Molecular & Biochemical Parasitology 249 (2022) 111476

Contents lists available at ScienceDirect

Molecular & Biochemical Parasitology

journal homepage: www.elsevier.com/locate/molbiopara

CRISPR/Cas9-based precision tagging of essential genes in bloodstream form African trypanosomes

Julie Kovářová [a,b,1,2], Markéta Novotná [a,1,3], Joana Faria [a,c,4], Eva Rico [a,5], Catriona Wallace [a], Martin Zoltner [a,d,6], Mark C. Field [a,e,7], David Horn [a,*,8]

[a] The Wellcome Trust Centre for Anti-Infectives Research, School of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, UK
[b] Institute of Parasitology, Biology Centre, Czech Academy of Sciences, Branišovská 31, 37005 České Budějovice, Czech Republic
[c] York Biomedical Research Institute, Department of Biology, University of York, Heslington, York YO10 5DD, UK
[d] Faculty of Science, Charles University in Prague, Biocev, Vestec, Czech Republic
[e] Institute of Parasitology, Biology Centre, Czech Academy of Sciences, Branišovská 31, 37005 České Budějovice, Czech Republic

ARTICLE INFO

Keywords:
Trypanosoma brucei
CRISPR-Cas9
Epigenetics
Histone modification
Gene regulation

ABSTRACT

Proteins of interest are frequently expressed with a fusion-tag to facilitate experimental analysis. In trypanosomatids, which are typically diploid, a tag-encoding DNA fragment is typically fused to one native allele. However, since recombinant cells represent ≪0.1% of the population following transfection, these DNA fragments also incorporate a marker cassette for positive selection. Consequently, native mRNA untranslated regions (UTRs) are replaced, potentially perturbing gene expression; in trypanosomatids, UTRs often impact gene expression in the context of widespread and constitutive polycistronic transcription. We sought to develop a tagging strategy that preserves native UTRs in bloodstream-form African trypanosomes, and here we describe a CRISPR/Cas9-based knock-in approach to drive precise and marker-free tagging of essential genes. Using simple tag-encoding amplicons, we tagged four proteins: a histone acetyltransferase, HAT2; a histone deacetylase, HDAC3; a cleavage and polyadenylation specificity factor, CPSF3; and a variant surface glycoprotein exclusion factor, VEX2. The approach maintained the native UTRs and yielded clonal strains expressing functional recombinant proteins, typically with both alleles tagged. We demonstrate utility for both immunofluorescence-based localisation and for enriching protein complexes; ^{GFP}HAT2 or ^{GFP}HDAC3 complexes in this case. This precision tagging approach facilitates the assembly of strains expressing essential recombinant genes with their native UTRs preserved.

1. Introduction

Trypanosoma brucei are eukaryotic parasites responsible for sleeping sickness and nagana, neglected tropical diseases affecting humans and cattle, respectively, in sub-Saharan Africa. Almost all genes in these African trypanosomes are organised in a polycistronic manner, meaning that many genes are transcribed in a single transcription unit, and co-transcriptionally processed into individual mRNAs [1]. Sites of transcription initiation and termination are subject to control by histone modifiers [2,3], enzymes that also impact the expression of variant surface glycoprotein (VSG) genes [4]. These latter genes are expressed in a monogenic manner, with VSG switching and antigenic variation allowing these parasites to continually evade host immune responses [5]. Notably, enzymes involved in gene expression and its control

* Corresponding author.
E-mail address: d.horn@dundee.ac.uk (D. Horn).
[1] ORCID: joint first authors
[2] ORCID: 0000-0001-8737-6403
[3] ORCID: 0000-0001-9760-9119
[4] ORCID: 0000-0001-6274-8143
[5] ORCID: 0000-0001-6388-514X
[6] ORCID: 0000-0002-0214-285X
[7] ORCID: 0000-0002-4866-2885
[8] ORCID: 0000-0001-5173-9284

https://doi.org/10.1016/j.molbiopara.2022.111476
Received 27 January 2022; Received in revised form 28 March 2022; Accepted 30 March 2022
Available online 1 April 2022

present a rich source of established [6] and potential drug targets in *T. brucei* and in the related parasitic trypanosomatids, *T. cruzi* and *Leishmania spp*. [7].

Experimental studies on genes and proteins of interest often involve the use of recombinant DNA techniques to add a fluorescent protein or epitope tag, thereby facilitating (immune)fluorescence microscopy, to establish localisation within the cell, or affinity purification combined with proteomics, to identify interacting partner proteins, for example. Such recombinant approaches have been widely adopted since it can be challenging to develop mono-specific antibodies to native proteins to facilitate such studies. However, there are some caveats to be considered when using conventional gene tagging approaches in trypanosomatids. First, the relatively low-efficiency homologous recombination-based approaches employed currently involve adding a selectable marker cassette adjacent to the tag, which disrupts potential regulatory untranslated regions (UTRs). This may impact native gene expression control, a particular concern in trypanosomatids given the reliance upon post-transcriptional controls in the context of global polycistronic transcription. Current tagging approaches typically replace native UTRs with UTRs from highly expressed genes, such as tubulin, actin, aldolase or procyclin. Second, it often remains unknown whether a single tagged allele is functional since the remaining allele may be sufficient to maintain function and/or viability.

Genetic manipulation in *T. brucei* typically relies upon homologous recombination, using the native DNA repair machinery. The efficiency of site-specific integration of recombinant DNA is greatly increased if the target locus is damaged by a double-strand break [8]. Thus, the natural DNA damage thought to yield conventional recombinants is likely random, occurring primarily during DNA replication, and explaining why current approaches are inefficient. In this context, a selectable marker allows for these infrequent natural breaks to be exploited, allowing even one in a million recombinant cells to be selected, following elimination of the non-transformed and antibiotic-sensitive background population. The emergence of high-efficiency CRISPR--Cas9 (clustered regularly interspaced short palindromic repeats and CRISPR-associated protein from *Streptococcus pyogenes*) technologies applied to *T. brucei* [9,10] presents opportunities to increase the efficiency of site-specific double-strand breaks and to address the caveats detailed above. Indeed, an episome-based Cas9 expression system has been used to precision tag at least one SCD6 allele in insect stage *T. brucei* [10]. In this case, a second episome carried both the single guide RNA (sgRNA) expression cassette and the repair template along with > 400 bp of flanking homology.

