BSF-specific up or downregulation, data from Figure 5B. **B.** Summary of organelles tending to contain proteins with PCF-specific up or downregulation, data from Figure 5C. This figure is an original figure produced by the authors for this article.



## Data availability

### Underlying data

Zenodo: *Trypanosoma brucei* bloodstream form tagging: Targeted subcellular protein localisation. <https://zenodo.org/record/7418663><sup>21</sup>.

This project contains the following underlying data:

- Computer code, 96 well plate layout information and primer sequences, along with an index of further Zenodo

DOIs containing images of gels of PCR products, raw and processed microscopy images and localisation annotations

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/) (CC-BY 4.0).

## Acknowledgements

We would like to thank Keith Gull and the TrypTag principal investigators team for their support in TrypTag-associated projects.

## References

- Dyer NA, Rose C, Egeh NO, *et al.*: **Flying tryps: survival and maturation of trypanosomes in tsetse flies.** *Trends Parasitol.* 2013; **29**(4): 188–196. [PubMed Abstract](#) | [Publisher Full Text](#)
- Szöör B, Silvester E, Matthews KR: **A Leap Into the Unknown - Early Events in African Trypanosome Transmission.** *Trends Parasitol.* 2020; **36**(3): 266–278. [PubMed Abstract](#) | [Publisher Full Text](#)
- Budzak J, Rudenko G: **Pedal to the Metal: Nuclear Splicing Bodies Turbo-Charge VSG mRNA Production in African Trypanosomes.** *Front Cell Dev Biol.* 2022; **10**: 876701. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Zíková A: **Mitochondrial adaptations throughout the *Trypanosoma brucei* life cycle.** *J Eukaryot Microbiol.* 2022; **69**(6): e12911. [PubMed Abstract](#) | [Publisher Full Text](#)
- Hughes L, Towers K, Starborg T, *et al.*: **A cell-body groove housing the new flagellum tip suggests an adaptation of cellular morphogenesis for parasitism in the bloodstream form of *Trypanosoma brucei*.** *J Cell Sci.* 2013; **126**(Pt 24): 5748–5757. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wheeler RJ, Scheumann N, Wickstead B, *et al.*: **Cytokinesis in *Trypanosoma brucei* differs between bloodstream and tsetse trypomastigote forms: implications for microtubule-based morphogenesis and mutant analysis.** *Mol Microbiol.* 2013; **90**(6): 1339–1355. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jensen BC, Ramasamy G, Vasconcelos EJR, *et al.*: **Extensive stage-regulation of translation revealed by ribosome profiling of *Trypanosoma brucei*.** *BMC Genomics.* 2014; **15**(1): 911. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jensen BC, Sivam D, Kifer CT, *et al.*: **Widespread variation in transcript abundance within and across developmental stages of *Trypanosoma brucei*.** *BMC Genomics.* 2009; **10**: 482. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Naguleswaran A, Doiron N, Roditi I: **RNA-Seq analysis validates the use of culture-derived *Trypanosoma brucei* and provides new markers for mammalian and insect life-cycle stages.** *BMC Genomics.* 2018; **19**(1): 227. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Siegel TN, Hekstra DR, Wang X, *et al.*: **Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites.** *Nucleic Acids Res.* 2010; **38**(15): 4946–4957. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Vasquez JJ, Hon CC, Vanselow JT, *et al.*: **Comparative ribosome profiling reveals extensive translational complexity in different *Trypanosoma brucei* life cycle stages.** *Nucleic Acids Res.* 2014; **42**(6): 3623–3637. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Buccitelli C, Selbach M: **mRNAs, proteins and the emerging principles of gene expression control.** *Nat Rev Genet.* 2020; **21**(10): 630–644. [PubMed Abstract](#) | [Publisher Full Text](#)
