Impact of inherent biases built into proteomic techniques: Proximity labeling and affinity capture compared

Claudia Maria do Nascimento Moreira, Cristina D. Kelemen, Samson O. Obado, Farnaz Zahedifard, Ning Zhang, Fabiola B. Holetz, Laura Gauglitz, Bruno Dallagiovanna, Mark C. Field, Susanne Kramer, Martin Zoltner

PII: S0021-9258(22)01169-3
DOI: https://doi.org/10.1016/j.jbc.2022.102726
Reference: JBC 102726

To appear in: Journal of Biological Chemistry

Received Date: 16 August 2022
Revised Date: 26 October 2022
Accepted Date: 14 November 2022


This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology.
Impact of inherent biases built into proteomic techniques: Proximity labeling and affinity capture compared.

Claudia Maria do Nascimento Moreira¹,², Cristina D. Kelemen³, Samson O. Obado⁴, Farnaz Zahedifard³, Ning Zhang⁵, Fabiola B. Holetz², Laura Gauglitz¹, Bruno Dallagiovanna², Mark C. Field⁶,⁶, Susanne Kramer¹* and Martin Zoltner³*

¹Department of Cell and Developmental Biology, Biocenter, University of Würzburg, Würzburg, Germany, ²Carlos Chagas Institute (ICC), FIOCRUZ/PR, Curitiba, Brazil, ³Department of Parasitology, Faculty of Science, Charles University in Prague, Biocev, Vestec, Czech Republic, ⁴The Rockefeller University, New York, NY10021, USA, ⁵School of Life Sciences, University of Dundee, Dundee, DD1 5ED, UK and ⁶Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic.

*Corresponding authors: MZ: zoltnerm@natur.cuni.cz, SK: susanne.kramer@uni-wuerzburg.de

Key words: Affinity capture, proteome, cryomilling, BioID, interactome.
Abstract

The characterization of protein-protein interactions (PPIs) is of high value for understanding protein function. Two strategies are popular for identification of PPIs direct from the cellular environment: Affinity capture (pulldown) isolates the protein of interest with an immobilized matrix that specifically captures the target and potential partners, while in BioID genetic fusion of biotin ligase facilitates proximity biotinylation and labelled proteins are isolated with streptavidin. Whilst both methods provide valuable insights, they can reveal distinct PPIs, but the basis for these differences is less obvious.

Here, we compare both methods using four different trypanosome proteins as baits: poly(A) binding proteins PABP1 and PABP2, mRNA export receptor MEX67 and the nucleoporin NUP158. With BioID, we found the population of candidate interacting proteins decreases with more confined bait protein localization, but the candidate population is less variable with affinity capture. BioID returned more likely false-positives, in particular for proteins with less confined localization, and identified low molecular weight proteins less efficiently. Surprisingly, BioID for MEX67 identified exclusively proteins lining the inner channel of the nuclear pore complex (NPC), consistent with the function of MEX67, while the entire NPC was isolated by pulldown. Similarly, for NUP158, BioID returned surprisingly few PPIs within outer rings of the NPC that were by contrast detected with pulldown, but instead returned a larger cohort of nuclear proteins.

These rather significant differences highlight a clear issue with reliance on a single method to identify PPIs and suggest that BioID and affinity capture are complementary rather than alternative approaches.
Introduction

Most proteins function as part of multi-subunit complexes and identification of protein-protein interactions (PPIs) is valuable for understanding function. Identification of PPIs can uncover a wide range of interactions, that include direct, indirect, static or dynamic binding. Furthermore, proteins can moonlight and engage in multiple different specific complexes, while complex composition can change in a temporal and/or spatial manner.

Presently there are two common methods used to identify PPIs in a cellular context: Affinity capture (colloquially pulldown) (1) and proximity-labelling (2). While evidence indicates that both methods provide valuable insights, there are considerable differences between them, both in technical requirements and consequently the interactome revealed. Each method has its adherents and while it would be fallacious to view one or the other approach as superior, it is unclear what choice of method implies in terms of the types of PPIs identified and hence critical assessment of a given dataset.

Affinity capture requires cell rupture with the lysate or extract then exposed to a solid phase with a specific affinity to the protein of interest or bait. The solid phase is commonly coupled to an antibody against the bait or alternatively the bait is genetically fused to a tag and purified using a solid phase with affinity to the tag. The bait along with co-purified interaction partners is then eluted and this eluate analyzed. A major disadvantage here is that interactions are determined from the cell extract and not the true cellular environment and thus some interactions may be lost, while spurious, non-physiological interactions may occur as a result of membrane breakage and decompartmentalization of protein complexes. Screening of extraction conditions including detergent solubilization, pH, ionic strength and other parameters is therefore necessary. A variation on this approach is cryomilling, which avoids the use of detergents during lysis, but still disrupts cellular organization (3, 4). In general, affinity capture requires a lot of optimization for each individual bait, and results tend to be more variable between experiments and/or labs due to minor differences in cell harvesting, lysis and buffer conditions.

In proximity labelling some of the pitfalls of affinity isolation are avoided by capturing interactions in vivo. Here the bait is fused to an enzyme that
converts a substrate into a reactive radical that is covalently linked to nearby proteins. Modified proteins are purified frequently under extremely stringent conditions. The most commonly used method is BioID and variants. The bait is coupled to BirA*, a mutant form of biotin ligase from *Escherichia coli*. Wild-type BirA converts biotin to biotinol-5’-AMP which is retained by BirA until it is transferred to acetyl-CoA carboxylase. A mutant version, BirA*, is modified to release biotinol-5’ AMP, causing biotinylation of lysine residues of proteins nearby (5–7). Additional BioID variants have been developed, one of which is TurboID which exhibits greater biotinylation efficiency (8). TurboID biotinylates within minutes after addition of exogenous biotin, in contrast to other BirA* variants (8), but also has some biotinylation activity in the absence of exogenous biotin, leading to an increased labelling radius (9).

Given the distinct underlying principles of these two methods, we have carried out a direct comparison between them, employing four different proteins from trypanosome mRNA metabolism that encompass a variation of location and positional constraints. For affinity capture we used data previously published by us (10, 11) and repeated the experiment for one target, nucleoporin NUP158, for optimal comparability. Cell lysis was performed by cryomilling, where rapid freezing and mechanical cell disruption at 77 K preserves PPIs and has been successful for isolating NPCs and many other complexes from many organisms (3, 12–14). All proteins were expressed as GFP/YFP fusions from the endogenous locus and captured with an anti-GFP single-chain nanobody (10, 11). For BioID we expressed the same proteins from their endogenous loci fused to the biotin ligase TurboID and used steady-state biotinylation, facilitated by biotin in the culture medium.

We found surprisingly little concordance between PPIs identified from affinity capture and BioID. With their less-confined cytoplasmic localizations BioID is error-prone for poly(A)-binding proteins, with many false identifications. In contrast, for the identification of MEX67 interactors BioID was more adequate and exclusively identified FG-repeat nucleoporins, the nuclear pore complex (NPC) components that line its inner surface, rather than the entire NPC, that was isolated by the affinity method. Likewise, NUP158 affinity capture identified most part of the NPC cellular structure including distant proteins that are indirectly but stably associated with the bait, which are outside the BioID
labelling radius. Whilst affinity capture delivers a snapshot of stable PPIs at the
time of lysis, proximity labelling records a history of protein interactions
occurring during the duration of labelling, which can bias against detection of a
stable ‘core’ complex in favor of dynamic associations. Altogether, our data
indicate that the utility of each method is context-dependent and hence should
be viewed as complementary rather than as alternatives.