Here, we tested the capacity of efficient and inducible Cas9-based editing to obviate the use of co-integrated selectable markers when tagging genes of interest in bloodstream form *T. brucei*. We achieved precision tagging of four essential genes, with both alleles tagged in three cases. Native untranslated regions remained intact and no selectable marker was integrated adjacent to the tagged genes. We report simple amplicon templates encoding a green fluorescent protein (GFP) tag, but the approach should be compatible with other fluorescent or epitope tags. Thus, high-efficiency, precision gene editing facilitates the generation of bloodstream form *T. brucei* strains expressing tagged essential proteins that remain subject to native UTR-based controls.

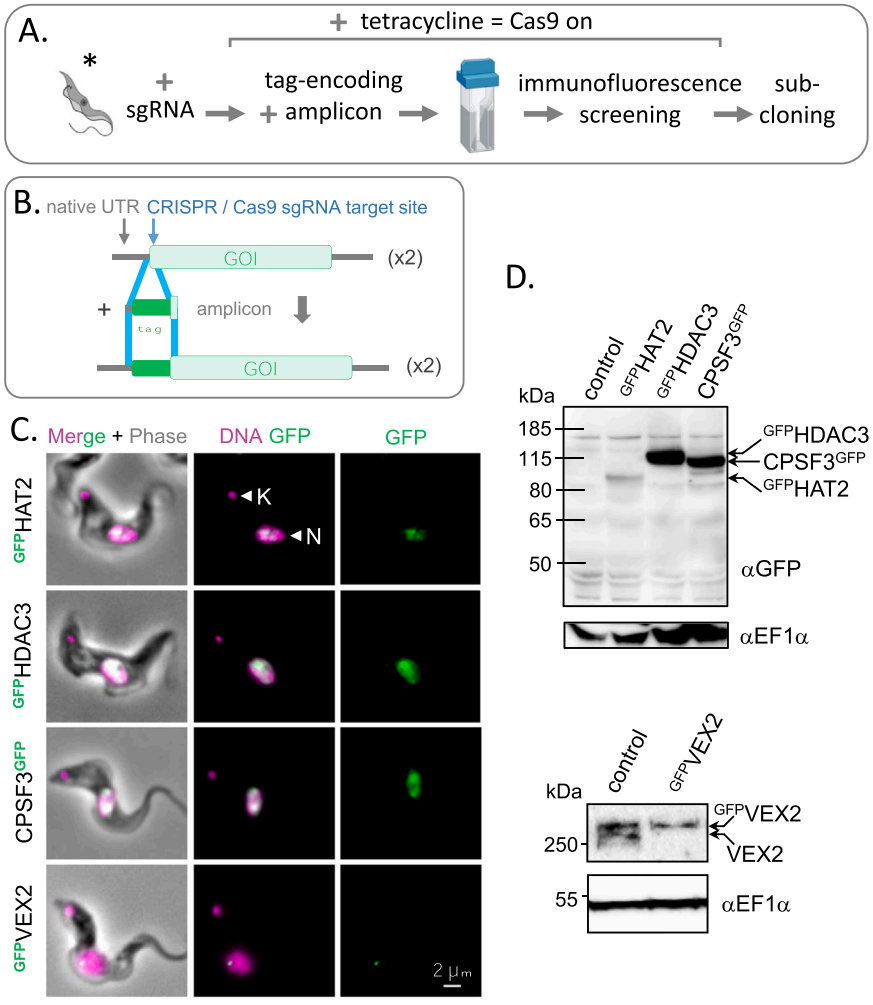


Fig. 1. Precision tagging driven by Cas9; A. The schematic shows the Cas9-driven precision-tagging protocol. * The $2T1^{T7\text{-}Cas9}$ strain expresses tetracycline repressor, T7 RNA polymerase and *Sp*Cas9 [9]. Tetracycline was added either 3 h prior to transfection (HAT2, HDAC3 and CPSF3) or immediately following transfection (VEX2) and maintained thereafter, to the point of sub-cloning. B. The schematic shows the Cas9-driven precise insertion of the GFP template adjacent to the gene of interest (GOI). 25 bp of target homology on either side of the GFP sequence is indicated by the blue bars. sgRNA, single guide RNA. C. The immunofluorescence microscopy images show nuclear localisation of each GFP-tagged protein. DNA was stained with DAPI (4′,6-diamidino-2-phenylindole) which stains the nucleus (N) and kinetoplast (K, mitochondrial DNA). D. The protein blots show expression of each GFP-tagged protein. ^{GFP}HAT2, 95 kDa; ^{GFP}HDAC3, 102 kDa; $CPSF3^{GFP}$, 112 kDa; ^{GFP}VEX2, 260 kDa. Elongation factor 1α served as a loading control, in addition to non-specific bands observed in the upper panel. VEX2 was detected using αVEX2 primary antibody. The control is a sample from a strain lacking a GFP tag in the upper panel and with a single VEX2 allele tagged with GFP in the lower panel; the latter strain was generated using a conventional selectable marker based approach.

2. Results and discussion

2.1. Precision tagging driven by Cas9

We previously established inducible Cas9-based gene disruption and precision base editing in *T. brucei* [9]. Here, we sought to use this high-efficiency system for precise integration of tag-encoding DNA fragments. The schematic in Fig. 1A illustrates the strategy. Briefly, *T. brucei* 2T1$^{T7\text{-}Cas9}$ cells, in which the Cas9 endonuclease can be induced upon addition of tetracycline, were transformed with an appropriate sgRNA sequence within a pT7^{sgRNA} plasmid. Cas9 was then induced, and the cells were subsequently transfected with the *GFP* repair template to initiate homology directed repair. Transfected cultures were periodically assessed for GFP fluorescence and, when appropriate, sub-clones were screened for GFP fluorescence. The sgRNA was designed to target a site as close as possible to the appropriate start or stop codon, based on available protospacer adjacent motifs (PAMs, 5′-NGG for *Sp*Cas9), while the *GFP* repair template was an amplicon generated by PCR, incorporating 25 bp of target homology on either side of the GFP sequence (Fig. 1B).

