Trends in Parasitology

CellPress OPEN ACCESS

Forum

TrypTag.org: from images to discoveries using genome-wide protein localisation in *Trypanosoma brucei*

Jack D. Sunter [1,*], Samuel Dean [2,*] and Richard John Wheeler [3,*]

TrypTag was a 4-year project to tag the N- and C-termini of almost all *Trypanosoma brucei* proteins with a fluorescent protein and record the subcellular localisation through images and manual annotation. We highlight the new routes to cell biological discovery this transformative resource is enabling for parasitologists and cell biologists.

The TrypTag project aimed to generate a microscopy map of the subcellular localisation of the vast majority of proteins encoded in the *Trypanosoma brucei* genome. Our target was 8721 protein-coding genes after excluding the known variant surface glycoprotein genes (critical for antigenic variation), pseudogenes, and identical duplicated genes. After optimising conditions for high-throughput cell line generation and imaging, we established that only the procyclic form (insect midgut proliferative form) was sufficiently tractable for genome-scale protein localisation in a reasonable time frame. We used endogenous N- or C-terminal tagging to generate cell lines, which we subsequently imaged at high quality under consistent conditions [1,2].

The project began in 2016, with a policy of making data rapidly openly accessible via TriTrypDB and TrypTag.org, and, in 2017, we published an announcement article outlining the resource [1]. Prior to publication of our genome-wide analysis [2], this article was used as the primary citation when TrypTag data were used. To examine the prepublication impact of TrypTag on the research community, we identified PubMed-indexed publications citing this article: By December 2022, these totalled 119 publications broadly falling into five categories (Figure 1, Table S1 in the supplemental information online):

1. Validation of a proteome, immunoprecipitation, or bioinformatic dataset or screen
2. Validation of protein localisation by an alternative method, such as immunofluorescence or another fluorophore tag
3. Identification of marker protein(s) for organelles as the basis for another assay
4. Identification or characterising novel protein(s) of interest based on localisation
5. Methodology or potential application of the resource

The power of the TrypTag resource for these applications stems from its high quality and coverage: 89% (7766) of *T. brucei* proteins were tagged on at least one terminus, and >75% were tagged at both. Overall, 5806 had a clear fluorescent signal, brighter than the background and/or localised to a structure atypical of the background signal. Finally, these data are not simply images. A high-quality open access localisation database was created through blinded and cross-checked manual annotation using a localisation ontology [matched with gene ontology (GO) cellular component terms], based on the expertise gained from classifying thousands of subcellular localisations.

To date, the majority of publications citing TrypTag used it for validation of either protein localisation or proteomic/bioinformatic screens and comparative genomics. Validation by an independent method is important and can add significant value; for example, TrypTag facilitated the interpretation of hits from a genome-wide screen, by the Horn group, of genes required for cell cycle progression [3]. However, we believe TrypTag is additionally powerful due to its genome-wide nature, as demonstrated by the Akiyoshi group, who determined there were unlikely to be any undiscovered kinetochore proteins not already identified by various proteomic approaches [4].

There are also a growing number of studies where TrypTag was a key resource for generating and/or testing hypotheses. Here, we highlight a range of discoveries enhanced or made possible by TrypTag.

Identification of markers

Markers for organelles in microscopy are critical for many analyses, historically dominated by monoclonal antibodies but now making use of epitope or fluorescent tags enabled by the use of advanced genetic tools to generate transgenic cell lines. However, as a neglected tropical pathogen, many organelles/structures were poorly covered by characterised proteins or available antibodies. TrypTag has been used to overcome this, providing a panel of validated markers, such as the flagellar pocket marker used by the Carrington group [5]. We were able to translate many of these to *Leishmania* spp., and they have proved valuable, e.g., our use of flagellar pocket markers [6]. We anticipate such markers will enable advanced readouts for many cell biological assays, drug treatments, or mutant phenotype analyses.

