Inchausti *et al. BMC Genomics* (2025) 26:916
https://doi.org/10.1186/s12864-025-12095-7

BMC Genomics

RESEARCH Open Access

Exploring a gene co-expression network throughout the *trypanosoma cruzi* life cycle

Lucas Inchausti[1,2], Álvaro Martín[3,4], Leticia Pérez-Díaz[2], Beatriz Garat[2], María Ana Duhagon[2,5], José Sotelo-Silveira[4,6], Javier G. De Gaudenzi[7,8*] and Pablo Smircich[1,2*]

Abstract

Background *Trypanosoma cruzi*, the causative agent of American trypanosomiasis (Chagas disease), is a protozoan parasite with a complex life involving multiple developmental stages in both its triatomine vector and mammalian host. Each stage is characterized by distinct morphological and functional traits. Intriguingly, *T. cruzi* exhibits polycistronic transcription, relying predominantly on post-transcriptional mechanisms for gene regulation.

Methods To delve deeper into the molecular aspects of this regulation, we performed gene co-expression network (GCN) analysis using transcriptomic data spanning all life-cycle stages of *T. cruzi*, offering insights into the coordinated expression patterns of functionally Linked gene groups. We examined the global network properties, identifying overrepresented functional pathways and highly connected hub genes. Additionally, we explored potential regulatory mechanisms within each module, focusing on conserved motifs in the 3' untranslated regions (3'UTRs) of co-expressed genes.

Results Our approach led to the identification of thirteen distinct co-expressed gene modules, each enriched in specific biological processes, including metabolism, pathogenesis, chromatin regulation, cytoskeleton modulation, and cellular movement. Finally, our study highlighted hub genes within each module. Combining a guilt-by-association approach with structural alignments and HMM-HMM profile comparisons, we assigned putative functions to previously uncharacterized proteins. Motif analysis of 3'UTR sequences in co-expressed genes revealed conserved elements and potential regulatory protein factors.

Conclusions GCN analysis is a powerful tool for studying gene expression regulation. Our findings provide new insights into the regulatory networks of *T. cruzi*, identifying key genes and mechanisms underlying coordinated gene expression.

Keywords *Trypanosoma cruzi*, Chagas disease, Gene co-expression network analysis, Post-transcriptional regulation, RNA-seq

*Correspondence:
Javier G. De Gaudenzi
jdegaudenzi@iib.unsam.edu.ar
Pablo Smircich
psmircich@fcien.edu.uy

Full list of author information is available at the end of the article

Author summary

This study explores gene regulatory mechanisms of *Trypanosoma cruzi*, the parasite responsible for Chagas disease. By analyzing gene expression patterns across the parasite's life cycle, we identified distinct groups of co-expressed genes and revealed associated biological processes. We also pinpointed previously unknown hub genes, shedding light on their potential functions. Finally, putative regulatory sequences and their associated proteins were found. These findings provide a deeper understanding of *T. cruzi* biology and offer valuable insights into potential targets for drug development against Chagas disease.

Background

Trypanosoma cruzi is an intracellular protozoan parasite that causes Chagas disease, a neglected tropical disease affecting millions of people worldwide [1, 2]. The life cycle of *T. cruzi* involves sequential stages in its insect vector and mammalian host, each with unique environmental conditions and physiological requirements. During its life cycle, the parasite undergoes several developmental transitions, including the noninfective replicative epimastigote form in the insect vector, the infective non replicative metacyclic trypomastigote form, the intracellular replicative amastigote and the infective non replicative bloodstream trypomastigote forms in the mammalian host [3].

RNA-Seq technology has allowed the genome-wide analysis of gene expression processes in *T. cruzi* and other kinetoplastids for understanding their biology [4–14]. However, most studies have focused on individual stages or specific experimental conditions, limiting our understanding of the global gene expression landscape and the interplay between genes across the entire life cycle. The accumulation of extensive transcriptomic data in recent decades has facilitated more in-depth analyses by constructing gene co-expression networks (GCN) that provide valuable insights into the functional and molecular complexities of biological processes. Prior studies have investigated gene co-expression networks (GCNs) in *Trypanosoma brucei*, shedding light on functionally relevant genes associated with parasite development within the tsetse fly vector [15]. More recently, Martinez-Hernandez et al. (2025) extended GCN analysis to Leishmania spp. by integrating genome-wide non-coding RNA (ncRNA) profiles with coding-gene expression across developmental stages, highlighting novel regulatory modules that may govern stage-specific differentiation and host-parasite interactions [16]. To date, no comparable GCN studies have been performed in T. cruzi.

This type of analysis has emerged as a powerful approach to uncover the modular organization and functional relationships between genes within a biological system. By integrating multiple RNA-seq datasets from different life cycle stages, it is possible to capture the dynamics of gene expression and identify modules of co-regulated genes associated with specific stages or biological processes.