We reasoned that Cas-driven precision tagging would be more easily achieved when targeting genes that are essential for cell viability. This is because breaks induced by Cas9 can either promote tag-integration or lead to gene disruption. The latter output would likely dominate in the case of a dispensable gene, given little impact on cell viability. In contrast, gene disruption would result in loss of viability in the case of an essential gene, while tag-integration would block further Cas9-induced breaks; see GFP repair template design section in Materials and Methods. We, therefore, selected four known essential genes for assessment; Histone AcetylTransferase 2 (HAT2) [11], Histone DeACetylase 3 (HDAC3) [4], Cleavage and Polyadenylation Specificity Factor 3 (CPSF3) [6] and VSG Exclusion 2 (VEX2) [12]. Following GFP template transfection, mixed cultures were assessed using immunofluorescence microscopy between two to nine days later, revealing between 1% and 51% GFP-positive cells (Table 1). Sub-cloning then yielded between 3% and 10% GFP-positive clones (Table 1). Further immunofluorescence microscopy confirmed the expected nuclear localisation for each protein, with a single sub-nuclear focus in the case of VEX2 (Fig. 1C), and protein blotting revealed the expression of each tagged protein, with relatively lower abundance of HAT2 (Fig. 1D). In the case of VEX2, the availability of an αVEX2 antibody allowed us to visualise both the native and recombinant VEX2 proteins simultaneously (Fig. 1D, lower panel). These results indicate that Cas9 can be used to drive the site-specific integration of simple GFP cassettes to generate fusions with essential genes in *T. brucei*.

2.2. Precision tagging of both alleles

Although the GFP template fragment was available only transiently, we maintained Cas9 expression, induced by tetracycline, throughout the analysis period above (see Fig. 1A). This was for negative selection, to drive the elimination of untagged cells due to loss of an essential gene, and to promote conversion from one tagged allele to two tagged alleles in those cells that had initially integrated one copy of the GFP cassette.

Table 1
Proportions of GFP positive cells and sub-clones.

GOI	GeneID	% GFP+ cells (days)	Subcloning at day	% GFP+ clones (n° screened)
HAT2	Tb927.11.11530	10 (3) / 17 (7) / 51 (9)	~15 (after freeze/thaw)	10 (20)
HDAC3	Tb927.2.2190	1 (2) / 1 (4) / 5 (8)	8	3 (34)
CPSF3	Tb927.4.1340	5 (5) / 9 (7)	7	8 (52)
VEX2	Tb927.11.13380	1 (2) / 7 (6) / 10 (9)	10	13 (24)

We used Southern blotting to assess both alleles of each *T. brucei* gene targeted for tagging in GFP positive clones. HAT2, HDAC3 and CPSF3 tagged strains were assessed in parallel, by digesting genomic DNA with *Nsi*I, while VEX2 tagged strains were assessed by digesting genomic DNA with *Age*I/*Eco*RV; using the selected probes, these digests yield distinct fragments depending upon whether the GFP cassette is integrated or not (Fig. 2A).

The Southern blots revealed modification of both native alleles, and fragments consistent with GFP insertion in all four cases (Fig. 2B). Quantification, using independent probes as loading controls, allowed us to determine whether one or both alleles were modified by GFP insertion. This analysis revealed GFP insertion at both alleles in all clones analysed, except for the HDAC3 clone, in which only one allele was fused to GFP (Fig. 2B); the second allele is likely disrupted by deletion of the region containing the probe in this case. DNA sequencing confirmed the expected GFP-junction sequences for all four tagged genes (Supplemental Fig. S1). We conclude that the Cas9 driven GFP integration approach described here typically yields cells with both targeted alleles tagged and with native untranslated regions intact. We can also conclude that the recombinant ^{GFP}HAT2, ^{GFP}HDAC3, CPSF3^{GFP} and ^{GFP}VEX2 proteins are functional, since these genes, for which no unmodified native alleles remain in the clones analysed, are known to be essential for viability. Other advantages relate to protein stoichiometry and strain stability, in that the full pool of the targeted protein in these cells is similarly tagged and tagged alleles are unlikely to be lost through gene conversion as no unmodified alleles remain.

Addition of a tag could produce hypomorphs, i.e. a gene/protein that has lower potency than the unmodified allele. To assess the impact of the tag in each case, we measured the growth of each tagged clone in comparison to untagged parental cells. This analysis revealed no growth defect in the HAT2, HDAC3 or VEX2-tagged strains and only a moderate growth defect in the CPSF3-tagged strain, which displayed a maximum of < 5% reduction in relative cell number over five days (Fig. 2 C). Thus, the GFP tags have little to no impact on those proteins' function in terms of cell viability.

2.3. Affinity purification of tagged protein complexes

We demonstrated utility of the current precision-tagged clones above using immunofluorescence analysis (Fig. 1D). To further demonstrate utility, we used tagged proteins as bait to affinity purify associated proteins following cryo-milling, subsequently identifying those proteins using quantitative proteomics. For this analysis, we selected *T. brucei* HAT2, which is a histone H4K10 acetyltransferase enriched at transcription initiation sites [2,11] and HDAC3, which has been linked to *VSG* silencing [4]. GFP-affinity purification, using the ^{GFP}HAT2 and ^{GFP}HDAC3 strains, revealed a specific set of associated proteins in each case (Fig. 3). Comparison with a recent analysis by others [3], revealed similar sets of co-purified proteins. Common HAT2 interacting factors include the putative acetyl-lysine binding bromodomain containing factors, BDF3 and BDF5, and a set of six hypothetical conserved proteins (Fig. 3A). Common HDAC3 interacting factors include BDF2, the H2AZ and H2BV histone variants known to be enriched at transcription initiation sites, the telomere-associated protein TelAP1, DNA primase, casein kinase, and a hypothetical conserved protein, 3.2460 (Fig. 3B). These findings are consistent with the established role for HAT2 in propagating histone H4K10 acetylation at transcription initiation sites [2,3,11,13] and for HDAC3 in removing acetyl groups from histones at both transcription initiation sites and at telomeric sites [3,4]. We also observe enrichment of H2AZ in association with HAT2 and enrichment of DNA ligase and SNF2 (sucrose nonfermenting 2) in association with HDAC3.