Asymmetries and subdomains

Complex asymmetry of the *T. brucei* cytoskeleton is an emerging theme, particularly shown through analysis of the cortical cytoskeleton [7]. Our analysis of the TrypTag dataset showed extensive examples of such spatial asymmetry and identified many examples of temporal (i.e., old versus newly formed copy of organelles) asymmetry [2]. In a separate study, we used TrypTag

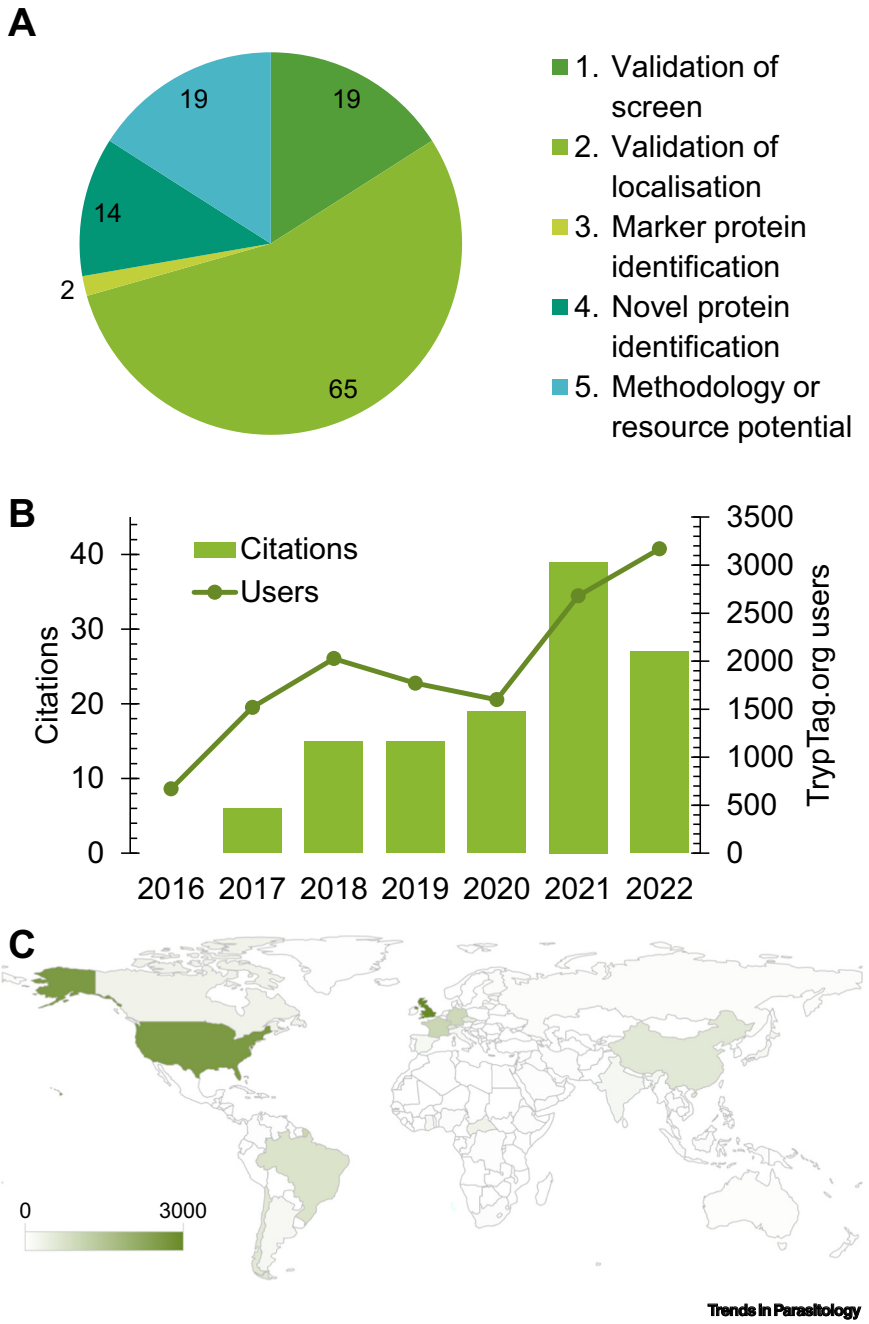

Figure 1. Citations of TrypTag prior to release of the final dataset. (A) Papers citing the TrypTag project announcement (PubMed-indexed papers indexed up to December 2022), organised by manual classification of broad use of the resource. (B) Number of papers and TrypTag.org website users (unique users reported by Google Analytics), organised by year of publication or visit. (C) Geographic distribution of TrypTag.org website users (user geolocation reported by Google Analytics).

to identify a novel proximal/distal asymmetry of the outer dynein arm (ODA) within the axoneme of *T. brucei* – these asymmetries were dependent on a paralogous pair of ODA docking complexes and, importantly, were conserved in *L. mexicana* and necessary for normal flagellum beating [8]. We argue that functional specialisation of organelle asymmetries and subdomains is an important growing area for which TrypTag is a critical resource.

Insights into different life cycle stages

Parasites adapt to different host environments by changing their gene expression program. One prominent adaptation in *T. brucei* is the vital bloodstream monoallelic expression of a variant surface glycoprotein (VSG) surface antigen and subsequent switching to confer antigenic variation. This does not occur in the procyclic form. Work by the Rudenko group showed how TrypTag can inform novel factors regulating monoallelic VSG transcription in bloodstream forms by well-judged interrogation of TrypTag procyclic localisations. By mining TrypTag for proteins localising to one to four foci in the nucleus (potential nuclear bodies), they identified Cajal body, spliced leader array body, and NUFIP body proteins, which they showed associate with and regulate the expression site body (ESB), the site of VSG transcription [9]. We used TrypTag to successfully identify the first and long-sought ESB-specific protein, reasoning that lack of detectable tagged protein signal in the procyclic was an important criterion, combined with bloodstream form localisation, mRNA abundance, and bioinformatic data [10]. This shows little or no expression of a tagged protein in TrypTag after tagging at both termini can indicate expression is restricted to another life cycle stage, and this can be leveraged to identify important proteins in different life cycle stages.