- Gunasekera K, Wüthrich D, Braga-Lagache S, *et al.*: **Proteome remodelling during development from blood to insect-form *Trypanosoma brucei* quantified by SILAC and mass spectrometry.** *BMC Genomics.* 2012; **13**: 556. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Butter F, Bucerius F, Michel M, *et al.*: **Comparative proteomics of two life cycle stages of stable isotope-labeled *Trypanosoma brucei* reveals novel components of the parasite's host adaptation machinery.** *Mol Cell Proteomics.* 2013; **12**(1): 172–179. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Urbaniak MD, Guther ML, Ferguson MA: **Comparative SILAC proteomic analysis of *Trypanosoma brucei* bloodstream and procyclic lifecycle stages.** *PLoS One.* 2012; **7**(5): e36619. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Billington K, Halliday C, Madden R, *et al.*: **Genome-wide subcellular protein localisation in the flagellate parasite *Trypanosoma brucei*.** *bioRxiv.* 2022; 2022.06.09.495287. [Publisher Full Text](#)
- López-Escobar L, Hänisch B, Halliday C, *et al.*: **Stage-specific transcription activator ESB1 regulates monoallelic antigen expression in *Trypanosoma brucei*.** *Nat Microbiol.* 2022; **7**(8): 1280–1290. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hirumi H, Hirumi K: **Continuous cultivation of *Trypanosoma brucei* blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers.** *J Parasitol.* 1989; **75**(6): 985–989. [PubMed Abstract](#) | [Publisher Full Text](#)
- Beneke T, Madden R, Makin L, *et al.*: **A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids.** *R Soc Open Sci.* 2017; **4**(5): 170095. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dean S, Sunter J, Wheeler RJ, *et al.*: **A toolkit enabling efficient, scalable and reproducible gene tagging in trypanosomatids.** *Open Biol.* 2015; **5**(1): 140197. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Halliday C, Dean S, Sunter JD, *et al.*: ***Trypanosoma brucei* bloodstream form tagging: Targeted subcellular protein localisation.** [Dataset] *Zenodo.* 2022. <http://www.doi.org/10.5281/zenodo.7418663>
- Schumann Burkard G, Jutzi P, Roditi I: **Genome-wide RNAi screens in bloodstream form trypanosomes identify drug transporters.** *Mol Biochem Parasitol.* 2011; **175**(1): 91–94. [PubMed Abstract](#) | [Publisher Full Text](#)
- Dean S, Sunter J: **Light Microscopy in Trypanosomes: Use of Fluorescent Proteins and Tags.** *Methods Mol Biol.* 2020; **2116**: 367–383. [PubMed Abstract](#) | [Publisher Full Text](#)
- Edelstein A, Amodaj N, Hoover K, *et al.*: **Computer control of microscopes using  $\mu$ Manager.** *Curr Protoc Mol Biol.* 2010; **Chapter 14**: Unit14.20. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- De Gaudenzi JG, Noé G, Campo VA, *et al.*: **Gene expression regulation in trypanosomatids.** *Essays Biochem.* 2011; **51**: 31–46. [PubMed Abstract](#) | [Publisher Full Text](#)
- Amos B, Aurrecochea C, Barba M, *et al.*: **VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center.** *Nucleic Acids Res.* 2022; **50**(D1): D898–D911. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Billington K, Halliday C, Madden R, *et al.*: **TrypTag: Genome-wide subcellular protein localisation in *Trypanosoma brucei*.** [Publisher Full Text](#)
- Sanchez MA: **Molecular identification and characterization of an essential pyruvate transporter from *Trypanosoma brucei*.** *J Biol Chem.* 2013; **288**(20): 14428–14437. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jones NG, Thomas EB, Brown E, *et al.*: **Regulators of *Trypanosoma brucei* cell cycle progression and differentiation identified using a kinome-wide RNAi screen.** *PLoS Pathog.* 2014; **10**(1): e1003886. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Woods K, Nic a'Bhaird N, Dooley C, *et al.*: **Identification and characterization of a stage specific membrane protein involved in flagellar attachment in *Trypanosoma brucei*.** *PLoS One.* 2013; **8**(1): e52846. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Olego-Fernandez S, Vaughan S, Shaw MK, *et al.*: **Cell morphogenesis of *Trypanosoma brucei* requires the paralogous, differentially expressed calpain-related proteins CAP5.5 and CAP5.5V.** *Protist.* 2009; **160**(4): 576–590. [PubMed Abstract](#) | [Publisher Full Text](#)
- Colasante C, Alibu VP, Kirchnerberger S, *et al.*: **Characterization and**