Our affinity purification analyses are broadly consistent with recently published data, while those proteins that were specifically enriched here may reflect differences in abundance and stoichiometry due to precision-tagging. In particular, HDAC3 association with DNA ligase and DNA primase, suggests a role for this histone deacetylase in

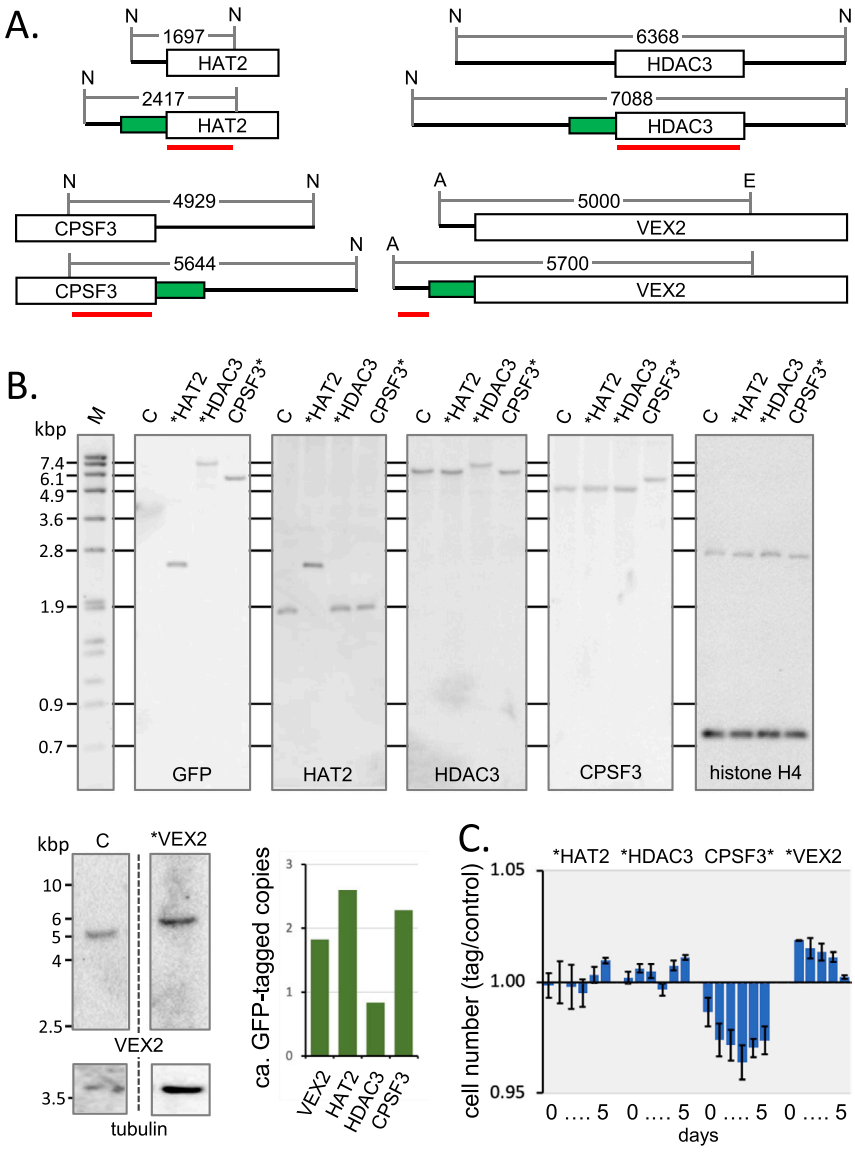
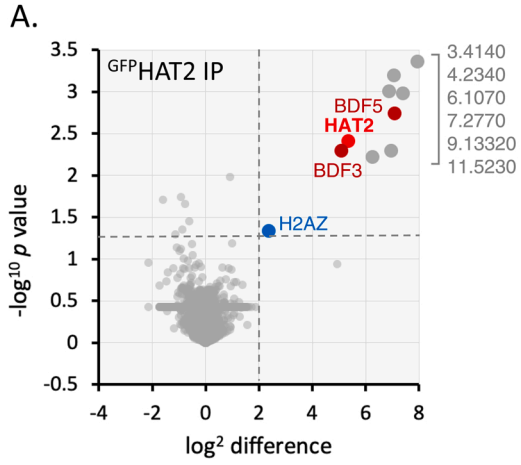
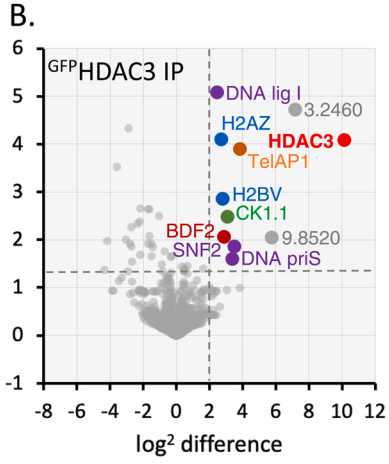


Fig. 2. Precision tagging of both alleles; A. The schematic maps show the size in bp of each band expected following digestion with either *Nsi*I (N; HAT2, HDAC3 and CPSF3) or with *Age*I/*Eco*RV (A/E; VEX2). GFP regions are indicated in green and the location of each probe is also indicated by red bars. B. The Southern blots were hybridised with the probes indicated. A histone H4 (HAT2, HDAC3 and CPSF3) or tubulin probe (VEX2) was subsequently used as a loading control for each blot; a representative rehybridised histone H4 blot is shown. Quantification was performed by measuring the intensity of each band relative to the loading control. The resulting values were then normalised to the wild-type control strain (C, WT), to determine the number of alleles tagged. * indicates the location of the GFP tag. C. Growth analysis. The plot reveals deviation for each precision-tagged strain relative to a control strain. Error bars, SD.





C.

GeneID	Product description	difference	unique peptides	-log10 p-value	
Tb927.11.5230	hypothetical protein, conserved	7.94	24	3.36	
Tb927.7.2770	hypothetical protein, conserved	7.38	28	2.98	
Tb927.11.13400	BDF5	7.10	47	2.73	
Tb927.4.2340	hypothetical protein, conserved	7.06	42	3.20	
Tb927.6.1070	hypothetical protein, conserved	6.96	33	2.29	
Tb927.9.13320	hypothetical protein, conserved	6.88	27	3.00	
Tb927.3.4140	hypothetical protein, conserved	6.25	23	2.22	
Tb927.11.11530	HAT2	5.36	28	2.41	HAT2
Tb927.11.10070	BDF3	5.10	17	2.29	
Tb927.7.6360	H2AZ	2.37	8	1.33	
Tb927.2.2190	HDAC3	10.17	39	4.08	
Tb927.3.2460	hypothetical protein, conserved	7.16	15	4.72	
Tb927.9.8520	hypothetical protein, conserved	5.75	16	2.05	
Tb927.11.9870	Telomere-associated protein 1	3.84	9	3.89	
Tb927.3.5440	SNF2 DNA repair protein, putative	3.53	49	1.86	
Tb927.7.2310	DNA primase small subunit, putative	3.38	7	1.59	
Tb927.5.790	CK1.1	3.12	27	2.48	
Tb927.10.7420	BDF2	2.92	12	2.05	
Tb927.11.7350	H2BV	2.80	9	2.86	
Tb927.7.6360	H2AZ	2.71	8	4.10	
Tb927.6.4780	DNA ligaseI, putative	2.46	5	5.07	