High content image analysis

The genome-wide coverage and high-quality standardised microscopy conditions for TrypTag make it powerful for automated image analysis. We demonstrated the power of this by measurement of protein partition between the nucleus, nucleolus, and cytoplasm, through which

we reidentified the canonical eukaryotic nuclear localisation signal KRXK for nuclear targeting and showed there is unlikely to be a similar short linear motif for nucleolar targeting – positive protein charge is important instead [11]. Our ongoing work uses high content analysis of TrypTag data to understand flagellum transition zone and paraflagellar rod molecular ultrastructure, and we believe the highly structured *T. brucei* cell will empower similar quantitative analyses for many organelles.

Bioinformatic pipeline development

A powerful application of TrypTag will be its incorporation into evolutionary genomic analyses. For example, the Keeling lab studied the horizontal gene transfer of viral glycosyltransferases into eukaryotes and noted two transfers into the kinetoplastids that they hypothesised were important for mitochondrial innovation; TrypTag provided the experimental characterisation by showing the two glycosyltransferases localised to the mitochondrion [12]. We believe that leveraging TrypTag will aid the deconvolution of protein targeting signals. The presence of the mNG tag may disrupt a targeting signal that is proximal to the terminus tagged, as described for the C-terminal glycosome localisation sequence. Genome-wide, we observed that protein tagging at a terminus expected to disrupt targeting did indeed do so [2]. For example, proteins localising to the endoplasmic reticulum (ER) by C-terminal tagging were unlikely to have an ER localisation by N-terminal tagging. This informs import mechanisms, identifying ER proteins using a canonical N-terminal signal peptide and, by exclusion, those that may be using atypical targeting. Our initial analysis did not identify any short linear targeting sequences for any organelle except the nucleus. This suggests that most protein targeting is not short motif dependent or that targeting motifs are restricted to a limited number of proteins, which then direct the targeting of the protein complex which they are assembled into. We found a poor correlation between