In this study, we conducted a large-scale GCN using publicly available RNA-seq datasets encompassing all four life cycle stages of *T. cruzi*. We identified co-expression modules representing groups of genes that are coordinately regulated and potentially involved in similar biological processes. By integrating network topology analysis, functional enrichment, and hub gene identification, we aimed to elucidate the molecular mechanisms underlying *T. cruzi* biology. These findings provide a global view of the transcriptomic landscape of *T. cruzi* across its life cycle stages, revealing key genes and putative cis and trans acting factors involved in stage-specific processes.

Methods

Data processing and quality control

RNA-seq data from different stages of *Trypanosoma cruzi* were obtained from various published studies: epimastigotes at different in vitro growth phases (7, 14, 21, and 28 days) by Smircich et al. [4], available from the National Center for Biotechnology (NCBI) Sequence Read Archive (SRA) under project PRJNA915394; metacyclic trypomastigotes by Cruz-Saavedra et al. [5], available in the European Nucleotide Archive (ENA) under project PRJEB33521; and intracellular amastigotes at 4, 6, 12, 24, 48, and 72 h postinfection and extracellular trypomastigotes by Li et al. [6], available in the NCBI SRA under projects PRJNA251582 and PRJNA251583, respectively. The number of replicates for each sample are shown in Supplementary Table 1. The quality of each sample was determined via FastQC [17] version 0.11.9.

Estimation of gene expression profiles

For the final quantification of reads assigned to *T. cruzi* genes from all 38 RNA-seq experiments, Kallisto software version 0.46.1 was used [18] with default parameters, performing pseudoalignment against the reference *T. cruzi* genome (CL-Brener Esmeraldo-like version 50, http://tritrypdb.org). After the gene counts were obtained, normalization and rlog transformation of the data were performed via the DESeq2 R package [19]. Briefly, rlog applies a log2 transformation but stabilizes variance across the mean by applying a shrinkage-based transformation, particularly effective for visualizing low-count genes.

Gene co-expression network construction

To construct the GCN, the CEMiTool R package was used [20]. CEMiTool implements and automates the

functionalities of the WGCNA R package [21]. Using this tool, we built an unsigned co-expression network, meaning that both positive and negative correlations between genes contribute equally to their connectivity.

The algorithm first computes a gene-by-gene Pearson correlation matrix from the gene expression data. It takes the absolute value of these correlations and transforms them into a similarity matrix, raising them to a power β to enhance strong connections while discarding weak ones. The optimal β is selected based on the scale-free topology criterion [22]. Next, a topological overlap matrix (TOM) is constructed, considering both direct and shared connections between genes.

Finally, the TOM matrix is transformed into a dissimilarity matrix $\mathrm{dissTOM} = 1 - \mathrm{TOM}$, and hierarchical clustering via average linkage is performed with the hclust function to create a dendrogram. The dynamic tree cut algorithm from the cutreeDynamicTree function of the dynamicTreeCut R package [23] is then used to group genes into co-expressed modules or clusters. For each module, a.txt file with the gene identifiers is created. For graph visualization igraph [24] and ggraph [25] R packages were used.

To better visualize coordinated expression patterns, each module was then partitioned according to the sign of the correlation between individual genes and the corresponding module eigengene. This approach allowed us to separate genes that increase or decrease together across the *T. cruzi* life cycle and to highlight shared expression trends within modules. Plots were generated using the ggplot2 R package.

Functional enrichment analysis of co-expressed gene modules

For the functional enrichment analysis, overrepresentation of Gene Ontology (GO) terms from the "biological process" ontology among genes from each module was assessed using the Gene Ontology Enrichment tool available in TriTrypDB [26]. Modules that exhibited overrepresented GO terms with $\mathrm{pvalue} < 0.05$ and a $\mathrm{Benjamini} - \mathrm{Hochberg\ adjusted\ p} - \mathrm{value} < 0.1$ were considered significant and retained for further analysis. To visualize the global expression dynamics of the co-expression modules across the *T. cruzi* life cycle, we calculated the module eigengenes using the moduleEigengenes function from the WGCNA R package [21]. Module eigengenes represent the first principal component of the expression values of all genes within each module and serve as a summary of their expression pattern. Module-stage relationships were determined by calculating Pearson correlation coefficients between modules and developmental stages traits (epimastigote, metacyclic trypomastigote, amastigote or cellular trypomastigote stages).

Functional inference of genes with unknown function

First, the 5 most connected genes (hub genes) were identified for each module showing overrepresentation of GO terms using the get_hubs function from the CEMiTool R package. The functional annotation of those genes was retrieved from TriTrypDB, and selected genes annotated as hypothetical proteins were further analyzed to determine their function employing two methodologies: HMM-HMM profile comparison and structural alignment. For the first methodology, DARK database (available at https://github.com/sradiouy/DARK) was used. This software allows visualization and interrogation of protein annotations produced by HMM-HMM comparison strategies.

For the second methodology, the predicted 3D structure for each hypothetical protein encoded by the previously identified hub genes was obtained from the AlphaFold web server [27]. The corresponding PDB file containing the 3D structure information for each protein was subsequently downloaded and used as input in the FoldSeek web server [28]. FoldSeek enables homology searches through structural alignment in numerous databases using TM-align and 3Di/AA algorithms. TM-scores > 0.4 and E-values < 0.05 were the criteria employed to accept a protein as a homologous.