HDAC3

Fig. 3. Affinity purification of tagged protein complexes; Statistical analysis of affinity purified proteins following immunoprecipitation, mass-spectrometry and label free quantification. To generate the volcano plots, the $-\log_{10}$ *P*-value was plotted versus the *t*-test difference (difference between means), comparing the bait experiments ^{GFP}HAT2 (A) and ^{GFP}HDAC3 (B) to untagged control cells lacking GFP. Samples were prepared in triplicate. The dashed lines indicate a cutoff (difference between means >2; and -$\log_{10}$*p*-value >1.3 including significantly enriched proteins, which are labelled. C. Statistical data and unique peptide numbers for significantly enriched proteins as determined by label-free quantification. Difference and *p*-values were derived from a Student's *t*-test in Perseus.

controlling DNA replication and/or repair. Indeed, sites of transcription initiation in *T. brucei*, where HDAC3 is enriched [3], largely coincide with origins of DNA replication initiation [14]. Further supporting this hypothesis, the *T. cruzi* orthologue of HDAC3-associated BDF2 displays nuclear accumulation in response to UV irradiation [15], while *T. brucei* BDF2 binds acetylated H2AZ [16], a histone variant that facilitates DNA replication in mammalian cells [17].

2.4. Concluding remarks

Protein abundance and stoichiometry can have major impacts on phenotype and, since native untranslated regions impact expression in trypanosomatids, a native gene tagging approach that preserves these regions will facilitate functional analyses. We show that Cas9-driven gene editing can be used to precisely tag both alleles of essential genes in *T. brucei*. The utility of the approach was demonstrated by tagging four distinct proteins. Microscopy revealed subcellular localisation, while affinity purification of ^{GFP}HAT2 and ^{GFP}HDAC3, followed by mass-spectrometry, revealed associations with specific histone variants, bromodomain proteins, and other conserved as well as uncharacterised factors. Precision tagged strains could similarly be used for chromatin immunoprecipitation, for structural studies or for in vitro analysis of purified proteins or complexes, for example.

The efficiency of 'double-allele' precision tagging remains relatively low, meaning that 20–50 sub-clones must typically be screened, and also that the current approach cannot readily be applied to dispensable genes. Fluorescence activated cell sorting could be used to enrich tagged cells, but this would be limited to sufficiently abundant/bright fluorescent tags that are detectable in live cells. The current precision-tagging approach is, therefore, not currently a readily scalable approach. It is, however, a useful addition to the Cas9-toolbox that can be applied to high priority drug targets and other genes of interest. Notably, scalable tagging approaches that employ selectable marker genes are yielding vast datasets relating to protein subcellular localisation in trypanosomatids [18,19]. These datasets will certainly facilitate prioritisation of genes for precision tagging, as will genome-scale fitness profiling data [20].

We use a GFP tag here, but the approach should be equally successful with any tag that can be readily prepared as an amplicon and is detectable by immunofluorescence microscopy. We demonstrate tagging at either the *N*- or *C*-terminus and suspect that the approach will work equally well for other proteins, at either terminus, or internally, where tolerated.

3. Materials and methods

3.1. T. brucei growth and manipulation

T. brucei Lister 427 bloodstream form 2T1$^{T7\text{-}Cas9}$ cells [9] were grown in HMI-11 medium (Gibco), supplemented with 10% foetal bovine serum (Sigma-Aldrich) at 37 °C in 5% CO_2 and split daily to a density of 1×10^5 cells/ml. Growth analysis was carried out in the absence of tetracycline and cells were counted using a haemocytometer. 1 μg/ml hygromycin (Sigma-Aldrich) and 2 μg/ml blasticidin (Melford) were added to culture media to select for trypanosomes containing the pRPa^{Cas9} construct, and pT7pol. Transfection with pT7^{sgRNA} constructs was carried out in 100 μl of cytomix using a 2 mm gap cuvette (BioRad) and an Amaxa nucleofecter (Lonza) set on the X-001 programme. Phleomycin selection (2 μg/ml) was applied 4 h post-transfection and clones were selected by limiting dilution in 48-well plates. Positive clones were selected after 5 days and phleomycin selection was subsequently maintained at 1 μg/ml. 1 μg/ml of tetracycline (Sigma-Aldrich) was added to induce Cas9 expression.

3.2. sgRNA construct design and assembly

Oligonucleotides encoding sgRNAs were designed by first identifying protospacer adjacent motif (PAM) sequences (5′-NGG) close to the start or stop codon, followed by selection of 20 bp PAM-preceding regions with > 40% GC-content. The oligonucleotide pairs were:

HAT2 - H2G5 (AGGGGGATGCGAGAGGTTGCGCAA) / H2G3 (AAACTTGCGCAACCTCTCGCATCC).

HDAC3 - D3G5 (AGGGTTGTTGAACCGCAACCGCAA) / D3G3 (AAACTTGCGGTTGCGGTTCAACAA).

CPSF3 - C2G5 (AGGGTGAACCGGCTCCCATTGACG) / C2G3 (AAACCGTCAATGGGAGCCGGTTCA).

VEX2 - V2G5 (AGGGAACAAGGATATGGATGACCA) / V2G3 (AAACTGGTCATCCATATCCTTGTT). Oligonucleotides were resuspended in 400 μl Milli-Q H_2O, resulting in 100 μM solution, and 5 μl of each pair were added to 6 μl 10x Tris-borate-EDTA (TBE) buffer and 44 μl Milli-Q H_2O which was then heated to 70 °C for 3 min. Mixtures were then cooled slowly to allow annealing prior to ligation with pT7^{sgRNA} [9] digested with *Bbs*I (New England BioLabs). Successful cloning was verified by DNA sequencing and the resulting constructs (10 μg / transfection) were linearised with *Not*I (New England Biolabs) and purified by ethanol precipitation prior to transfection.