Results and Discussion

Establishing TurboID in trypanosomes: We selected four trypanosome proteins
as bait to establish BioID with TurboID biotin ligase fusion. Poly(A) binding
proteins PABP1 and PABP2 are cytoplasmic and by light microscopy appear
unconfined within the cytoplasm under standard culture conditions (10, 15). The
nuclear export receptor MEX67 shuttles between the nucleus and the
cytoplasm with predominant localization at NPCs (16, 17). The NPC protein
NUP158 (orthologue to Nup145 in S. cerevisiae and Nup98/96 in H. sapiens)
is localised to the outer rings of the NPC (3). We selected these bait proteins
for the following reasons: (i) We anticipate differential levels of non-specific
background biotinylation dependent on how rigidly a protein is confined, and
including a range of protein localizations (not confined, semi-confined,
confined) allows this to be considered, and (ii) Prior work means we have
validated cell lines and in some cases mass spectrometry (MS) data (10, 11).

All four proteins were expressed from their endogenous loci fused C-
terminally to TurboID and hemagglutinin (HA), replicating the tagging-strategy
used for cryomill affinity capture, except that in that case fusion was to eYFP.
We used two control cell lines: Unmodified parental cells and cells expressing
eYFP fused to TurboID and HA (using an inducible expression system). All
experiments were done in procyclic form (PCF) cells, the T. brucei life-cycle
stage in the insect host. Generated lines were subjected to western blotting to
detect biotinylated proteins by a streptavidin probe (Figure 1A). Almost no
biotinylated proteins were detected in parental cells, while the tagged cell lines,
including the eYFP control, had many biotinylated proteins. There were major
differences in the number of biotinylated proteins obtained for each bait and
also in the identity of the detected proteins, indicating specificity. Notably,
PABP2-TurboID resulted in a significantly higher number of biotinylated proteins than PABP1-TurboID in both life cycle stages, consistent with previous findings that PABP2 interacts with a significantly bigger cohort of proteins compared to PABP1 (10). Additionally, we confirmed correct localization of the TurboID bait proteins by anti-HA immunofluorescence and, in parallel, traced biotinylation via a fluorescent streptavidin probe (Figure 1B). For the cytoplasmic PABPs we observed cytoplasmic localization of the two probes and for NUP158, co-localization at distinct structures of the nuclear envelope. MEX67 TurboID yielded a similar streptavidin labelling pattern as NUP158 with some additional weak signal from the nucleus, but immunofluorescence signals were restricted to nucleoplasm and nucleolus. The absence of immunofluorescence from the nuclear envelope can be explained by the phase-separated environment created by the FG-type nucleoporins, that presumably prevents antibody body binding in this region of the pore, but not streptavidin binding. Altogether, the observed labelling pattern indicates high spatial labelling selectivity of our TurboID approach. Hence, we purified biotinylated proteins by streptavidin affinity for each cell line in triplicate and analyzed by LC-MSMS.

Defining confidence intervals: Proteins enriched in the BioID samples were grouped into confidence intervals (SigA, SigB, SigC) defined by cut-off curves in volcano plots (Figure 2) and based on statistical analysis in Perseus (18) (for details see Methods part). All BioID samples were compared to parental cells to identify proteins naturally biotinylated as well as proteins that non-specifically bind to the affinity matrix. A second control, cells expressing a GFP-TurboID fusion, served to identify proteins biotinylated in a non-specific manner.

For the abundant cytoplasmic proteins PABP1 and PABP2, we chose strict parameters to define significance intervals for both the bait proteins and the GFP control (SigA: FDR=0.01; s0=0.1, SigB: FDR=0.05; s0=0.1). This defined 35 proteins as GFP-positive, 13 in SigA and 24 in SigB (Table S1A and B, Figure 2, Figure 3A), which were removed from the list of PABP1 and PABP2 candidate interactors. The usage of less strict parameters for the GFP-control (SigC: FDR=0.05; s0=2; as used for NUP158 or MEX67 below) would have resulted in the subtraction of bona fide PABP1/2 interactors, as for example
PAB1 binding protein PBP1 and even the bait PABP1 itself. Among the GFP-positive proteins were many involved in translation or associated with cytoskeleton or membranes (19–24), as well as large proteins (26%>100 kDa; whole genome: 10%>100 kDa). Only two of this cohort were unique to the GFP BioID experiment, likely reflecting non-specific interactions while the remaining proteins were also identified in other BioID experiments (usually in more than one; nine in all five experiments).

We also applied SigA and SigB cutoffs to the less abundant proteins with a confined localization at the NPC, NUP158 and MEX67. However, we chose less strict parameters to define the GFP control (sigA: FDR=0.01; s0=0.1, SigC: FDR=0.05; s0=2) (Table S1 a and c; Figure 2, Figure 3B). With these parameters, 139 proteins were defined as GFP-positive (13 of these in SigA). About 50% of these 139 proteins were uniquely enriched in the GFP control. The others were removed from the list of NUP158 or MEX67 enriched proteins, except for proteins that were in SigA for the bait proteins and in SigC for GFP (listed in Figure 3B). The latter cohort contained an enrichment in proteins with nuclear localization and even two proteins with localization to the NPC, including the bona fide T. brucei NPC component NUP132 (11, 25). In contrast, very few nuclear-localized proteins were among confirmed GFP-positive proteins (Figure 3B).

In conclusion, we determined the appropriate filtering parameters to define the GFP control data for each individual BioID bait to exclude obvious false positives, while including known interaction partners. Notably, being GFP positive does not necessarily exclude the protein from being a true interacting protein; MARP2 (TbBBP268) was identified as GFP positive here but was identified by BioID with basal body protein TbPAC11 and TbBBP46 as baits. MARP2 localization to the basal body indicates that this is likely a true interactor (21). Likewise, some of the translation initiation factors and the RNA helicase within the cohort of GFP positives, may be true PABP interactors, highlighting the need for careful consideration of the dataset.

Cytoplasmic poly(A) binding proteins: T. brucei has two cytoplasmic poly(A) binding proteins. Both PABP1 and PABP2 have cytoplasmic localizations under physiological conditions, but PABP2 localizes to stress granules when cells are
starved (26, 27) and can also localize to the nucleus under certain conditions (26, 28). We have previously determined an interactome for PABP1 and PABP2 by cryomill affinity capture using GFP single chain antibodies, from cells expressing C-terminal fusions of the PABPs to eYFP from endogenous loci (10). Data from four experiments (two with high and two with low salt conditions) were analyzed to yield a high confidence list of PABP PPIs that included only proteins that were at least two-fold enriched in all four experiments, had no more than one zero-detection in the bait sample replicates and were no obvious contaminants. This resulted in a cohort of 13 proteins co-immunoprecipitated with PABP1, of which two proteins (eIF4E4 and eIF4G3) dominated with extremely high enrichment ratios of >150. PABP2 co-immunoprecipitated 26 proteins (Figure 4A). All proteins co-immunoprecipitated with either PABP1 or PABP2 have either a known function in T. brucei mRNA metabolism, are known mRNA binders (Tb927.7.7460 and Tb927.11.14750, (29)) or have a localization reflecting RNA granules (Tb927.4.4940, (20)), suggesting a cohort of genuine PPIs. In contrast, the number of proteins identified with BioID was larger and included 178 (PABP1) and 250 (PABP2) proteins in SigA and 263 (PABP1) and 330 (PABP2) proteins in SigB (Table S2).