Obtaining 3'UTRs

The genome and annotation files of the CL Brener Esmeraldo-Like strain of *T. cruzi* were downloaded from TriTrypDB (v. 50), and together with the previously described sequencing reads, UTRme software [29] was used to predict the 3'UTRs of each gene, retaining the longest reported 3'UTR for each gene.

Search for motifs in 3'UTRs

To identify sequence motifs in the 3'UTRs of genes in each module, the fasta files generated with UTRme were used as inputs for XSTREME software [30] from MEME Suite [31]. A background sequence file is used for comparison to determine motif overrepresentation. For this purpose, a fasta file was created for each module's input file containing the remaining 3'UTRs of genes that did not belong to the evaluated module.

Once the motifs were obtained for each module, the FIMO software [32] from the MEME Suite was utilized to determine the motif proportions in both the modules and the backgrounds. Motifs that had at least twice the representation in the module compared with the background were retained, i.e., $\frac{proportioninthemodule}{proportioninthebackground} \geq 2$. Furthermore, this program reports RNA-binding proteins (RBPs) previously identified by Ray et al. (2013) that bind to the motif.

Orthologous RBP identification in *T. cruzi*

To identify the orthologous proteins from *T. cruzi* corresponding to the reported RBPs obtained via MEME Suite, BLASTP [33] was employed, and stringent criteria were applied, considering hits with E-values lower than 5E-04, a subject coverage greater than 0.5, and selecting the hit with higher sequence identity.

3'UTR motif and RBP interaction analysis

To characterize the identified motifs and associated RBPs in detail, an analysis using RPIseq [34] was performed to assess the interaction between the identified motifs and RBPs. RPIseq takes the motif and RBP sequences and determines the likelihood of interaction between the RNA-RBP pair via a random forest (RF) and a support vector machine (SVM) classifier. The criterion for identifying potential interactions was satisfied when both the RF and SVM classifiers obtained interaction scores greater than 0.5. Tertiary structures of both RNA motifs and RBPs were obtained using RNAComposer [35] and AlphaFold [27], respectively. Finally, docking simulations were performed using HDOCK [36] and HADDOCK [37] to further evaluate the likelihood of the predicted RNA–RBP interactions.

Results

Data acquisition and processing

A total of 38 RNA-seq samples were used to conduct this study, encompassing the four main life cycle stages of *T. cruzi*: epimastigotes, metacyclic trypomastigotes, amastigotes, and cell derived trypomastigotes, which were obtained from three independent studies (Supplementary Table 1).

For each sample, a quality analysis of the reads was performed using FastQC software, which revealed that all the samples presented good sequence quality, as determined by their phred scores > 28, and did not show any sequence overrepresentation. Therefore, the study was conducted with all 38 transcriptomes.

Network construction and characterization

To estimate gene expression profiles from all 38 RNA-seq samples, Kallisto was employed, which returned a table with pseudocounts assigned to each of the 10,338 genes of *T. cruzi*. Normalization and rlog transformation were performed via the DESeq2 R package [19].

Once gene expression levels were estimated, the identification of co-expressed gene clusters was performed. First, a gene expression correlation adjacency matrix among different samples was built. Before the network was constructed, the soft-thresholding power (β) was determined, which was used to adjust the network to a scale-free topology by eliminating noise produced by weak correlations in the adjacency matrix. This type of topology prevails in biological networks [38, 39] and is characterized by an uneven distribution of connections between nodes (genes) in the network: a few genes (called hub genes) are expected to have many connections, whereas many genes have few connections. To achieve this, different values of $\beta(1-20)$ were evaluated. For each value, a linear model was fitted between the logarithm of connectivity, k, and the logarithm of the node frequency with connectivity k, $p(k)$. The goodness of fit of this model was assessed using the scale-free topology index, R^2, which quantifies how well the data adhere to a power-law distribution. Values of R^2 closer to 1 indicate a better fit to the scale-free topology. Finally, the smallest β that obtained an $R^2 > 0.9$ was chosen $(\beta = 7)$ to obtain a network that is as connected as possible in terms of the mean connectivity of network nodes (Fig. 1B) while maintaining a scale-free topology (Fig. 1A). The linear fit between $log10(k)$ and $log10(p(k))$ is illustrated in Fig. 1C.

Once the network was constructed, 14 modules of genes were identified through hierarchical clustering via the dynamic tree cut algorithm. Among the 14 modules, one contained 10 genes that could not be assigned to any of the other 13 modules because no association in their co-expression levels was detected, so it was excluded from further analyses. Among the 13 retained modules, module M1 had the most genes (2646), whereas module M13 had the fewest genes [74] (Fig. 2).

Functional characterization of modules

Once the GCN was constructed, a functional enrichment analysis was conducted to assess the overrepresentation of GO terms from the "biological process" ontology. Among the 13 evaluated modules, 10 were enriched in biological processes associated with metabolism, pathogenesis, chromatin regulation, cytoskeleton regulation, and cellular movement, among others (Table 1).

It is interesting to observe that several GO terms specific to developmental changes are overrepresented in the modules, suggesting that they successfully capture gene expression dynamics underlying the major biological transitions throughout the *T. cruzi* life cycle [6, 12, 40].