3.3. GFP repair template design, amplification, and transfection

GFP repair template primers were designed to amplify the green fluorescent protein (GFP) coding sequence and to add 25-bp homology arms targeting the gene of interest. The following oligonucleotide pairs were used:

HAT2 -

H2TGFP5 (AGTGGATGCGAGAGGTTGCGCAATGTCTAGA*GTGAGCAAGGGCGAGGAG*) /

H2TGFP3 (CGTCTTTTTTTGCTGCTAACGACGC*CTTGTACAGCTCGTCCATGC)*.

HDAC3 - D3TGFP5 (CATTTGTTGAACCGCAACCGCAATGTCTAGA*GTGAGCAAGGGCGAGGAG*) / D3TGFP3 (TTTCTGCTTCACGCGTTTCTTTGCCC*TTGTACAGCTCGTCCATGC*).

CPSF3 - C1TGFP5 (TGATTCACGGTGCTGAACCGGCTCC**G**AT**C**GA**T**-GAA**GG**A**GC**C*GTGAGCAAGGGCGAGGAG*) / C1TGFP3 (ACCAGTTTTTCCCCATACTCTCTGT*TTACTTGTACAGCTCGTCCATG*).

VEX2 - V2TGFP5 (TTTGTTATCTGTGAACAAGGATATGTCTAGA*GTGAGCAAGGGCG*) / V2TGFP3 (ACCTCGGGAAATGAGTACAACCATG**A**T**C**G**T***C**TTGTACAGCTCGTCCATGC*).

Underlined are *Xba*I sequences (TCTAGA), included to disrupt the sgRNA target region, in bold are synonymous mutations, included to disrupt the PAM or to disrupt the sgRNA target region, and in italics are the GFP template-matched sequences. Primer pairs were used to PCR-amplify each GFP cassette from a plasmid DNA template using Phusion polymerase (New England BioLabs). PCR products, ≥ 15 μg / electroporation, were ethanol precipitated and pellets were washed with 70% ethanol and resuspended in 100 μl Amaxa buffer. The DNA solution was mixed with 1×10^7 sgRNA-transformed cells. These cells were either pre-induced to express Cas9, with 1 μg/ml of tetracycline for 3 h (HAT2, HDAC3, CPSF3) or induced immediately following transfection (VEX2). The mixture was transferred to a 2 mm gap cuvette (BioRad) and transfected using an Amaxa nucleofecter (Lonza) set on the Z-001 programme. Tetracycline was subsequently maintained at 1 μg/ml to sustain Cas9 expression.

3.4. Immunofluorescence microscopy

1×10^6 cells were fixed in 3% paraformaldehyde (PFA) for 15 min at 37 °C and subsequently washed twice with 1 ml phosphate buffered saline (PBS), centrifuging at 4600 g for 1 min following each wash. Cells were resuspended in 30 μl of an ice cold 1% solution of bovine serum albumin (BSA). 5 μl were added and attached to each well of poly-L-

lysine coated slides for 15 min. Cells were rinsed in 1 x PBS for 5 min before being permeabilised in 0.5% Triton-X 100 in PBS for 15 min. Three washes were carried out in PBS before cells were blocked with a 1:1 foetal bovine serum (FBS):PBS solution for 15 min. Two subsequent 5-min washes in 1 x PBS were carried out. Primary antibody was then applied to the samples. Rabbit α GFP IgG (Invitrogen) was added to a 3% FBS in PBS solution in a 1:250 (VEX2 – low abundance) or 1:500 (HAT2, DAC3, CPSF3) dilution and slides were left at room temperature for a minimum of 1 h. Three more washes were carried out with 1 x PBS before application of the secondary antibody. Goat α rabbit IgG with Alexafluor-488 (Invitrogen) was added to a 3% FBS in PBS solution in a 1:2000 dilution and left in the dark at room temperature for 1 h. After 3 washes, slides were mounted in Vectashield with 4′,6-diamidino-2-phenylindole (DAPI) (Vector Laboratories) and the cover slip sealed on top. Slides were viewed by fluorescence and bright field microscopy on a Zeiss Axiovert 200 M microscope with an AxioCam MRm camera and the ZEN Pro software (Carl Zeiss, Germany). Z-stacks were acquired, with 20–30 slices at 0.2 μm intervals. Images were deconvolved using ZEN Pro software and further processed in Fiji version 2.0.0-rc-2.

3.5. Protein blotting

2×10^7 cells were washed in 1 x PBS by centrifuging at 1000 g twice for 10 min and once for 5 min. 15 μl of NaCl/Tween buffer (20 mM Hepes pH 7.4, 0.1% Tween-20, 1 mM $MgCl_2$, 10 μM $CaCl_2$, 100 mM NaCl) was added to each pellet. The samples were sonicated and centrifuged for 15 min at 4 °C. The supernatant was collected, and LDS sample buffer (Invitrogen) was added. The samples were heated to 70 °C for 10 min. Proteins were separated by SDS-PAGE at 200 V for 50 min, using 4–12% Bis-Tris gel (Invitrogen). Proteins were transferred to a PVDF membrane which was washed with 5% milk in TBS-Tween. Rabbit α GFP IgG (Invitrogen) was added to the membrane at a 1:2000 dilution in a 5% milk in TBS-Tween solution and left on a roller at 4 °C overnight. Following washing, secondary antibody, goat α rabbit IgG conjugated with horse radish peroxidase (HRP) was applied to the membrane and left for 1 h. ECL-Prime western blotting detection agents (GE Healthcare) were mixed in a 1:1 solution and applied to the membrane, which was incubated for 1 min then imaged. For VEX2 detection, 4×10^6 cells were washed in 1 x PBS by centrifuging at 1000 g twice for 10 min and once for 5 min 20 μl LDS sample buffer (Invitrogen) was added to each pellet. The samples were heated to 70 °C 10 min and proteins separated at 200 V for 1 h, using 4–12% Bis-Tris gels (Invitrogen). Proteins were transferred to a PVDF membrane, which was washed with 1% milk in PBS-Tween. Rabbit α VEX2 [12] was added to the membrane at a 1:1000 dilution in a 1% milk in PBS-Tween solution and left on a roller at 4 °C overnight. Washes, incubation with secondary antibody and development were performed as described above.