We first asked which of the affinity purified proteins were also identified with BioID (Figure 4A). Of 13 proteins identified in the PABP1 pulldown, four were in SigA of the BioID experiment, including the bait and the two highly enriched proteins eIF4E4 and eIF4G3. A further three proteins were in SigB. Six proteins were not enriched in the BioID experiment. Four of these are low molecular weight proteins (Figure 4A) and may be missed due to their lower number of detectable unmodified peptides, which we recognized as a systematic problem of BioID (discussed in conclusion). The remaining two proteins, PUF9 and UPF1, were the least enriched proteins in the PABP1 pulldown and hence may suggest a less stable or physically more distant and indirect association with PABP1. Of the 26 proteins identified in the PABP2 pulldown, 15 are also in BioID SigA, including the bait and six proteins with highest enrichment ratios in the pulldown. A further six proteins were in the SigB cohort. Of the remaining five, three are low molecular weight and may not be detected as suggested above. The fourth is a large nuclear protein (CBP110)
that possibly remains in the cytoskeleton fraction removed prior to incubation with the streptavidin beads. The fifth protein is, again, UPF1.

We next asked how many proteins identified by BioID were also present in the pulldown datasets (Figure 4B). Of the 178 proteins identified in the PABP1 BioID (SigA), 25 were present in at least two replicates of the pulldown, 30 were enriched in one replicate, 50 were not enriched but detected and 73 were not detected. Of the 250 proteins from the PABP2 BioID (SigA) we detected 72 in at least two of the replicates of the pulldown, a further 31 in one replicate, 44 not enriched but detected and 103 proteins undetected. PABP interacting proteins are expected to be mainly involved in mRNA metabolism. To evaluate these data further, we manually assessed evidence for involvement in mRNA metabolism for each protein, based on conserved domains, known homologues and published data. 40 and 49% of proteins PPIs identified by BioID in confidence group SigA for PABP1 and PABP2 respectively had evidence for a role in mRNA metabolism. As expected, the number of proteins with a role in mRNA metabolism was highest (>75%) among the BioID proteins that were also identified by affinity capture and lowest (<25%) among the BioID proteins undetected by affinity isolation.

While there is overlap between BioID and pulldown PPI datasets, BioID identifies a significantly larger number of potential interactors, partly explained by the complexity of PABP interactions and the dynamic nature of ribonucleoprotein complexes in which the PABPs function. Consistent with this, a recent study (30) applied BioID to elucidate the PPIs of two T. brucei RNA-binding proteins, RBP9 and RBP10, uncovering a similar number (>200) of high confidence interactors.

BioID appears to have a false positive rate among the high confidence SigA cohort of about 50%. This is likely due to the less-confined (cytoplasmic) localization of PABP1 and PABP2 and perhaps additionally by their function: PABPs are mobile and interact with many complexes which themselves may have peripheral proteins not obviously involved in mRNA metabolism. Moreover, PABPs are involved in translation and thus in close proximity to many nascent proteins that they potentially biotinylate despite the absence of direct interaction. Note, however, that ribosomal proteins were not found biotinylated, excluding random biotinylation of the entire polysomal complex.
**MEX67:** MEX67 forms a heterodimer with Mtr2 and this complex is the major mRNA export complex and is conserved across most eukaryotes (31, 32). MEX67 binds its mRNA targets both directly or indirectly via adaptor proteins and mediates mRNA export to the cytoplasm by interacting with FG-type nucleoporins. While mRNA export in trypanosomes differs in several aspects from the pathway described in opisthokonts (33), TbMEX67 (Tb927.11.2370) has been well studied in the past and appears to have a conserved function in mRNA export (17, 33–35). Two datasets from pulldowns are available for *T. brucei* MEX67. A classical immunoprecipitation with a C-terminal PTP-tag identified Mtr2 and importin 1, a transport receptor for nuclear import (35), while a cryomill affinity capture using a C-terminally GFP-tagged MEX67 as bait identified many nucleoporins (11). We re-analyzed the latter affinity capture data (11) for three different extraction conditions to generate a list of enriched proteins (Table S4b). We have then performed BioID with MEX67 to compare the PPIs identified by these different methods.

In contrast to the PABP1 and PABP2 datasets, the number of MEX67 PPIs identified with BioID (99) was similar to the number of PPIs identified by pulldown (118 proteins) (Figure 5A). We suggest that this is likely a consequence of the more confined localization of the MEX67 protein, which results in less bystander labeling.

MEX67 BioID identified only nine proteins with a non-nuclear localization, spread over both SigA and SigB cohorts, and the majority of proteins were localized to the nucleoplasm (69 proteins) or the NPC (17 proteins) (Figure 5A). SigA contained 15 core nucleoporins, while SigB showed enrichment of proteins localized to the nucleoplasm. Proteins with nuclear localization only identified by BioID could be proteins that interact with nuclear proteins only transiently, preventing their detection in the pulldown experiment.

Pulldown-identified PPIs differ from the corresponding BioID PPIs in containing a larger number of nucleoporins and a smaller number of nucleoplasmic proteins, and, in particular for confidence group B a higher number of non-nuclear proteins. Also specific to the pulldown PPIs were proteins localized to the endoplasmic reticulum (ER) and nuclear envelope: these likely co-precipitated with the NPCs, but are not in close contact with
MEX67, and hence not sampled with BioID (Figure 5A). However, it should be considered, that any transmembrane, vesicular or intraluminal (perinuclear cisterna) proteins might also be inaccessible to the BioID labeling but not to direct physical connection mediated affinity capture.

While MEX67 co-precipitated essentially the entire NPC, MEX67 BioID selectively identified six of nine FG-repeat proteins exposed to central channel (Figure 5B). The only non-FG-repeat NUP identified by BioID is NUP132, which is also enriched with the GFP-bait control (SigB) and may thus have higher exposure to random biotinylation. Thus, for MEX67 BioID captured only proteins directly interacting with MEX67, while the pulldown identified the entire NPC, likely driven by high affinity interactions between individual nucleoporins.

Similar to PABPs, there was little concordance between the PPIs identified by the two different methods, and only 15 proteins out of 202 were common (Figure 5C, Table S3B). Additional to MEX67, there were 12 proteins with localization to the NPC or nuclear envelope, namely (i) eight nucleoporins (compare Figure 5B), (ii) Tb927.6.890, unique to Kinetoplastida, but possessing a SAC3/GANP/THP3 domain at its N-terminus; Sac3 from yeast is a well-known interactor of MEX67-Mtr2 that localizes to the cytoplasmic side of the NPC (36), (iii) Tb927.7.2330 and Tb927.11.5560, two proteins of unknown function, and finally (iv) Tb927.10.7680, a GTPase activating protein (RabGAP-TBC domain, TBC-RootA (37)); TBC-RootA is likely involved in the GTP-dependent mRNA transport via the small GTPase Ran (11). Ran, Ran-binding protein and MEX67b (38) all high confidence interactors in the pulldown, were also enriched in the BioID PPI dataset, but below a significance cut-off. Two further proteins common to both BioID and pulldown are cyclophilin (localization to the nucleolus, unknown function) and KMP11 (a protein of the basal body and flagellum attachment zone) (20, 39), both likely contaminants. Surprisingly, the major MEX67 interactor, Mtr2, was only identified by pulldown, likely due to the low molecular weight of Mtr2 (15.8 kDa).