- Developmentally Regulated Localization of the Mitochondrial Carrier Protein Homologue MCP6 from *Trypanosoma brucei*.** *Eukaryotic Cell*. 2006; **5**(8): 1194–1205.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
33. Sykes S, Szempruch A, Hajduk S: **The Krebs Cycle Enzyme  $\alpha$ -Ketoglutarate Decarboxylase Is an Essential Glycosomal Protein in Bloodstream African Trypanosomes.** *Eukaryot Cell*. 2015; **14**(3): 206–215.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Calvo-Álvarez E, Bonnefoy S, Salles A, *et al.*: **Redistribution of FLAgellar Member 8 during the trypanosome life cycle: Consequences for cell fate prediction.** *Cell Microbiol*. 2021; **23**(9): e13347.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
35. Smithson L, Ihuoma Akazue P, Findlater L, *et al.*: **Diversity in new flagellum tip attachment in bloodstream form African trypanosomes.** *Mol Microbiol*. 2022; **118**(5): 510–525.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

# Open Peer Review

Current Peer Review Status:   

---

## Version 2

Reviewer Report 05 June 2023

<https://doi.org/10.21956/wellcomeopenres.21501.r58162>

© 2023 Zikova A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Alena Zikova** 

Biology Centre, Institute of Parasitology, Czech Academy of Sciences, České Budějovice, Czech Republic

I would like to thank the authors for considering the reviewers' suggestions. They have done their best, and I have no further reservations or requests.

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Trypanosoma biology

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 24 May 2023

<https://doi.org/10.21956/wellcomeopenres.21501.r58161>

© 2023 Clayton C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Christine Clayton** 

Centre for Molecular Biology (ZMBH), Heidelberg University, Heidelberg, Germany

This is now considerably improved, though I still think that the data aren't as accessible as they could be. The "how-to" guide in Zenodo is more useful though it still requires some computing expertise (ability to use line command) to do all but download tables. I now see where the localizations are on the Table (which was requested by all reviewers) but getting the microscopy would be pretty complicated. I'm afraid many people won't bother trying, which is a pity.

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Trypanosome gene expression

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

## Version 1

Reviewer Report 20 March 2023

<https://doi.org/10.21956/wellcomeopenres.20611.r55329>

© 2023 Urbaniak M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Michael D. Urbaniak** 

Biomedical and Life Sciences, Faculty of Health and Medicine, Lancaster University, Lancaster, England, UK

The Tryptag project that has generated images of mNeon-green tagged protein in procyclic form (PCF) *Trypanosoma brucei* has been immensely useful in determining protein sub-cellular localisation and by inference function. In this work the authors have extended the project to include the clinically relevant bloodstream form (BSF) for selected protein sets, with a focus on those that are BSF enriched at transcript level or did not generate distinct signal in the PCF, in addition to certain nuclear proteins. This is a laudable effort and I have no doubt that it has been performed to the highest technical standards and will be of great utility to the community. However, the overall presentation of the data is disappointing and format of the available underlying data is unsuitable.

The manuscript is rather short, which as a data-set paper would be fine, but the brevity extends to a lack of specific details in the results and discussion. There are several times that we are told the rather vague "many" without the precision of hard numbers, percentages or readily available tabulated data to back up these statements. Identifying the results that support the assertions made should be made clearer and not requires a deep dive into the underlying data.