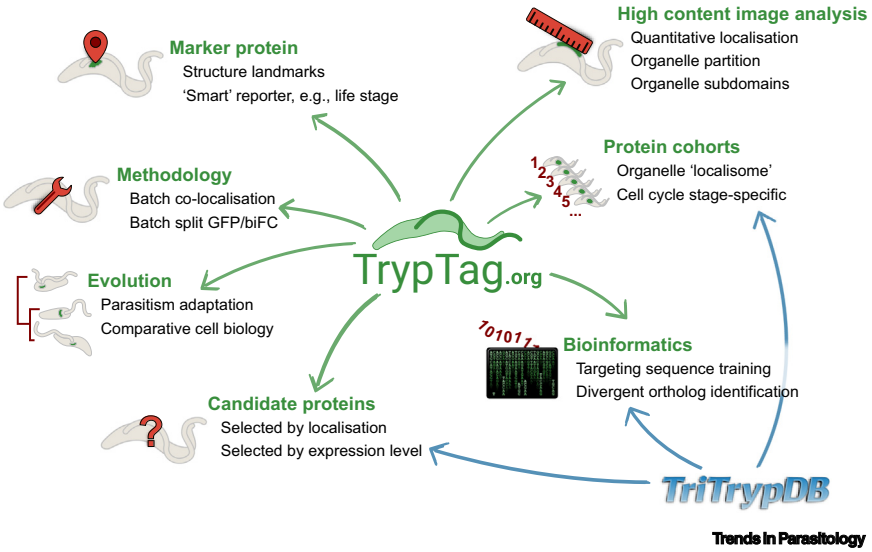

Figure 2. Potential future TrypTag applications. Broad overview of some of the potential TrypTag applications, also highlighting where TriTrypDB-assisted searching is likely of particular benefit.

TrypTag and subcellular localisation prediction algorithms, likely due to their overfitting to the relatively closely related common model organisms. This underscores the power of TrypTag for the development of bioinformatic tools, as trypanosomes provide critical evolutionary context due to their phylogenetic position as an early-branching eukaryote.

The work above, showcasing the power of TrypTag, has relied on prepublication access to the TrypTag resource. Final annotation is now complete; it is openly available and fully searchable, and the full raw data are available for download. In our initial genome-wide analysis, we show that this allows analysis of *T. brucei* organelle complexity, organelle evolution associated with parasitism, and which organelles are enriched in vital proteins, and that it highlights many cell structures of interest [2].

TrypTag has improved the functionality of trypanosomes as a research system, which will attract new scientists and ideas to the field. We anticipate this will drive innovative uses by the cell biology and parasitology research communities, in addition to more genome-wide analyses along with applications of the types we outline in the preceding text (Figure 2).

The broader power of genome-wide microscopy protein localisation is clear from the impact of the existing human and yeast datasets [13–15] and provides complementary advantages to advanced proteomic methods, such as Localisation of Organelle Proteins by Isotope Tagging (Box 1). TrypTag adds *T. brucei* to this collection of species, providing a protein localisation resource for an early-branching eukaryote, a flagellate, a parasite, and a neglected tropical pathogen.

Box 1. A brief history of genome-wide protein localisation resources

The first genome-wide subcellular protein localisation resource produced was in budding yeast, *Saccharomyces cerevisiae*, published in 2003 [13]. This was followed 6 years later with a resource for fission yeast, *Schizosaccharomyces pombe* [15]. It was not until 2017 that genome-wide subcellular protein localisations in several human cell lines were published [14]. TrypTag provided genome-wide localisations by N-terminal tagging in 2017, both termini in early 2020, and the full annotated and analysed summary in 2022 [2].