103 proteins were unique to the pulldown, 61 of these in confidence group A (ConfA). Most (40 proteins) of these localized to the NPC or nuclear envelope and included most nucleoporins (compare Figure 5B), (putative) transport proteins including Ran-binding protein 1, importin 1, importin beta, Ran RTB2
and many hypothetical proteins. Four proteins localized to the nuclear envelope and ER. Seven had nuclear localizations, including the RNA binding proteins DRBD4, DBP2B and DRBD18 while ten had non-nuclear or unknown localizations.

In contrast, of 84 proteins unique to the BioID dataset, 68 had a nuclear localization, four with NPC and only 12 had non-nuclear or unknown localization. NPC-associated proteins unique to the BioID were exportin-1 (XPO1), NMD3 and Ran-binding protein RanBPL. Interestingly, NMD3 is implicated in nuclear export of mRNA of procyclin-associated genes, a process shown to be disrupted by silencing of NMD3, XPO1 or MEX67 (40). This finding extended the prototypic function of NMD3 and XPO1 in rRNA export and is suggestive of overlapping, interdependent nuclear export routes for mRNA and rRNA, for which there is also evidence in yeast (41). While the absence of NMD3 and XPO1 in the pulldown argues against a stable interaction with MEX67, their high enrichment in BioID datasets suggests transient interactions with MEX67.

Among the nuclear-localized proteins were many RNA-binding proteins and proteins with functions in nuclear mRNA processing, including spliceosomal factors (U5-200K, RBSR4/U2AF65, splicing factor 3B, U4/U6 small nuclear ribonucleoprotein PRP3), an exosome subunit RRP6, the non-coding poly(A) polymerase NPAPL, the cap guanylyltransferase-methyltransferase 1 (CGM1) and proteins involved in transcription. However, the majority of proteins known to be involved in nuclear mRNA metabolism (33) was not identified with either method (Figure 5D).

To summarize, both BioID and pulldown detected the expected interactions of MEX67 with FG-type nucleoporins. In fact, these are likely the major interactions, as in yeast, MEX67 can be fused to such a nucleoporin and still fulfil its essential functions (42). Apart from these FG-type nucleoporins, there was little coincidence between the PIPs identified by the two methods. While the pulldown mainly co-immunoprecipitated the entire NPC, BioID identified nuclear proteins involved in mRNA metabolism and many RNA binding proteins, potentially reflecting transient interactions between MEX67 and its mRNA substrates that are missed in the pulldown.
Altogether, BioID is a valuable tool for interrogation of MEX67 function as it clearly discriminates components of the NPC engaging in direct interaction, as the subset NUPs exposed the inner pore channel. On the other hand, the pulldown detects many indirect interactions due to high stability of NPC subunit interactions. The latter provides equally meaningful information, as MEX67 indeed, is considered to exhibit some of the characteristics of a mobile nucleoporin, due to its significant presence at the NPC (42). Furthermore, BioID appears to identify transient interactions, as indicated by the enrichment of proteins with a function in mRNA binding and nuclear localization, albeit requiring validation.

The nucleoporin NUP158: NUP158 is an FG-NUP/alpha-solenoid and orthologue of outer ring nucleoporins ScNup145 and HsNup98-96. Cryomill affinity capture was repeated as described (11) but subjecting the entire elution to triplicate LC-MSMS-analysis. We detected co-precipitated 23 nucleoporins, 8 falling into SigA and 5 into SigB (Figure 6B; Figure S1; Table S3). Among these are NUP109, Sec13, NUP41, NUP82, NUP89, NUP132 and NUP152; together with NUP158 these eight proteins form the outer ring complex (11). Further significant interactors are the nuclear basket protein NUP110, NUP76, NUP48 and inner ring proteins NUP65, NUP96 and NUP225. Many additional NPC-associated proteins such as NUP140, the inner ring proteins NUP62, NUP53b and NUP119, the FG NUPs NUP 64 and NUP98 and the lamina protein NUP-1 were also enriched but fall outside the SigA/B cutoff. 18 additional proteins were identified as significant interactors (Table S3), of which five localize to the NPC and further five to the nucleus (20).

NUP158 BioID identified 53 proteins (Figure 6A; Table S3): 10 proteins in SigA (including 5 nucleoporins), a further 12 in SigA that also appear in the GFP-control SigC detections, and 31 proteins in SigB. In comparison to the PABP and MEX67 BioID experiments, this is the shortest list of biotinylated proteins, likely reflecting the highly confined localization of NUP158. Of these 53 proteins, 13 localized to the NPC, 16 to the nucleus and five to the ER / nuclear envelope (20).

Concordance between BioID and pulldown PPIs was once again poor (Figure 6B). Many nucleoporins identified as interaction partners in the
pulldown were not identified in the BioID (NUP140, NUP76, NUP41, NUP89, SEC13, NUP82, NUP152, NUP110) and, vice versa, others were unique to the BioID (NUP144, NUP149, NUP53a) or outside the pulldown SigA/B cutoff (NUP98 and NUP64). The only nucleoporins in common were FG-NUPs NUP109 and NUP13, and there was just one additional interactor Tb927.11.13080 shared between pulldown and BioID datasets (Table S3). This latter is a protein of unknown function but localizes to the NPC (20)). Hence both methods identify meaningful, but significantly different PPIs for NUP158. The pulldown identifies adjacent outer ring proteins that are partly absent in BioID. A potential reason for the absence could be that NUP158, located within the rigid structure forming the outer rings of the NPC, has a limited labelling radius due to immobility, steric constraints or even the local biochemical environment. Therefore, although pulldown is superior identification of PPIs for a strictly localized nucleoporin, additional valuable data can be obtained with BioID.

Conclusions: Collectively, we found surprisingly little concordance between proteins identified by BioID and by affinity capture. This strongly suggests that these methods should be viewed as providing a distinct PPI for a given bait rather than offer equivalents or alternatives. A similar comparative interactomics study to the one here has been performed for several chromatin-associated protein complexes in human cell lines (43, 44) with broadly similar outcome; BioID generally produced larger interactomes and the concordance between pulldown and BioID was poor, albeit that both techniques identified meaningful PPIs. Our dataset however analyzed a cohort of bait proteins with considerably greater variation of location and positional constraints to provide additional insight.