3.6. Southern blotting

Genomic DNA was extracted using a DNA extraction kit (Qiagen) according to the manufacturer's instructions. Approximately 10 μg of genomic DNA were digested with a 5-fold excess of *Nsi*I or *Age*I/*Eco*RV overnight at 37 °C. The samples were then run overnight in a 0.8% agarose gel. The gel was sequentially incubated with 0.25 M HCl, 1.5 M NaCl, 0.5 M NaOH, and 3 M NaCl, 0.5 M Tris.HCl pH 7. DNA was then transferred overnight onto a nylon membrane (Amersham), using 10 x SSC (saline sodium citrate: 300 mM sodium citrate, 1 M NaCl). Nucleic acids fixation was achieved through UV-crosslinking. Probe labeling, hybridization, washes and development were performed using DIG high prime DNA labeling and detection starter kit II (Roche) following the manufacturer's instructions. VEX2 5′-UTR, and parts of HDAC3, HAT2 or CPSF3 ORF sequences were used as probes, which were applied at 25 ng/ml. The VEX2 5′-UTR probe was amplified using OneTaq (NEB) and the following primers: fwd: GCGGCCGCACCTTCACAACACCGTACGTAATG; rev: AGATCTATCCTTGTTCACAGATAACAAAAGTTAGCGTGACTGTTTGC. The resulting amplicon was cloned into pGEM T-easy (Promega), sequenced, and the fragment to be used as probe generated through plasmid digestion with *Age*I/*Eco*RI.

Probes for HDAC3, HAT2 or CPSF3 were amplified using Q5 polymerase (NEB) and the following primers: HDAC3 fwd: GGCAAAGAAACGCGTGAAG, rev: GAATTGGGATTGAGGCACTG; HAT2 fwd: GACGGAGGTGACATCAATG, rev: GCCAGATTGTTGTCAGTCAC; CPSF3 fwd: GAAGCTGTACCAGACATTTG, rev: GTTCAGCACCGTGAATCATA. The PCR products were used as probes. Membranes were then stripped and rehybridized with a probe targeting histone H4 or tubulin, as a loading control. Densitometry was performed using Fiji v. 2.0.0.

3.7. Affinity purification and mass spectroscopy

Cultures of 8 L for each cell line (^{GFP}HAT2, ^{GFP}HDAC3, and parental strain) were harvested (1300 g, 10 min, RT), and washed 1 x in PBS. Cell pellets were mixed 1:1 with PBS containing protease inhibitor cocktail (Roche) and flash-frozen into small pellets in liquid nitrogen, which were subsequently used for cryo-milling in a planetary ball mill (Retsch) [21] in order to obtain fine sample powder. For co-immunoprecipitation, two smidgen spoons of powder were mixed with 1 ml of extraction buffer (20 mM HEPES, 100 mM NaCl, 0.1% Triton X-100, cOmplete protease inhibitor cocktail (Roche), DNase, pH 7.4) in LoBIND eppendorf tubes. The mixture was sonicated with a microtip sonicator (Misonix Ultrasonic Processor XL), 5 pulses of 3 s each at amplitude 15, at 4 °C. Subsequently, insoluble material was removed by centrifugation (20,000 g, 10 min, 4 °C). 6 μl magnetic anti-GFP nanobody beads (GFP-Trap Magnetic Agarose, Chromotek) were used per sample and washed 2 x with extraction buffer prior to use. Beads were incubated with samples for 2 h rotating at 4 °C, and then washed 3 x with the extraction buffer. With the 3rd wash, six pull-out samples were combined for one final sample. All buffer was removed, and then beads incubated with 40 μl of 4 x LDS buffer (Thermo Scientific) for 10 min at 70 °C. The beads were removed, and eluates loaded on a NuPAGE 4–12% gel (Invitrogen). Following short separation on the gel (until the sample migrated ~1 cm into the gel), the protein containing gel slices were cut out, washed 3 x for 15 min in gel fixative (10% acetic acid, 45% methanol), and then subjected to reductive alkylation and in-gel tryptic digest using routine procedures. Samples from a non-tagged parental strain were used as controls and three replicates were performed for each strain. Liquid chromatography mass spectrometry (LC-MS) was performed at the Proteomics Facility at the University of Dundee, UK. Samples were analysed on a Dionex UltiMate 3000 RSLCnano System coupled to a Q Exactive HF Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Scientific). Data analysis was carried out by mapping peptides to proteins and label-free quantification using Max Quant v1.6.12.0 [22,23]. Differential protein abundance and statistical analysis were performed using Perseus v1.6.12.0 [24] and volcano plots drawn using GraphPad Prism8 software. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [25] partner repository with the dataset identifier PXD032995 (www.ebi.ac.uk/pride).

Funding

This work was supported by an Investigator Award to D.H. (217105/Z/19/Z) and a Centre Award (203134/Z/16/Z) from the Wellcome Trust. M.N. was supported by a four-Year PhD Studentship (222326/Z/21/Z) from the Wellcome Trust. The FingerPrints Proteomics Facility is supported by the 'Wellcome Trust Technology Platform' award (097945/B/11/Z).

CRediT authorship contribution statement

J.K. Conceptualization, Formal analysis, Investigation, Methodology, Writing – review & editing. **J.F.** Conceptualization, Formal analysis, Investigation, Methodology, Writing – review & editing. **M.N.** Conceptualization, Formal analysis, Investigation, Writing – review & editing. **E.R.** Conceptualization, Formal analysis, Investigation, Methodology. **C.W.** Investigation. **M.Z.** Investigation. **M.C.F.** Supervision. **D. H.** Conceptualization, Formal analysis, Funding acquisition, Supervision, Writing – original draft.

Acknowledgements

We thank Haley Licon who, during an internship, first attempted VEX2-precision tagging. We also thank the FingerPrints Proteomics Facility at the University of Dundee.

Conflict of interest statement

None declared.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.molbiopara.2022.111476.