I find the format of the data presented in Figure 2 & 3 hard to interpret visually and would rather the data were tabulated for clarity, with additional supplementary table identifying which gene have these characteristics i.e. Set 1 "BSF upregulated" of 289 gene ~75% appear to match the threshold (>2.5-fold, P > 0.05) but their identity is unclear.

In the discussion regarding the differences in localisation of ~15% of proteins between PCF and BSF there is a statement that a "weak cytoplasmic signal" is simply background and "cytoplasm,

nuclear lumen and flagellar cytoplasm localisation” is observed in Pcf for unfused mNG. It would be useful to see examples images of this type (in comparison to “genuine” cytoplasmic signal) and to have greater clarity as to which proteins these caveats relate to. The localisation of unfused mNG in BSF should also be presented and discussed as this is a critical control.

The Bsf upregulated proteins with higher fluorescent signal in BSF than PCF “includes many novel or uncharacterised proteins localising to many different organelles”. The identity and localisation of these proteins should be clearly given in a supplementary table.

#### Underlying data is opaque

Whilst all the underlying data is technically available the data provided in the Zendo link is difficult to navigate or gain any sort of high-level view of the data without investing time and effort. Combined with the lack of detail in the manuscript itself, it is very difficult to see summary data (i.e. a list of proteins tagged and whether the localisation could be determined), let alone compare with previous PCF tagging. Simply browsing images is difficult to the point of being technically challenging and most likely beyond everyone except the most dedicated and data-savvy.

It is essential that the authors provide summary data in a user-friendly format that can be easily viewed and interpreted by the casual reader to ensure maximum dissemination and impact of the work. I would also urge the authors to integrate these results into the existing Tryptag resource so that they are searchable and can be readily viewed, and to work to integrate them into the TritypDB functional genomic database.

#### **Minor points:**

1. The introduction states “However, genome-wide mapping of the global changes are broadly limited to gene expression level, most extensively determined at the mRNA level<sup>7-11</sup> which does not correlate fully with protein abundance<sup>12</sup>. Few studies consider later steps in protein production: translation (mRNA ribosome footprinting)<sup>13</sup> and protein abundance (quantitative proteomics)<sup>14</sup>. ”

This is not an accurate reflection of the literature, as there are three SILAC quantitative proteomic studies alone that compare Bsf and Pcf protein abundance (PMID: 23090971<sup>1</sup> and references 14 & 26).

2. Reference 13 appear to be erroneous – it is certainly not an mRNA ribosome footprinting study.
3. The language in the manuscript would benefit from minor changes to grammar to ensure clarity, particularly in the abstract and introduction, i.e:

“Results: We have confirmed the localisation of known [stage-specific proteins] and identified the localisation of novel stage-specific proteins.”

“This showed how informative localisation can be for holistic mapping of potential protein function, although naturally [localisation] does not determine specific molecular function.”

“We also previously used high throughput tagging [of] BSF-upregulated genes to identify

ESB1”

### References

1. Butter F, Bucerius F, Michel M, Cicova Z, et al.: Comparative proteomics of two life cycle stages of stable isotope-labeled *Trypanosoma brucei* reveals novel components of the parasite's host adaptation machinery. *Mol Cell Proteomics*. 2013; **12** (1): 172-9 [PubMed Abstract](#) | [Publisher Full Text](#)

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**

No

**Are the conclusions drawn adequately supported by the results?**

Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Trypanosome cell signalling and post-transcriptional regulation

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Reviewer Report 01 March 2023

<https://doi.org/10.21956/wellcomeopenres.20611.r54787>

© 2023 Zikova A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Alena Zikova**

Biology Centre, Institute of Parasitology, Czech Academy of Sciences, České Budějovice, Czech

## Republic

No one doubts the importance of genome-wide localization studies. We can clearly see their power in the impact of the TrypTag project<sup>1</sup>, which was finally published in Nature Microbiology, but whose resources we have long been able to use via the TrypTag or Tritypodb platforms. Since the whole genome-scale analysis of subcellular localization was performed on the procyclic forms (PCF), one of the two easily cultured forms of *T. brucei*, it is obvious that future studies will also focus on the second form, the bloodstream stage (BSF).