To further explore this relationship between co-expression modules and *T. cruzi* life cycle stages, we examined the module-eigengene expression profiles of each module across the different developmental stages (Fig. 3). Also, genes in each module were partitioned into two subsets according to the sign of their correlation with the module eigengene, and their expression trajectories were plotted separately (Supplementary Fig. 1).

This analysis revealed clear stage-specific patterns. For example, modules M4 and M13 -two highly connected groups in the network (Fig. 2B)- showed the highest expression in trypomastigotes, modules mainly

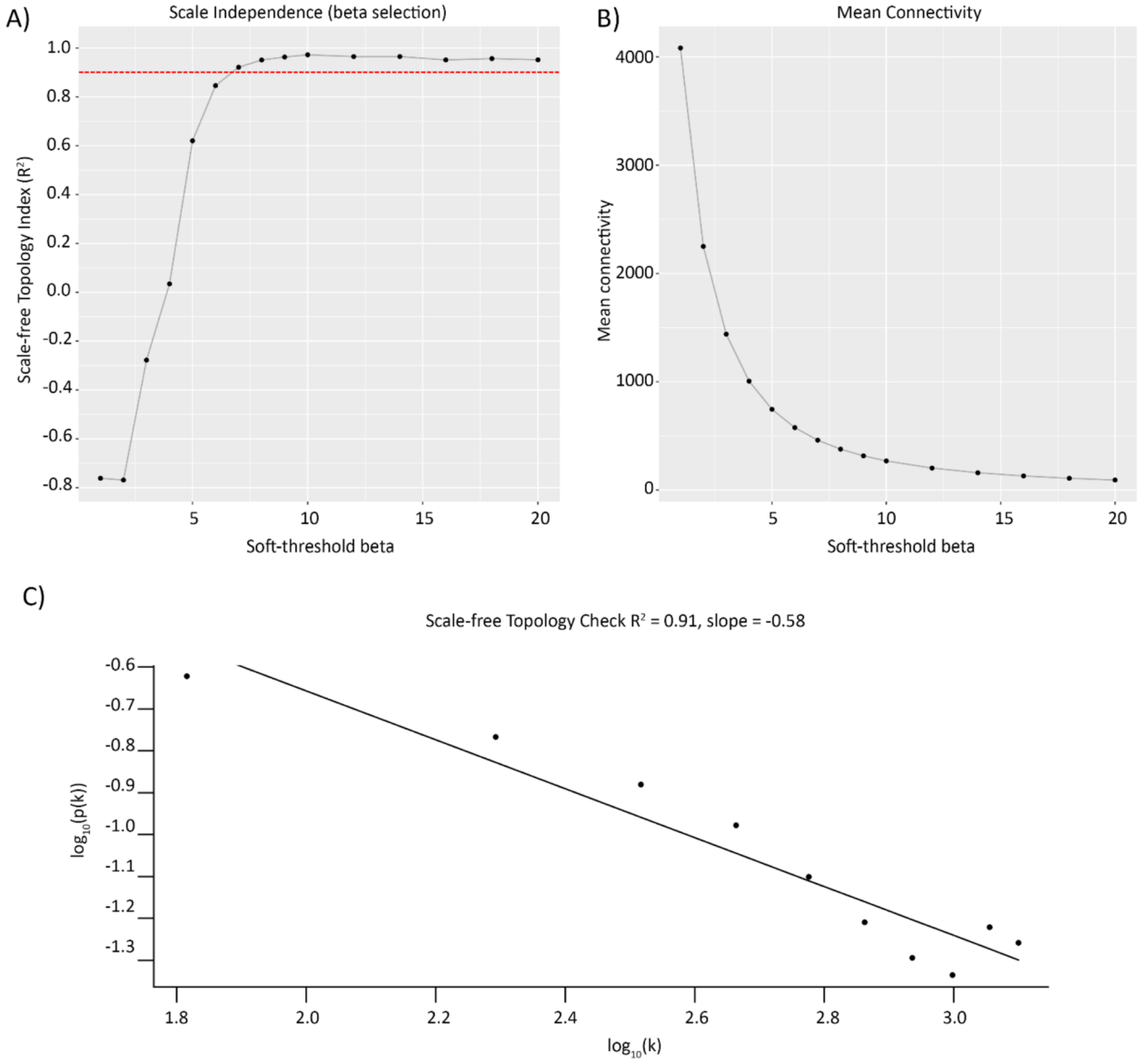

Fig. 1 Selection of the parameter β for the construction of the GCN with a scale-free topology. The lowest β was selected when the scale-free topology index (R^2) was > 0.9. **A** Fit index of the network to a scale-free topology (R^2) as a function of the different evaluated β values ($R^2 = 0.9$ marked in red), **B** average connectivity of the network as a function of the different evaluated β values, and **C** linear fit between the logarithm of the frequency of nodes with connectivity k and the logarithm of connectivity k for the network obtained with β = 7

composed of surface protein coding genes, showing their involvement in processes active during this stage, such as infection (see Discussion for a more comprehensive analysis of the modules) [40].