For soluble, unconstrained cytoplasmic proteins with a high potential for nonspecific interactions, BioID has a significant level of bystander labeling. By contrast, BioID with MEX67, a more restricted bait, identified exclusively FG-NUPs of the inner pore channel, while pulldown co-precipitated the entire NPC. Thus, the optimal method depends on the localization of the protein of interest. Ideally, both methods are used in parallel and recently a hybrid tag was introduced combining both biotin ligase and an epitope tag for pulldown (45).
Several variants of BioID aiming to overcome some of the pitfalls of BioID are available and include an inducible system with target-specific biotinylation only occurring when the biotin ligase is attached to the bait using a dimerizing agent (46) and a split BioID, where biotinylation can only occur once two proteins carrying partial BioID tags interact (conditional biotinylation) (47).

A major weakness of BioID beyond bystander identifications is that many low-molecular weight proteins are unidentified. Most strikingly here was the absence of Mtr2 from the MEX67 BioID data. In general, small proteins are harder to detect by MS than large proteins as there are simply less peptides to detect, but this applies to both pulldown and BioID. However, firstly, BioID relies on surface exposed lysine residues, and probability of occurrence depends on protein size amongst other parameters. Secondly, in pulldown analysis peptides of the entire protein are potentially available for MS, as proteins are fully, or near fully, eluted from the beads, but in BioID peptides are eluted from beads by trypsin digestion and any biotinylated peptides remain attached to the beads (Table S4). For low-molecular weight proteins, the likelihood that lysine residues are inaccessible is lower than for larger proteins, as small proteins are less likely to ‘bury’ lysine residues in their core. This can explain the systematic absence of expected interactors with a low molecular weight from BioID but not from the pulldown. One potential way to increase elution of biotinylated peptides would be to add excess biotin before the proteolytic digest, to saturate all biotin binding sites and hence restrict re-binding of biotinylated peptides to the beads after their release by the protease. There is also an alternative approach for purification: trypsination of cell lysates prior to purification followed by purification of biotinylated peptides using a biotin-antibody rather than streptavidin has been shown to result in a higher recovery rate of biotinylated peptides (48, 49). However, this approach comes at the cost of increased background from non-specific binding, excluding the use of stringent washes as facilitated by the strong biotin streptavidin interaction.

Altogether, this study offers guidance for choosing the most appropriate method for protein complex characterization, dependent on localization and positional constraints. Furthermore, it highlights potential pitfalls concerning experimental design and data interpretation. Overall, BioID and affinity capture are complementary with each approach elucidating a unique subset of PPIs for
a given target protein. Hence combination of both methods can be leveraged for more complete interactome mapping.

**Experimental Procedures**

*Cell culture and genetic manipulation: T. b. brucei* Lister 427 procyclic cells were cultured in SDM-79 (51). The generation of transgenic trypanosomes was done using standard methods (52). All fusion proteins were expressed from their endogenous locus as described (53), modified to result in fusing TurboID and one hemagglutinin (HA) tag to the C-terminus of the protein. eYFP was expressed fused to TurboID-HA using an inducible expression system based on tetracycline (53).

*Western blot:* Western blotting was performed using standard methods. Detection of biotinylated proteins was done with Streptavidin-IRDye®680LT (LI-COR) (1:10,000). BiP was detected using anti-BiP (1:200,000) (kind gift of James D Bangs, University at Buffalo, US).

*Immunofluorescence:* 1X107 cells at 5x106 cells/ml were washed once in 1 ml SDM79 without hemine and serum and resuspended in 500 µl PBS (phosphate buffered saline). For fixation, 500 µl 8% paraformaldehyde was added for 20 minutes while rotating. 7 ml PBS with 20 mM glycine were added, cells were pelleted, resuspended in 150 µl PBS and spread on polylysine-coated slides (in circles drawn with a hydrophobic pen). After 15 minutes cells had settled to the slide, surplus PBS was removed, and cells permeabilized with 0.5% TritonX100 in PBS. Slides were rinsed in PBS and cells then blocked in 3% BSA in PBS for 30 minutes, followed by 60 minutes incubation with rabbit-mAb-anti-HA C29F4 (Cell Signaling Technology) 1:500 and with Streptavidin-Cy3 (Jackson Laboratories) 1:200 in PBS/3%BSA. Slides were washed in PBS (3 x 5 minutes) and incubated with anti-rabbit Alexa Fluor Plus 488 1:500 in PBS/3%BSA for 1 h; the last 10 minutes were done in the presence of DAPI (0.1 µg/ml). Slides were washed 3 x 5 minutes in PBS and embedded into ProLongTM Diamond Antifade Mountant (Thermofisher). Images were recorded as Z-stacks (100 images with 100 nm distance) on a custom build TILLPhotonics iMic microscope equipped with a Sensicam camera (PCO AG, 6.45m/pixel) und 100x oil immersion (NA1.4) objectives (Olympus). Filter sets were (i) ex: 320–
380 nm, dc: 400–430 nm, em: 438–486 nm (DAPI), (ii) ex: 430–474 nm, dc: 585 nm, em: 489–531 nm (Alexa Fluor Plus 488) and (iii) ex: 540–580 nm, dc: 585 nm, em: 592–664 nm (Cy3). For each image, the exposure times were 50 ms for DAPI and 500 ms for other all other fluorophores. Images were deconvolved using Huygens Essential software (SVI, Hilversum, The Netherlands) and are either presents single plane or as Z-stack projection (sum slices), as indicated.

**Affinity enrichment of biotinylated proteins and on-bead tryptic digests:** Cells were maintained at a density of 1×10^6 to 10^7 cells per ml and harvested at a cell density of 5×10^8 cells per ml. The eYFP control was induced by addition of tetracycline at a concentration of 1 μg/ml 24 h prior to harvesting. No extra-biotin was added for induction of labelling, as we found the biotin concentration in the SDM79 medium (827 nM) to be sufficient for high level biotinylation.

5×10^8 cells were used in each replicate. Cells were harvested at 1,400 g, washed once with serum-free medium and pellets rapidly frozen in liquid nitrogen and stored at -80°C. For isolation of biotinylated proteins, each cell pellet was resuspended in 1 ml lysis buffer (0.5% octylphenoxypolyethoxyethanol (IGEPAL), 0.1 M piperazine-N,N′-bis(2-ethanesulfonic acid) (PIPES)-NaOH pH 6.9, 2 mM ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N″-tetraacetic acid, 1 mM MgSO_4_, 0.1 mM ethylenediaminetetraacetic acid (EDTA), complete EDTA-free protease inhibitor cocktail (Roche)) and incubated for 15 min at room temperature in an orbital mixer. Soluble and non-soluble fractions were separated by centrifugation (14,000 g, 5 min, 4°C) and the soluble fraction incubated with 100 μl streptavidin-linked Dynabeads (MyOne Streptavidin C1, Thermofisher) for 1 hour at 4°C under gentle mixing. Beads were washed twice in 1 ml buffer 1 (2% (w/v) SDS in water) once in 1 ml buffer 2 (0.1% (w/v) deoxycholate, 1% Triton X-100, 1 mM EDTA, 50 mM HEPES pH 7.5, 500 mM NaCl), once in 1ml buffer 3 (250 mM LiCl, 0.5% IGEPAL, 0.5% (w/v) deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8.1) and once in 1 ml buffer 4 (50 mM Tris-HCl pH 7.4, 50 mM NaCl); each washing step was eight minutes at room temperature (RT) under orbital shaking. Beads were then prepared for tryptic digestion by washing three times in 500 μl ice-cold 50 mM NH_4_ HCO_3_, resuspension in 40 μl of the same buffer supplemented with 10 mM dithiothreitol and incubation in a
thermomixer at RT for 1h. Iodoacetamide was added to a concentration of 20 mM, followed by incubation in the dark at RT for 30 min. Finally, 5 μg/ml proteomics-grade trypsin (SOLu-Trypsin, SigmaAldrich) was added to the beads. The digest was done overnight at 30°C in a thermomixer (1000 rpm). After removal of the first eluate, beads were resuspended in 50 μl 50 mM NH₄HCO₃ supplemented with 10 mM dithiothreitol and 5 μg/ml mass spectrometry (MS) grade trypsin and incubated in a thermomixer at 37°C for 1h. The eluate was combined with the first eluate, and both were lyophilized in a Speed-vac (Christ alpha 2-4). LoBind tubes (Eppendorf) were used throughout.