References

[1] E. Ullu, K.R. Matthews, C. Tschudi, Temporal order of RNA-processing reactions in trypanosomes: rapid *trans* splicing precedes polyadenylation of newly synthesized tubulin transcripts, Mol. Cell Biol. 13 (1) (1993) 720–725.
[2] T.N. Siegel, D.R. Hekstra, L.E. Kemp, L.M. Figueiredo, J.E. Lowell, D. Fenyo, X. Wang, S. Dewell, G.A. Cross, Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*, Genes Dev. 23 (9) (2009) 1063–1076.
[3] D.P. Staneva, R. Carloni, T. Auchynnikava, P. Tong, J. Rappsilber, A. A. Jeyaprakash, K.R. Matthews, R.C. Allshire, A systematic analysis of *Trypanosoma brucei* chromatin factors identifies novel protein interaction networks associated with sites of transcription initiation and termination, Genome Res. 31 (11) (2021) 2138–2154.
[4] Q.P. Wang, T. Kawahara, D. Horn, Histone deacetylases play distinct roles in telomeric *VSG* expression site silencing in African trypanosomes, Mol. Microbiol. 77 (5) (2010) 1237–1245.
[5] J. Faria, V. Luzak, L.S.M. Muller, B.G. Brink, S. Hutchinson, L. Glover, D. Horn, T. N. Siegel, Spatial integration of transcription and splicing in a dedicated compartment sustains monogenic antigen expression in African trypanosomes, Nat. Microbiol. 6 (3) (2021) 289–300.
[6] R.J. Wall, E. Rico, I. Lukac, F. Zuccotto, S. Elg, I.H. Gilbert, Y. Freund, M.R.K. Alley, M.C. Field, S. Wyllie, D. Horn, Clinical and veterinary trypanocidal benzoxaboroles target CPSF3, Proc. Natl. Acad. Sci. U.S.A. 115 (38) (2018) 9616–9621.
[7] G.M. Peralta, E. Serra, V.L. Alonso, Update on the biological relevance of lysine acetylation as a novel drug target in trypanosomatids, Curr. Med. Chem. (2021).
[8] L. Glover, D. Horn, Site-specific DNA double-strand breaks greatly increase stable transformation efficiency in *Trypanosoma brucei*, Mol. Biochem. Parasitol. 166 (2) (2009) 194–197.
[9] E. Rico, L. Jeacock, J. Kovarova, D. Horn, Inducible high-efficiency CRISPR-Cas9-targeted gene editing and precision base editing in African trypanosomes, Sci. Rep. 8 (1) (2018) 7960.
[10] J.J. Vasquez, C. Wedel, R.O. Cosentino, T.N. Siegel, Exploiting CRISPR-Cas9 technology to investigate individual histone modifications, Nucleic Acids Res. 46 (18) (2018), e106.
[11] T. Kawahara, T.N. Siegel, A.K. Ingram, S. Alsford, G.A. Cross, D. Horn, Two essential MYST-family proteins display distinct roles in histone H4K10 acetylation and telomeric silencing in trypanosomes, Mol. Microbiol 69 (4) (2008) 1054–1068.
[12] J. Faria, L. Glover, S. Hutchinson, C. Boehm, M.C. Field, D. Horn, Monoallelic expression and epigenetic inheritance sustained by a *Trypanosoma brucei* variant surface glycoprotein exclusion complex, Nat. Commun. 10 (1) (2019) 3023.
[13] D. Schulz, M.R. Mugnier, E.M. Paulsen, H.S. Kim, C.W. Chung, D.F. Tough, I. Rioja, R.K. Prinjha, F.N. Papavasiliou, E.W. Debler, Bromodomain proteins contribute to maintenance of bloodstream form stage identity in the African trypanosome, PLoS Biol. 13 (12) (2015), e1002316.
[14] C. Tiengwe, L. Marcello, H. Farr, N. Dickens, S. Kelly, M. Swiderski, D. Vaughan, K. Gull, J.D. Barry, S.D. Bell, R. McCulloch, Genome-wide analysis reveals extensive functional interaction between DNA replication initiation and transcription in the genome of *Trypanosoma brucei*, Cell Rep. 2 (1) (2012) 185–197.
[15] G.V. Villanova, S.C. Nardelli, P. Cribb, A. Magdaleno, A.M. Silber, M.C. Motta, S. Schenkman, E. Serra, *Trypanosoma cruzi* bromodomain factor 2 (BDF2) binds to acetylated histones and is accumulated after UV irradiation, Int. J. Parasitol. 39 (6) (2009) 665–673.
[16] X. Yang, X. Wu, J. Zhang, X. Zhang, C. Xu, S. Liao, X. Tu, Recognition of hyperacetylated N-terminus of H2AZ by TbBDF2 from *Trypanosoma brucei*, Biochem. J. 474 (22) (2017) 3817–3830.
[17] H. Long, L. Zhang, M. Lv, Z. Wen, W. Zhang, X. Chen, P. Zhang, T. Li, L. Chang, C. Jin, G. Wu, X. Wang, F. Yang, J. Pei, P. Chen, R. Margueron, H. Deng, M. Zhu, G. Li, H2A.Z facilitates licensing and activation of early replication origins, Nature 577 (7791) (2020) 576–581.
[18] T. Beneke, E. Gluenz, LeishGEdit: a method for rapid gene knockout and tagging using CRISPR-Cas9, Methods Mol. Biol. 2019 (1971) 189–210.
[19] S. Dean, J.D. Sunter, R.J. Wheeler, TrypTag.org: a trypanosome genome-wide protein localisation resource, Trends Parasitol. 33 (2) (2017) 80–82.
[20] S. Alsford, D.J. Turner, S.O. Obado, A. Sanchez-Flores, L. Glover, M. Berriman, C. Hertz-Fowler, D. Horn, High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome, Genome Res. 21 (6) (2011) 915–924.
[21] S.O. Obado, M. Brillantes, K. Uryu, W. Zhang, N.E. Ketaren, B.T. Chait, M.C. Field, M.P. Rout, Interactome mapping reveals the evolutionary history of the nuclear pore complex, PLoS Biol. 14 (2) (2016), e1002365.
[22] J. Cox, M.Y. Hein, C.A. Luber, I. Paron, N. Nagaraj, M. Mann, Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ, Mol. Cell Proteom. 13 (9) (2014) 2513–2526.
[23] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, Nat. Biotechnol. 26 (12) (2008) 1367–1372.
[24] S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M.Y. Hein, T. Geiger, M. Mann, J. Cox, The Perseus computational platform for comprehensive analysis of (prote) omics data, Nat. Methods 13 (9) (2016) 731–740.
[25] Y. Perez-Riverol, J. Bai, C. Bandla, D. Garcia-Seisdedos, S. Hewapathirana, S. Kamatchinathan, D.J. Kundu, A. Prakash, A. Frericks-Zipper, M. Eisenacher, M. Walzer, S. Wang, A. Brazma, J.A. Vizcaino, The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences, Nucleic Acids Res 50 (D1) (2022) D543–D552.