Identification and functional analysis of the hub genes

Hub genes are those that exhibit high connectivity with others in GCN. Owing to this high connectivity, there is evidence that hub genes are important for the organization and functioning of the biological network and may play a crucial role in the biological processes underlying each module [21, 39].

In this context, genes from each module were sorted based on their connectivity (Supplementary Table 2). The top five most connected genes (top 5 hub genes) were identified for the 10 modules that exhibited overrepresented GO terms, and their annotations from TriTrypDB were obtained (Table 2 and Supplementary Fig. 2). Thus, a total of 50 hub genes were analyzed, 16 of which were annotated as hypothetical proteins. Of these, 3 were annotated as conserved hypothetical protein pseudogenes, whereas 9 were annotated as conserved hypothetical proteins.

In the case of module M1, the overrepresented GO terms are located at high levels within the hierarchical

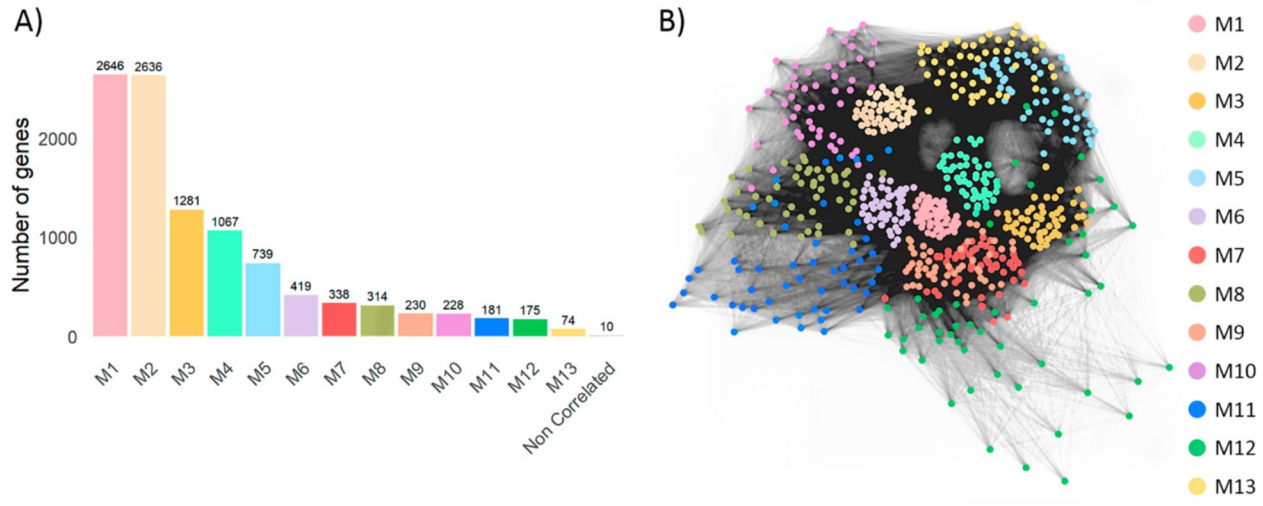

Fig. 2 **A** Number of genes identified in each of the 14 modules obtained. The "Non-Correlated" module contained 10 genes that could not be assigned to any of the other modules, so it was excluded from further analyses. **B** Co-expression graph of the top 50 most connected genes from each module. In the graph each node represents a gene, and each Line connects two genes if there is a 1-dissTOM value > 0.05

Table 1 Modules with overrepresented Gene Ontology (GO) terms and their most significantly overrepresented GO terms with *p* values < 0.05 and Benjamini–Hochberg adjusted *p* values < 0.1

Module	Overrepresented GO terms	Benjamini–Hochberg adjusted *p* value (< 0.1)
M1	metabolic process, biological process, organic substance metabolic process	2.10E-3, 2.10E-3, 2.26E-3
M4	obsolete pathogenesis	5.77E-14
M5	obsolete pathogenesis	2.14E-41
M7	oxoacid metabolic process, pyruvate metabolic process, ATP metabolic process, ribonucleotide metabolic process	2.86E-4, 2.86E-4, 1.95E-3, 2.32E-3
M8	chromosome organization, DNA packaging, DNA conformation change, nucleosome assembly, cellular component organization	6.30E-5, 8.88E-4, 2.65E-3, 1.64E-2
M9	regulation of supramolecular fiber organization, regulation of cytoskeleton organization, cilium or flagellum dependent cell motility	1.24E-2, 1.24E-2, 1.24E-2, 1.24E-2
M10	regulation of mRNA stability, regulation of mRNA catabolic process, regulation of mRNA metabolic process	7.07E-2, 7.07E-2, 7.07E-2
M11	cell–matrix adhesion, cell-substrate adhesion	1.92E-2, 1.92E-2
M12	movement of cell or subcellular component, microtubule-based process	2.83E-2, 2.83E-2
M13	obsolete pathogenesis	9.38E-02

structure of the GO DAG, rendering it challenging to establish a close association between the annotated hub genes and the biological processes associated with the module. Interestingly, four of its five hub genes were annotated as conserved pseudogenes. In this regard, recent studies have shown that 94% of *T. cruzi* pseudogenes are transcriptionally active [41] and that pseudogenes could have developed new functions as regulators of gene expression, serve as substrates for the evolution of new genes, or even be reactivated to perform specific biological functions [42]. However, further investigation into pseudogenes in *T. cruzi* is still necessary to determine their potential biological functions.