Mass spectrometry and proteomics analysis: BioID eluted peptides were resuspended in 50 mM NH₄HCO₃ and passed over C₁₈ stage tip columns as described (54), then analyzed by liquid chromatography-tandem mass spectrometry (LC-MSMS) on an Ultimate3000 nano rapid separation LC system (Dionex) coupled to an LTQ Q-exactive mass spectrometer (Thermo Fisher Scientific). Resulting spectra were processed using the intensity-based label-free quantification (LFQ) in MaxQuant version 1.6.16 (55, 56). Minimum peptide length was set at six amino acids allowing a maximum of 2 missed cleavages and false discovery rates (FDR) of 0.01 were calculated at the levels of peptides, proteins and modification sites based on the number of hits against the reversed sequence database. Data from TurboID were additionally searched for lysine biotinylated peptides (+ 226.078 Da). LFQ data were analyzed using Perseus (18). For statistical analysis, LFQ values were log₂ transformed and missing values imputed from a normal distribution of intensities around the detection limit of the mass spectrometer. These values were subjected to a Student’s t-test comparing an untagged control (wt cells) triplicate sample group to the TurboID tagged-protein triplicate sample groups, including the GFP control. -log₁₀ t-test p-values were plotted versus t-test difference to generate multiple volcano plots (Hawaii plot, see Figure 2). Potential interactors were classified according to their position in the Hawaii plot, applying cut-off curves for ‘significant class A ’ (SigA; FDR=0.01, s0=0.1), ‘significant class B ’ (SigB; FDR=0.05, s0=0.1) and ‘significant class C ’ (SigC; FDR=0.05, s0=2), respectively. The cut-off is based on the false discovery rate
(FDR) and the artificial factor s0, controlling the relative importance of the t-test p-value and difference between means (At s0 = 0 only the p-value matters, while at non-zero s0 the difference of means contributes).

**Cryomill affinity capture:** For affinity capture of PABP1, PABP2 and MEX67 we used data from our previous studies (10, 11). MEX67 data were re-analyzed for three different extraction conditions (higher stringency buffer 1: 20mM HEPES, pH 7.4, 250mM NaCl, 0.5% Triton, plus protease inhibitors (Roche) and two low stringency stabilizing buffers: 2A - 20mM HEPES, pH 7.4, 250mM sodium citrate, 0.5% Tween-20, plus protease inhibitors (Roche), and buffer 2B - 20mM HEPES, pH 7.4, 20mM NaCl, 50mM sodium citrate, 0.5% Tween-20, plus protease inhibitors (Roche)) and compared to a negative control, omitting the GFP tag. Potential interactors were assigned to two confidence intervals termed Confidence A (ConfA) for proteins enriched >2 fold under all three conditions, and Confidence B (ConfB) for proteins enriched >2 fold in at least one condition. The NUP158 cryomill affinity enrichment purification analysis was repeated essentially as described in (11) but subjecting the entire elution to triplicate LC-MSMS-analysis. In brief, NUP158 endogenously tagged with eYFP at the C-terminus, was extracted in buffer 3 (20 mM HEPES pH7.4, 250 mM NaCl, 0.5% CHAPS, complete EDTA-free protease inhibitor cocktail (Roche)), captured on magnetic anti-GFP nanobody beads (GFP-Trap Magnetic Agarose, Chromotek, Germany) and washed four times with buffer 3. Proteins were eluted by on-bead tryptic digest and analyzed by LC-MSMS on an Ultimate3000 nano rapid separation LC system (Dionex) coupled to an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific). Data were analyzed as described (57) and detailed above.

All proteomics data have been deposited at the ProteomeXchange Consortium via the PRIDE partner repository (58) with the data set identifier PXD031245. The proteomics data for the PABP1 and PABP2 pulldown has the data set identifier PXD008839.
Figure legends

Figure 1: Biotinylation in the different BioID cell lines. A) Western blot probed with fluorescently labelled streptavidin to detect biotinylated proteins in the control cell lines (WT=wild type, eYFP=TurboID fused to eYFP) or cell lines expressing the respective bait proteins (PABP1, PABP2, MEX67, NUP158) fused to TurboID. The expected molecular weight of all bait proteins fused to TurboID is indicated. B) Localization of bait proteins and biotin labeling. Cells were probed with anti-HA antibody and IRDye 800CW streptavidin to detect biotinylated proteins, then DAPI stained analyzed by fluorescence microscopy.

Figure 2: Hawaii plot for the statistical analysis of BioID experiments. Hawaii plot (multiple volcano plots) of LFQ results of the BioID experiments for PABP1, PABP2, MEX67, NUP158 and the GFP-control. All samples were prepared in triplicate. To generate the volcano plots, the $-\log_{10} P$-value was plotted versus the $t$-test difference (difference between means), comparing each respective bait experiment to the wt control. Potential interactors were classified according to their position in the plot, applying cut-off curves for ‘significant class A ’(SigA; drawn in red; FDR=0.01, s0=0.1) and 'significant class B ' (SigB; drawn in blue; FDR=0.05, s0=0.1), respectively, and, for the GFP control 'significant class C ' (SigC; drawn in pink; FDR=0.05, s0=2). Bait proteins are indicated by a green dot.

Figure 3: Proteins biotinylated by the GFP-TurboID control. A cell line expressing GFP-TurboID was used to detect background biotinylation. We have used different thresholds to define a protein as GFP-positive for the different bait proteins (Sig A, B and C; details in the text). For further comparison, we have added respective TurboID data for the trypanosome mRNA decapping enzyme ALPH1 (50), which localizes to the cytoplasm and posterior pole granules. A) To control the BioID experiment of PABP1 and PABP2, proteins biotinylated by GFP-TurboID were defined using strict parameters (35 proteins, falling into SigA and SigB): only the proteins shown
here were considered GFP positive. All GFP positive proteins were removed from the list of proteins identified in the BioID with PABPs as baits, independent of the significance group. B) To control the BioID experiment of MEX67 and NUP158, we used less strict parameters to define the control (139 proteins, falling into SigA and SigC, were considered GFP-positive). However, proteins that were identified with very high confidence with MEX67 and NUP158 BioID were included, if the GFP control was in significance group C. Protein localization is mainly taken from TrypTag (20).