The authors made a good selection of the initial sets of proteins of interest, namely i) genes that are upregulated (2.5 times,  $p < 0.05$  according to the study by Jensen *et al*, & some others), ii) genes for which no localization in PCF was shown, and iii) genes for which nucleoplasmic or nucleolar localization was shown (actually, a rationale for this third group should be included, as it must not be obvious to many, including me, I just conclude that it is related to VSG).

Based on the localization data obtained, the authors identify organelles that are subject to stage-specific regulation. They also conclude that in most cases localization does not change when known for both life cycle stages, and that if it was not possible to localize the protein to PCF, it is most likely a BSF protein whose localization can be detected in these forms.

Overall, the manuscript is quite short, but one was more looking forward to the supplement to evaluate the study in general and for specific IDs. However, I cannot evaluate the supplemental data available from Zenodo. I downloaded everything, clicked on everything, but I did not see any complete tables or images. After opening the TrypTag website, I was sad to see that the data from this study is not yet available on this website.

I have two main criticisms, which I will address below: i) some vagueness of the text, 2) the lack of supplements and tables to support the figures shown.

Ad1) - the whole text is interspersed with statements like:

'we observed convincing fluorescent signal in BSF for many tagged proteins [link to Fig 3B, which is firstly in percentages, so you have to do the math, and secondly it is not even clear how many genes are 100%]';

'...for many in Set 2...';

'...many of these genes encode...'

'...it includes many novel or uncharacterized proteins...'

One wonders how many is "many." Since the data sets were very discrete, I do not see why the authors cannot be more specific. The same is true for the cut-off for genes up-regulated in BSF, it is only mentioned once that the cut-off is 2.5 with a p-value  $< 0.05$ .

This certain ambiguity relates to Figure 2 and 3, which are a mystery to me. Example - Figure 2A, where the y-axis is in percent and labeled as BSF upregulated proteins - the first column - 8171 is 100% (I guess), so I conclude that 10% are upregulated BSF - that is about 800 proteins, but the authors selected 289 genes..., the second column - 289 is 100%, if so I do not know what this column is supposed to show, and the same is true for column 3 and 4.

It would be nice if the authors could redesign the diagrams to make them easier to understand (but it is possible that only I have this problem). Nonetheless, I would definitely recommend accompanying these graphs with supplemental Excel spreadsheets highlighting the gene IDs, including the gene description, the determined localization in BSF and PCF.

This will also help in determining for how many hypothetical proteins subcellular localizations were found. The authors mention this in the abstract, "we have confirmed the localization of known and identified the localization of novel stage-specific proteins" but practically, there is no table or graph (how many and where they are) for this result.

Figure 1 - I would recommend putting the number of genes in each group in parentheses.

Ad2) Lack of supplements and tables that would support the figures shown - this was mentioned above, but I would add that a table would also be useful for Figure 5. While it is nice to see the specific organelle adaptation, scientists will always want to look for their favorite gene IDs.

This is also related to the question of whether it is possible to make the data available on TrypTag; I understand that linking to Tritrypdb will take some time.

Last but not least, the phenomenon of dissimilar localization between PCF and BSF is very interesting. The authors claim that this is the case in 15% of the genes studied, which is not low. A list of their IDs and a comparison of localization would be very interesting. It is a pity that the authors did not take the published examples as proof of concept, I know of at least two examples - MCP 6 (PCF mito, BSF glycosomes, Colasante *et al.*, 2006<sup>2</sup>) and alpha-ketoglutarate decarboxylase (PCF mito, BSF glycosome, Sykes *et al.*, 2015<sup>3</sup>).