For modules M4, M5, and M13, which had overrepresented GO terms associated with pathogenesis, all annotated hub genes were found to encode surface proteins of the trans-sialidase type, mucin, or MASP, which are known to be significant virulence factors of this parasite [12]. This pattern is also reflected in the network visualization, where these modules appear as closely connected clusters, indicating a potential co-regulation or functional interplay (see Fig. 2B and Supplementary Fig. 2).

Upon examining the functionalities of the hub genes within module M7, which are linked to various metabolic processes, such as the oxoacid metabolic process, pyruvate metabolic process, ATP metabolic process, and ribonucleotide metabolic process, it was revealed that two of the hub genes had functions related to these processes. TcCLB.508387.20 codes for an enzyme responsible for polyamine metabolism, whereas TcCLB.506367.30 encodes an oxidoreductase family enzyme, participating in methionine metabolism.

Module M8, which is involved in biological processes associated with nuclear components and chromatin regulation, harbored hub genes with functional annotations, including TcCLB.511417.70, a histone methyltransferase, a key enzyme involved in the epigenetic modification of chromatin, thereby regulating its compaction level. Additionally, TcCLB.511871.30 encodes a

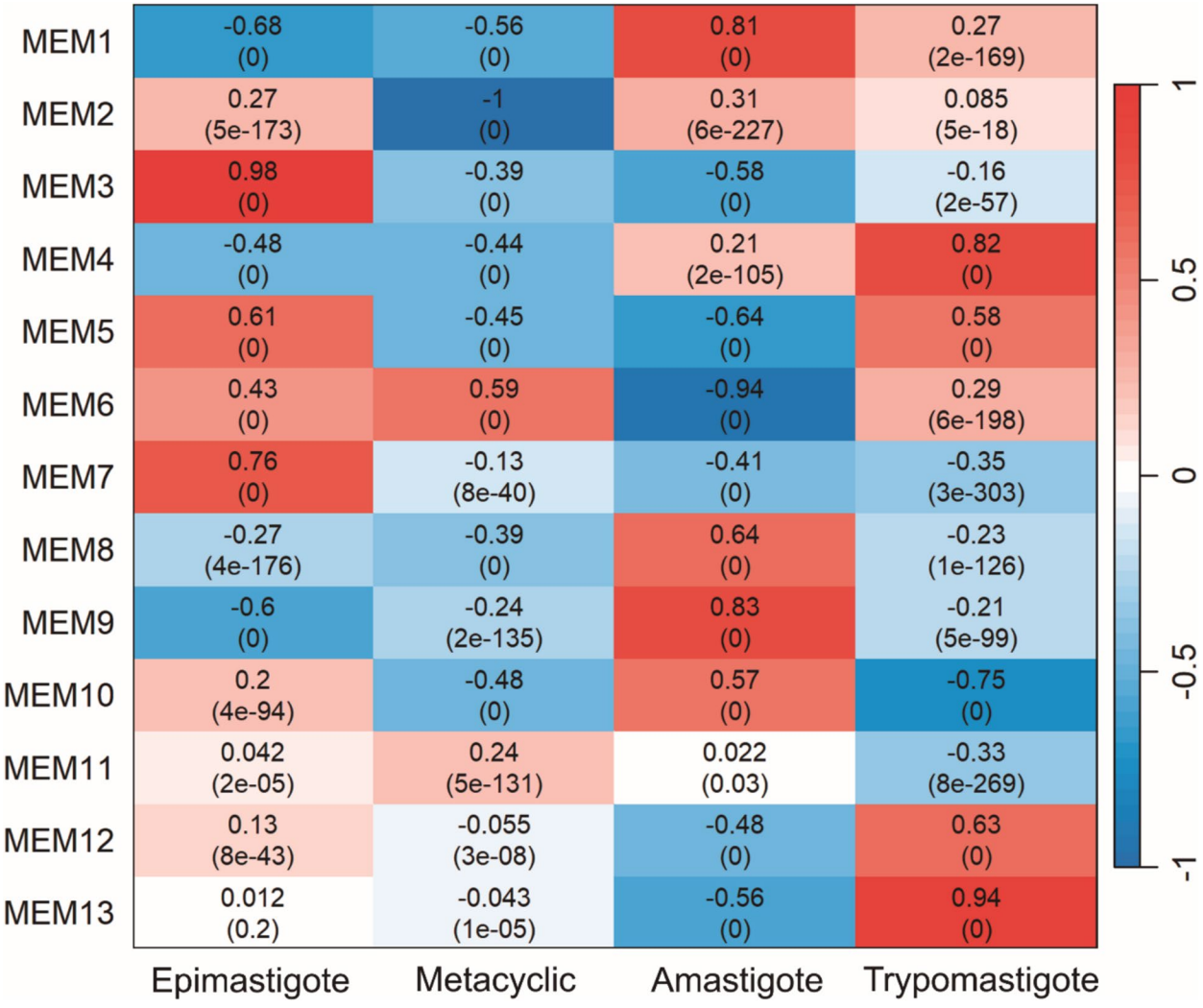

Fig. 3 Trait correlation analysis.Heatmap showing the correlation between module-eigengenes and life stage of the 13 co-expression modules identified. Rows correspond to the 13 identified modules, and columns represent the different life cycle stage samples. The upper number in each cell indicates the correlation value and the number in parentheses indicates the associated *p*-value