Figure 4: Poly(A) binding proteins: comparing BioID and pulldown. A) All proteins significantly enriched in all four replicates of the PABP1 and PABP2 pulldown experiment are shown ranked according to their enrichment ratio (10). Proteins that are also identified in the respective BioID experiment and matching significance group A or B criteria, are shown with black or grey filling, respectively. The molecular weight of all proteins is shown below, as we noticed a certain bias for loss of small proteins in BioID. Possible reasons of why a protein is absent from the BioID data are indicated. B) All proteins identified with the PABP BioID experiments in significance group A are listed, split into four groups, dependent on whether these proteins were found enriched in more than one of the pulldown experiments, enriched in only one of the pulldown experiments, detected in the pulldown experiments but not enriched or not detected. For each group of proteins, the fraction of proteins involved in mRNA metabolism is shown as a pie diagram (Table S2b,c).

Figure 5: MEX67: comparing BioID and pulldown. A) Proteins identified in the MEX67 pulldown and BioID experiments are grouped according to their cellular localizations based on TrypTag (20). A corresponding color-code scheme is shown on the left. B) Schematic representation of the trypanosome NPC modified from (11). Proteins identified in the BioID or pulldown experiments are filled in red or light red for significance group A or B, respectively. C) Comparison between proteins identified in the pulldown and the BioID. All proteins identified with both methods are listed, as well as selected proteins for proteins exclusively identified in the BioID or pulldown. D) List of proteins involved in selected nuclear mRNA maturation processes (33)
and in translation (eIF4F, PABPs). Only three of these proteins were identified in the BioID experiment (red asterisk) in significance group A (black, bold) or B (black, regular). TREX: couples transcription and export; CBC: cap binding complex; EJC: exon junction complex; TRAMP: Trf4-Air2-Mtr4, Polyadenylation. Full details on the MEX67 BioID experiment are listed in Table S3b. A corresponding color-code scheme is shown on the left.

**Figure 6: BioID with NUP158. A)** All proteins identified in the NUP158 BioID are listed, grouped according to their intracellular localizations based on TrypTag (20). **B)** Schematic representation of the trypanosome NPC modified from (11). Proteins identified in the NUP158 BioID or pulldown experiments are filled in red or light red for significance intervals A or B, respectively. The bait protein NUP158 is filled in yellow. Proteins enriched in the pulldown falling outside the SigA/B cutoff are filled in grey.

**Data availability**
All proteomics data have been deposited at the ProteomeXchange Consortium via the PRIDE partner repository (58) with the data set identifier PXD031245. The proteomics data for the PABP1 and PABP2 pulldown has the data set identifier PXD008839.

**Supporting information**
**Figure S1: Volcano plot for the statistical analysis of the NUP158 pulldown.** Samples were prepared in triplicate. To generate the volcano plot, the $-\log_{10} P$-value was plotted versus the $t$-test difference (difference between means), comparing the bait experiment (C-terminally eYFP- tagged NUP158) to the wt control. Potential interactors were classified according to their position in the plot, applying cut-off curves for ‘significant class A ‘(SigA; drawn in red; FDR=0.01, s0=0.1) and ‘significant class B ‘(SigB; drawn in orange; FDR=0.05, s0=0.1), respectively. Known nucleoporins and SigA interactors are labelled. The bait protein is indicated by a green dot.

**Table S1: Mass spectrometry data BioID GFP control.**
Table S2: Mass spectrometry data PABP1 and PABP2.
Table S3: Mass spectrometry data MEX67 and NUP158.
Table S4: Lysine biotinylated peptides detected in TurbolD experiments

Acknowledgements

We thank Silke Braune for excellent technical assistance and the TrypTag consortium, which is supported by the Wellcome Trust [108445/Z/15/Z] for freely providing data which enabled part of this work. We are grateful to the FingerPrints proteomics facility at the University of Dundee, which is supported by the 'Wellcome Trust Technology Platform' award [097945/B/11/Z], to Mark Carrington (University of Cambridge, UK) for plasmids, to Jay Bangs (University at Buffalo, US) for provision of the BiP antibody.

Funding and additional information

The work was financed by a bilateral GACR/DFG grant [project 21-19503J and KR4017/9-1 to MZ and SK, respectively], the bilateral DAAD/CAPES grant to FH and SK [project ID 57390761], the Czech Ministry of Education (project OPVVV/0000759 to MZ), the Wellcome Trust (204697/Z/16/Z to MCF) and the Medical Research Council (MR/P009018/1 to MCF). The Mex67 work was funded by: NSF 1818129 (to Samson Obado), NIH R01 AI140429 (to Samson Obado and Michael P. Rout), NIH R01 GM112108 (to Michael P. Rout (MPR)), NIH P41 GM103314 (to Brian T. Chait (BTC)), NIH U54 GM103511 and NIH P41 GM109824 (to MPR and BTC); and Wellcome Trust grant 082813/Z/07/Z (to MCF). We thank MPR and BTC for sharing the MEX67 pulldown data.

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article
References


monitor the composition of spatiotemporally defined protein complexes. *Nature communications.* 8, 1–14


FIGURE 1

A

kDa

170
130
100
70
55
35
25

Strep

BiP

100 99 200 99 64

Mw [kDa] (incl. tag)

100
99
200
99
64

B

Strep α-HA DAPI Strep α-HA DAPI Strep α-HA DAPI Strep

PABP1

PABP2

NUP158

MEX67

3 µm
FIGURE 2

The figure shows scatter plots for the t-test difference and -log10 p-value for different proteins:
- **PABP1**
- **PABP2**
- **MEX67**
- **NUP158**
- **GFP**

Each plot represents the distribution of t-test differences across different conditions, with the -log10 p-value on the y-axis. The green dots indicate significant differences.
Proteins associated with a structure or membrane
Proteins with a function in translation

BB: basal body; NL: nuclear lumen; C: cytoplasm
FP: flagellar pocket; CK: cytoskeleton; N: nucleus; FC: flagellar cytoplasm; NP: nuclear pore; FL:

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>MW (kDa)</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tb927.10.10280</td>
<td>268.0</td>
<td>BB</td>
</tr>
<tr>
<td>Tb927.11.12800</td>
<td>39.1</td>
<td>NL</td>
</tr>
<tr>
<td>Tb927.2.2740</td>
<td>16.3</td>
<td>C, NL, FL</td>
</tr>
<tr>
<td>Tb927.6.5010</td>
<td>40.8</td>
<td>C, NL, FL, spots</td>
</tr>
<tr>
<td>Tb927.9.6560</td>
<td>80.2</td>
<td>C</td>
</tr>
<tr>
<td>Tb927.5.4460</td>
<td>94.1</td>
<td>C</td>
</tr>
<tr>
<td>Tb927.10.1500</td>
<td>86.9</td>
<td>C</td>
</tr>
<tr>
<td>Tb927.3.5530</td>
<td>279.2</td>
<td></td>
</tr>
<tr>
<td>Tb927.4.2070</td>
<td>511.3</td>
<td>C, CK</td>
</tr>
<tr>
<td>Tb927.4.4550</td>
<td>92.1</td>
<td>C</td>
</tr>
<tr>
<td>Tb927.7.3980</td>
<td>97.3</td>
<td>C, points, cell tip</td>
</tr>
<tr>
<td>Tb927.10.11230</td>
<td>34.7</td>
<td>N</td>
</tr>
<tr>
<td>Tb927.10.5610</td>
<td>52.5</td>
<td></td>
</tr>
<tr>
<td>Tb927.10.3110</td>
<td>65.3</td>
<td>N</td>
</tr>
<tr>
<td>Tb927.6.4410</td>
<td>41.7</td>
<td>C, FC</td>
</tr>
<tr>
<td>Tb927.7.2300</td>
<td>132.0</td>
<td>NP</td>
</tr>
<tr>
<td>Tb927.8.2820</td>
<td>141.1</td>
<td>cell tip, FP</td>
</tr>
<tr>
<td>Tb927.10.2770</td>
<td>43.0</td>
<td></td>
</tr>
<tr>
<td>Tb927.10.12800</td>
<td>77.7</td>
<td></td>
</tr>
<tr>
<td>Tb927.10.1450</td>
<td>329.6</td>
<td>biolobe</td>
</tr>
<tr>
<td>Tb927.10.1510</td>
<td>259.1</td>
<td>C</td>
</tr>
<tr>
<td>Tb927.11.1270</td>
<td>48.0</td>
<td>C</td>
</tr>
<tr>
<td>Tb927.11.13280</td>
<td>34.7</td>
<td>C</td>
</tr>
<tr>
<td>Tb927.11.1966</td>
<td>42.6</td>
<td>Axoneme, C</td>
</tr>
<tr>
<td>Tb927.3.2600</td>
<td>245.0</td>
<td>C</td>
</tr>
<tr>
<td>Tb927.3.3030</td>
<td>41.3</td>
<td>FP collar</td>
</tr>
<tr>
<td>Tb927.6.3100</td>
<td>48.1</td>
<td>iFT particle</td>
</tr>
<tr>
<td>Tb927.7.2840</td>
<td>50.8</td>
<td>CK</td>
</tr>
<tr>
<td>Tb927.7.4420</td>
<td>32.1</td>
<td>C, N, nucleolus</td>
</tr>
<tr>
<td>Tb927.7.4970</td>
<td>42.4</td>
<td></td>
</tr>
<tr>
<td>Tb927.8.6260</td>
<td>60.3</td>
<td></td>
</tr>
<tr>
<td>Tb927.8.2030</td>
<td>48.2</td>
<td>CK</td>
</tr>
<tr>
<td>Tb927.8.4500</td>
<td>84.6</td>
<td>C</td>
</tr>
<tr>
<td>Tb927.8.4950</td>
<td>165.1</td>
<td>BB, cell tip, FP</td>
</tr>
<tr>
<td>Tb927.9.13920</td>
<td>11.1</td>
<td></td>
</tr>
</tbody>
</table>

Proteins associated with a structure or membrane

BB: basal body; NL: nuclear lumen; C: cytoplasm
FP: flagellar pocket; CK: cytoskeleton; N: nucleus; FC: flagellar cytoplasm; NP: nuclear pore; FL:

<table>
<thead>
<tr>
<th>Significance group A (SigA)</th>
<th>Significance group B (SigB)</th>
<th>Not significantly enriched</th>
</tr>
</thead>
<tbody>
<tr>
<td>13 proteins</td>
<td>126 proteins</td>
<td></td>
</tr>
</tbody>
</table>

Significance group A (SigA)
Significance group B (SigB)
Not significantly enriched
A

13 proteins enriched in PABP1

26 proteins enriched in PABP2

Enrichment ratio (fold)

MW [kDa]

1 High MW protein; absent due to cytoplasmic localisation
2 Low MW protein
3 Unclear why absent in BioID

1 High MW protein; absent due to cytoplasmic localisation
2 Low MW protein
3 Unclear why absent in BioID

BioID enriched (Sig A)
BioID enriched (Sig B)
Not enriched in BioID
MW <25 kDa
MW >25 kDa

B

178 proteins enriched in PABP1 BioID

250 proteins enriched in PABP2 BioID

Total

Enriched in > two cryomill replicates
Enriched one cryomill replicate
Not enriched by cryomill but detected
Not detected in cryomill data

Involved in mRNA metabolism (strong evidence)
Involved in mRNA metabolism (some evidence)
No evidence for involvement in mRNA metabolism
**Figure 5**

A. Diagram illustrating the NPC/nuclear envelope and BioID pulldown confidence interval.

B. BioID and pulldown of vertebrate homologues to these proteins have direct interaction with the nuclear export factor NXT1.

C. Table listing selected proteins and their functions.

D. Diagram showing the TREX and EJC complexes and their associated proteins.
**BioID**

**Nuclear pore complex, envelope**
- Tb927.3.3180 NUP98
- Tb927.4.4310 NUP64
- Tb927.11.980 NUP158
- Tb927.7.2330 Protein of unknown function (DUF1663)
- Tb927.11.13080 hypothetical protein, conserved
- Tb927.10.8170 NUP144
- Tb927.2.4230 NUP1
- Tb927.10.9020 Gcd10p family
- Tb927.7.2300 NUP132
- Tb927.11.11080 NUP149
- Tb927.10.8650 RanBPL
- Tb927.10.7680 GTPase activating protein
- Tb927.11.15560 NUP53a
- Tb927.11.14340 XPO1
- Tb927.11.15990 NUP109
- Tb927.8.5770 ELP3a

**Nucleus**
- Tb927.10.11210 hypothetical protein, conserved
- Tb927.10.12110 hypothetical protein, conserved
- Tb927.10.7420 bromodomain factor 2 protein
- Tb927.2.5240 pre-mRNA splicing factor 19
- Tb927.8.3310 Elongator-like Protein 3b
- Tb927.10.5330 ATP-dependent RNA helicase, U5-200K
- Tb927.1.3270 hypothetical protein, conserved
- Tb927.6.3870 RNA-binding protein, SGN1
- Tb927.10.14390 cap guanyltransferase-methyltransferase 1
- Tb927.1.3450 hypothetical protein

**Localisation includes ER**
- Tb927.11.13230 VAMP-associated protein, VAP (mainly FAZ, FP)
- Tb927.5.4500 ras-like small GTPase, Sec13
- Tb927.9.9550 EnigmRB2, ERBP2
- Tb927.10.13210 dolichyl-P-Man:GDP-ManGlcNAc2-PP-dolichyl
  beta-1,4-mannosyltransferase
  (function indicates ER localisation)

**Other localisations**
- Tb927.3.1240 Serine/threonine-protein phosphatase 2A
- Tb927.3.1030 RNA-binding protein
- Tb927.4.960 WD domain, G-beta repeat
- Tb927.9.7110 GRAM domain containing protein
- Tb927.5.4380 Kinetoplastid-specific Protein Phosphatase 1
- Tb927.10.3100 glycerol-3-phosphate acyltransferase, putative

*Also found in GFP BioID significance group B*

**Significance Group A (SigA)**

**Significance Group B (SigB)**

**B**

**BioID**

**Pulldown**

- NUP132 detected in BioID with GFP as bait, but only SigB

- IgA
- IgB
- enriched but not SigA/B (pulldown only)
- Bait

- NUP132 detected in BioID with GFP as bait, but only SigB
CRediT author statement

Susanne Kramer: Conceptualization, Supervision, Writing - Original draft preparation.
Martin Zoltner: Conceptualization, Supervision, Writing - Original draft preparation.