Otherwise, I believe that all the supplementary information (list of genes, their descriptions, localizations, and IFA images) is probably available via the files uploaded to Zenodo, but for students/researchers who do not know how to use a programming language, this information is actually inaccessible and reduces the impact of the study on the scientific community. An addition of a detailed description on how to dissect these data would help.

## References

1. Billington K, Halliday C, Madden R, Dyer P, *et al.*: Genome-wide subcellular protein map for the flagellate parasite *Trypanosoma brucei*. *Nat Microbiol.* 2023. [PubMed Abstract](#) | [Publisher Full Text](#)
2. Colasante C, Alibu VP, Kirchberger S, Tjaden J, *et al.*: Characterization and developmentally regulated localization of the mitochondrial carrier protein homologue MCP6 from *Trypanosoma brucei*. *Eukaryot Cell.* 2006; **5** (8): 1194-205 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Sykes S, Szempruch A, Hajduk S: The krebs cycle enzyme  $\alpha$ -ketoglutarate decarboxylase is an essential glycosomal protein in bloodstream African trypanosomes. *Eukaryot Cell.* 2015; **14** (3): 206-15 [PubMed Abstract](#) | [Publisher Full Text](#)

## Is the work clearly and accurately presented and does it cite the current literature?

Partly

## Is the study design appropriate and is the work technically sound?

Yes



**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

I cannot comment. A qualified statistician is required.

**Are all the source data underlying the results available to ensure full reproducibility?**

No

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Trypanosoma biology

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Reviewer Report 03 February 2023

<https://doi.org/10.21956/wellcomeopenres.20611.r54582>

© 2023 Clayton C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Christine Clayton** 

Centre for Molecular Biology (ZMBH), Heidelberg University, Heidelberg, Germany

Results from the Tryptag project constitute an immensely valuable resource. The ability to select proteins localized to particular structures or sub-structures within the trypanosome has already resulted in several publications and greatly facilitates interpretation of various types of data - such as distinguishing interaction partners identified by mass spectrometry from likely contaminants. A gap, however, was that the project was done using the procyclic form. This meant that N-terminally tagged proteins that are expressed exclusively in the bloodstream form could not be detected. Since mRNA regulation usually relies on 3'-untranslated regions, tagging at the C-terminus (which also replaces the 3'-UTR) might solve that issue but could also result in aberrant expression. In this manuscript the authors now describe localizations for bloodstream-form-specific proteins, with an emphasis on N-terminal tagging, except where N-terminal signal sequences were expected.

The paper is fine as far as it goes but currently of limited use because I found it impossible to access any of the underlying data, which are currently available only via a (to me, impenetrable) Zenodo folder. This was incredibly frustrating, I couldn't even see which genes had been analyzed.

There are two important gaps that must be filled:

1. The new localizations do not appear to be available via the Tryptag web site. I found only one image set for Tb927.11.1230, for example. The data must be uploaded onto Tryptag, unless there is a compelling reason why this is not possible. There should also be links to TritrypDB.

I am not a total idiot with regard to code but after following the Zenodo link (and downloading the zipped folder) I hadn't a clue what to do next. If it really is possible to get the data this way, it is essential to provide a really clear guide suitable for people who are not able to deal with code at all. But uploading onto Tryptag would be way preferable.

2. The paper must include a supplement which is simple spreadsheet listing the genes analysed, success of tagging, new results and previous PCF results. The authors presumably have this list since it would be needed to generate their figures. At the same time perhaps the authors could provide such a spreadsheet for the entire procyclic form dataset, since at present it is possible to interrogate (and download from) the Tryptag database only with individual genes or localizations. While I appreciate that no-one should undertake further experiments on a protein without consulting the original images, broader analyses absolutely require these spreadsheets.