Acknowledgments

TrypTag was funded by a Wellcome Trust biomedical resource grant (108445/Z/15/Z), and R.J.W. is supported by the Sir Henry Dale Fellowship (211075/Z/18/Z). We thank Ross Madden, Phil Dyer, Karen Billington, and Clare Halliday for their work generating the localisation data. We also thank Mark Carrington, Christiane Hertz-Fowler, Sue Vaughan, and Keith Gull for their support and advice.

Declaration of interests

The authors declare no competing interests.

Supplemental information

Supplemental information associated with this article can be found online at https://doi.org/10.1016/j.pt.2023.02.008.

[1]Department of Biological and Medical Sciences, Oxford Brookes University, Oxford, UK
[2]Warwick Medical School, University of Warwick, Coventry, UK
[3]Peter Medawar Building for Pathogen Research, Nuffield Department of Medicine, University of Oxford, Oxford, UK

*Correspondence:
jsunter@brookes.ac.uk (J.D. Sunter),
samuel.dean@warwick.ac.uk (S. Dean), and
richard.wheeler@ndm.ac.uk (R.J. Wheeler).

https://doi.org/10.1016/j.pt.2023.02.008

References

1. Dean, S. *et al.* (2017) TrypTag.org: a trypanosome genome-wide protein localisation resource. *Trends Parasitol.* 33, 80–82
2. Billington, K. *et al.* (2023) Genome-wide subcellular protein localisation in the flagellate parasite *Trypanosoma brucei*. *Nat. Microbiol.* Published online February 20, 2023 https://doi.org/10.1038/s41564-022-01295-6
3. Bachmaier, S. *et al.* (2022) A multi-adenylate cyclase regulator at the flagellar tip controls African trypanosome transmission. *Nat. Commun.* 13, 5445
4. Nerusheva, O.O. *et al.* (2019) Identification of four unconventional kinetoplastid kinetochore proteins KKT22-25 in *Trypanosoma brucei*. *Open Biol.* 9, 190236
5. Martos-Esteban, A. *et al.* (2022) Black-necked spitting cobra (*Naja nigricollis*) phospholipases A2 may cause *Trypanosoma brucei* death by blocking endocytosis through the flagellar pocket. *Sci. Rep.* 12, 6394
6. Halliday, C. *et al.* (2019) Cellular landmarks of *Trypanosoma brucei* and *Leishmania mexicana*. *Mol. Biochem. Parasitol.* 230, 24–36
7. Sinclair, A.N. *et al.* (2021) The *Trypanosoma brucei* subpellicular microtubule array is organized into functionally discrete subdomains defined by microtubule associated proteins. *PLoS Pathog.* 17, e1009588
8. Edwards, B.F.L. *et al.* (2018) Direction of flagellum beat propagation is controlled by proximal/distal outer dynein arm asymmetry. *Proc. Natl. Acad. Sci. U. S. A.* 115, E7341–E7350
9. Budzak, J. *et al.* (2022) An assembly of nuclear bodies associates with the active VSG expression site in African trypanosomes. *Nat. Commun.* 13, 101
10. López-Escobar, L. *et al.* (2022) Stage-specific transcription activator ESB1 regulates monoallelic antigen expression in *Trypanosoma brucei*. *Nat. Microbiol.* 7, 1280–1290
11. Jeilani, M. *et al.* (2022) Nucleolar targeting in an early-branching eukaryote suggests a general mechanism for ribosome protein sorting. *J. Cell Sci.* 135, jcs259701
12. Irwin, N.A.T. *et al.* (2022) Systematic evaluation of horizontal gene transfer between eukaryotes and viruses. *Nat. Microbiol.* 7, 327–336
13. Huh, W.-K. *et al.* (2003) Global analysis of protein localization in budding yeast. *Nature* 425, 686–691
14. Thul, P.J. *et al.* (2017) A subcellular map of the human proteome. *Science* 356, eaal3321
15. Hayashi, A. *et al.* (2009) Localization of gene products using a chromosomally tagged GFP-fusion library in the fission yeast *Schizosaccharomyces pombe*. *Genes Cells* 14, 217–225