protein from the 2OG-Fe(II) oxygenase superfamily that is known to be involved in DNA repair, regulation, and modification across various organisms [43]. Moreover, TcCLB.506795.44, a bifunctional NAD(P)H-hydrate repair enzyme, plays a role in DNA repair and is part of the nicotinamide nucleotide proofreading system [44].

In the case of module M9, associated with the regulation of cytoskeleton organization and flagellar mobility, only two of its top 5 hub genes were found to have assigned functions: TcCLB.508955.10 and TcCLB.511215.119. The former gene encodes a syntaxin, which is involved in exocytosis and endocytosis processes [45], whereas the latter gene encodes a Rod2 flagellar protein, which is involved in flagellar mobility [46].

Regarding module M10, which is linked to mRNA regulation processes, the hub genes TcCLB.508169.90 and TcCLB.509455.100 were found to encode subunit 1 of the eIF3 translation initiation factor and an exonuclease, respectively, the latter of which catalyzes RNA degradation from both the 5' and 3' ends.

Finally, concerning modules M11 and M12, as well as some of the hub genes from the other modules, no apparent associations were discerned between their functions and the overrepresented GO terms.

From Table 2, a reduced list of hub genes with the "hypothetical protein" description was obtained, and an attempt was made to approach their function in silico. To this end, we applied a guilt-by-association approach by analyzing subnetworks centered on each hypothetical protein hub gene. Each subnetwork consisted of the nine most strongly connected neighboring genes. The functions of these neighboring genes were examined, allowing us to infer preliminary functional annotations for the hypothetical hubs. As expected, the inferred

Table 2 Top 5 hub genes from the 10 modules with overrepresentation of Gene Ontology (GO) terms, along with their functional annotation obtained from TriTrypDB, and the overrepresented GO terms of the module

Module	Top 5 hub genes	Annotation	Overrepresented GO terms
M1	TcCLB.509205.100	surface protease GP63, putative	metabolic process, biological process, organic substance metabolic process
	TcCLB.508207.10	glycine dehydrogenase (pseudogene), putative	
	TcCLB.510849.30	hypothetical protein, conserved (pseudogene)	
	TcCLB.508835.10	hypothetical protein, conserved (pseudogene)	
	TcCLB.506113.80	hypothetical protein, conserved (pseudogene)	
M4	TcCLB.510553.30	Mucin-associated surface protein (MASP), subgroup S101	obsolete pathogenesis
	TcCLB.455171.9	trans-sialidase, putative	
	TcCLB.507905.39	Mucin-associated surface protein (MASP)	
	TcCLB.509081.20	Mucin-associated surface protein (MASP), subgroup S122	
	TcCLB.511553.30	Mucin-associated surface protein (MASP), subgroup S122	
M5	TcCLB.510275.272	hypothetical protein	obsolete pathogenesis
	TcCLB.511173.470	trans-sialidase, Group V, putative	
	TcCLB.509233.10	hypothetical protein	
	TcCLB.506599.420	Mucin-associated surface protein (MASP), subgroup S117	
	TcCLB.508541.90	Mucin-associated surface protein (MASP), subgroup S078	
M7	TcCLB.508387.20	methylthioadenosine phosphorylase, putative	oxoacid metabolic process, pyruvate metabolic process, ATP metabolic process, ribonucleotide metabolic process
	TcCLB.508771.50	hypothetical protein, conserved	
	TcCLB.503487.70	Inhibitor of apoptosis-promoting Bax1, putative	
	TcCLB.506367.30	1,2-Dihydroxy-3-keto-5-methylthiopentene dioxygenase, putative	
	TcCLB.509157.220	hypothetical protein, conserved	
M8	TcCLB.506563.10	pumilio/PUF RNA binding protein 9, putative	chromosome organization, DNA packaging, DNA conformation change, nucleosome assembly, cellular component organization
	TcCLB.511417.70	Histone-lysine N-methyltransferase, H3 lysine-76 specific	
	TcCLB.511871.30	2OG-Fe(II) oxygenase superfamily, putative	
	TcCLB.506795.44	Bifunctional NAD(P)H-hydrate repair enzyme	
	TcCLB.507129.30	C-14 sterol reductase, putative	
M9	TcCLB.508731.40	hypothetical protein, conserved	regulation of supramolecular fiber organization, regulation of cytoskeleton organization, cilium or flagellum dependent cell motility
	TcCLB.510347.29	hypothetical protein, conserved	
	TcCLB.508955.10	syntaxin, putative	
	TcCLB.511215.119	Paraflagellar rod protein 2	
	TcCLB.506927.20	hypothetical protein, conserved	
M10	TcCLB.503811.45	hypothetical protein, conserved	regulation of mRNA stability, regulation of mRNA catabolic process, regulation of mRNA metabolic process
	TcCLB.508169.90	eukaryotic translation initiation factor 3 subunit I	
	TcCLB.506661.40	zinc finger domain, LSD1 subclass, putative	
	TcCLB.509455.100	exonuclease, putative	
	TcCLB.506285.40	Mucin-associated surface protein (MASP) (pseudogene)	
M11	TcCLB.506287.209	DNA ligase, putative	cell–matrix adhesion, cell-substrate adhesion
	TcCLB.511623.20	hypothetical protein, conserved	
	TcCLB.511729.60	MORN repeat-containing protein 1	
	TcCLB.506735.10	mitochondrial processing peptidase alpha subunit, putative	
	TcCLB.510759.120	rieske iron-sulfur protein, mitochondrial precursor, putative	