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Partly

**If applicable, is the statistical analysis and its interpretation appropriate?**

I cannot comment. A qualified statistician is required.

**Are all the source data underlying the results available to ensure full reproducibility?**

No

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Trypanosome gene expression

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.**

# Comments on this article

## Version 1

Author Response 19 Apr 2023

### Richard Wheeler

We thank the reviewers for their careful consideration of this work. Unfortunately, it is clear that a major theme from the reviewers' comments was concerns about data availability.

We are sorry that this us seen as an issue, especially as the main priority of this paper is to describe this open dataset. I assure the reviewers that all the microscopy and localisation annotation data is available at Zenodo, under record number: 7258722.

For example, the list of localisation annotations (and gene IDs tagged) is in the file called "localisations.tsv".

The current link for this is <https://zenodo.org/record/7843865/files/localisations.tsv?download=1>, but it is better to access it via the Zenodo website to ensure the latest version of the record is accessed. There was clearly frustration that images are not easily visible. Please understand that this is a non-trivial data management issue.

While it would be good to have all of the image data simply listed for download by gene ID, the total dataset is too large even for specialised data deposition websites like Zenodo. Hence the spread over multiple Zenodo entries.

To access the microscopy data you just need to look up the gene ID in "id\_doi\_index.tsv", go to the indicated Zenodo ID and download the microscopy data. The current link for this index [https://zenodo.org/record/7843865/files/id\\_doi\\_index.tsv?download=1](https://zenodo.org/record/7843865/files/id_doi_index.tsv?download=1)

This is the best solution we have been able to set up. Unfortunately scientific infrastructure is simply not well-suited to projects with this scale of microscopy data. Please note that the emphasis placed on the .zip file (which is only necessary to replicate our image management) by the Zenodo preview is out of our control (see <https://github.com/zenodo/zenodo/issues/1477>)

We have updated the Zenodo description to try to help, however there is little more we can do than this. We are not keen to change the paper text to explicitly describe the Zenodo deposition file structure as the two are separately indexed entities would ultimately lose synchrony. We have decided that this data will not be made available on TrypTag.org, as the bloodstream form data is neither genome-wide nor carried out in the same strain:

TrypTag.org was designed for genome-wide information and is a poor search/user interface for identifying data with sparse genome coverage. A dedicated bloodstream form website would be 95% empty.

Adding the bloodstream form data to the existing procyclic form/TrypTag.org database is complex as the bloodstream form tagging was carried out in *T. brucei* Lister 427 instead of TREU927. There

is no good or simple way to handle this with a mixed database as gene IDs are not synchronised between these strains. TriTrypDB is aware of this dataset, and we hope will be available soon. There were several requests for data to be included as supplemental figures/tables. However, Wellcome Open Research articles cannot have supplementary information of that type: "Extended data should be deposited in an approved repository and listed as part of the data availability statement." (<https://wellcomeopenresearch.org/for-authors/article-guidelines/preparing-a-research-article/#supp>)

To address the various requests for a tabular format of the classes/criteria used through the paper we have added a new file to the Zenodo repository, named "geneselection.tsv", which contains this information. Christine Clayton These comments almost entirely concerned data access, which I hope we have clarified sufficiently above. There are two specific additional notes: No coding ability is needed to access the bloodstream form tagging data. Just look up the gene ID in the id\_doi\_index.tsv, and then go to the appropriate Zenodo deposition and download the zip file. Including TrypTag/procyclic form localisation data in the bloodstream form Zenodo deposition is an unnecessary duplication. TrypTag localisation annotations are available in that project's master Zenodo deposition, also named "localisations.tsv". Note that The TREU927 gene IDs identified as orthologs for our analysis are, however, provided for easy cross-reference. Alena Zikova Again, for queries related to data access, please see above.

Re. "candidates for *T. brucei*-specific BSF nuclear structure adaptation.": This, indeed, was motivated by the potential for nuclear architecture change associated with variant surface glycoprotein gene transcription. We have clarified this in the text.

Re. usage of "many": We have updated the text to include specific numbers and percentages at all such statements.

Re. Figure 2/3: We have updated the legend to indicate that the numbers at the top of each column indicate the total number of genes each category on the x axis, with the y axis indicating proportion of the proteins in that category matching the specified criterion. Note the totals vary depending on data availability. Eg. in Figure 3, the number is different for each successive panel as A is all tagging attempts, B is only successfully generated cell lines and D is only where both PCF and BSF tagging gave a cell line.

Regarding the number of upregulated genes in the "All" gene set in Figure 2A being much larger than the ~300 we analysed here. You are absolutely correct, but bear in mind that this is all genes, ie. including genome including 427-specific genes, VSGs, ISGs, (GR)ESAGs etc. which were excluded from our analysis as indicated in Figure 1.

Re. Figure 1: We have added the number of genes in each gene set to Figure 1. Note that these are the numbers indicated at the start of the results and for each set in Figure 3A.

We have also this information to the geneselection.tsv file

Re. Extended data for Figure 5: In this work, we do not do any additional mRNAseq to quantify transcript abundance - the data source we use for generating Figure 5 is Jensen et al. 2014. This is

already stated in detail for BSF vs. PCF, and we have updated the text to clarify that PCF vs. BSF is done in the same way.

We do not want to re-publish previously published data and the PCF or BSF-upregulated gene lists can be generated easily from the Jensen *et al.* 2014 supplemental data. We have, however, listed whether this criterion was met in `geneselection.tsv`

Re. Dissimilar localisation in PCF and BSF: The 85% similar rate is followed by an explanation of why those differences were then disregarded as differences in localisation, due to either similarity to background autofluorescence or a 'suspect' whole cell signal. Therefore our estimated number of 'true' differences is 1 protein, well under 1%, and this protein is specifically shown in Figure 4C, 4D in updated figure.

Again, we have included a column in the `geneselection.tsv` file indicating the similar/dissimilar manual comparison.

Thank you for examples of MCP6 and  $\alpha$ -KDE1, and we have added these references. Note that we did already highlight FLAM8 as a previously-characterised protein with a highly comparable localisation change between BSFs and PCFs.

Re. Data access: As noted above, we emphasise that no coding whatsoever is needed to access image data. Just look up the gene ID in `id_doi_index.tsv` table and download the zip from the corresponding Zenodo deposition. Michael Urbaniak

Re. Use of "many" and lack of specificity: We reserved use of these qualitative phrases when the figures showed the quantitation of these results, however for clarity have added the relevant measures to the text wherever we've made such a statement.

Re. "weak cytoplasmic signal" and "cytoplasm, nuclear lumen and flagellar cytoplasm" localisations: We have added example images of PCFs expressing no tagged protein and soluble mNG, and equivalent bloodstream form localisations (new Figure 4A) - when localisations like these are seen they are potentially, but not necessarily, spurious.

We noted that it was unclear from the text that we previously analysed the origin of these potentially spurious localisations in detail in PCFs, and have clarified the text to address this.

Re. "The identity and localisation of these proteins should be clearly given in a supplementary table.": The full listing of all gene IDs and localisation is provided in `localisations.tsv`, and now supported by the new `geneselection.tsv` file.

Re. Citation of quantitative proteomics: Apologies, we intended to make it clear that the reference was a single example. Including all three citations as a comprehensive overview is of course absolutely no problem.

Re. Reference 13: You are absolutely correct that this is an error, thank you for highlighting this. The Vasquez and Jensen 2014 papers are the ribosome profiling papers that should be cited here,

and Jensen 2014 was our primary reference dataset for mRNA seq.

Re. Grammatical changes: We have addressed these minor corrections.

**Competing Interests:** No competing interests were disclosed.

---