Table 2 (continued)

Module	Top 5 hub genes	Annotation	Overrepresented GO terms
M12	TcCLB.511127.90	Domain of unknown function (DUF4586), putative	movement of cell or subcellular component, microtubule-based process
	TcCLB.503903.70	hypothetical protein, conserved	
	TcCLB.453917.9	hypothetical protein, conserved	
	TcCLB.510285.20	PQQ-like domain/WD domain, G-beta repeat/Utp21 specific WD40 associated putative domain containing protein, putative	
	TcCLB.511693.70	hypothetical protein, conserved	
M13	TcCLB.506683.110	trans-sialidase, Group VIII, putative	obsolete pathogenesis
	TcCLB.508165.170	Mucin-associated surface protein (MASP), subgroup S030	
	TcCLB.510025.30	Mucin-associated surface protein (MASP) (pseudogene)	
	TcCLB.508165.400	mucin TcMUCII, putative	
	TcCLB.510105.310	Mucin-associated surface protein (MASP), subgroup S085	

functions are consistent with the GO terms associated with the modules to which these hub genes belong. These results are presented in Supplementary File 1. To further explore the potential roles of this set of putatively relevant hypothetical proteins, we employed two additional approaches. First, through a comparison of HMM-HMM profiles, second, through a homology search via structural alignment. Finally, we obtained all related information in TriTrypDB for each of these genes (user comments, orthology matches with other trypanosomatids, or assigned GO terms).

For the first methodology, DARK software was used (https://github.com/sradiouy/DARK). This software enables visualization and interrogation of protein annotations produced by HMM-HMM comparison strategies [47]. For homology searches through structural alignment, the three-dimensional structure of each of the proteins encoded by the hub genes was obtained from the AlphaFold database [27]. The FoldSeek web server was subsequently used to perform homology searches through structural alignment of proteins via two different algorithms: TM-align and 3Di/AA [28]. The results obtained are summarized in Table 3 and detailed in Supplementary Table 3.

The functions of 10 out of 16 unannotated hub genes were inferred using at least one of the two methods. When contrasted to the guilt-by-association results, a positive association was observed: TcCLB.510275.272 and TcCLB.509233.10, belonging to module M5 enriched in the GO term "obsolete pathogenesis," and connected to MASPs and trans-sialidases in the subnetworks (Supplementary File 1), were annotated by DARK as a trans-sialidase and a MASP, respectively.

The gene TcCLB.509157.220 from module M7, associated with various metabolic processes, was annotated as a transmembrane protein and a fatty acid exporter. Interestingly, TriTrypDB shows a homolog with a Diacylglycerol O-acyltransferase from *T. cruzi* Berenice strain which is consistent with our results.

Additionally, from module M9, which is associated with cytoskeleton organization and flagellar mobility, TcCLB.508731.40, was annotated both by DARK and FoldSeek as a mitochondrial protein; in this case, the relationship between the predicted function and the module is not as evident, although the mitochondrion is a relevant organelle for cell flagellar mobility. This gene shows sequence homology to “mithocondrial glycoproteins” from *Leishmania spp*. and *T. brucei*. Furthermore, TcCLB.510347.29 was annotated by FoldSeek as a "HORMA domain-containing protein"; recent studies have discovered new roles in regulating centriole dynamics [48], and TcCLB.506927.20 was annotated by both algorithms as an actin-interacting protein. Interestingly, all three proteins are associated in the subnetwork to cytoskeleton related proteins (Supplementary File 1).

With respect to module M11, which is associated with cellular adhesion processes, TcCLB.511623.20 was annotated by FoldSeek as a "pleckstrin homology-like domain, family B, member 2" (PHLDB2), although it had a TM-score < 0.4. It has been reported that this protein acts as a microtubule-anchoring factor that binds to the CLASP protein and is involved in the interaction between the distal ends of microtubules and the cell cortex [49]. This protein is connected in the network to MORN1 (Supplementary File 1), a protein associated with the spindle poles [50], supporting its putative function.

Finally, concerning module M12 associated with biological processes linked to cellular movement, TcCLB.511127.90 was annotated by DARK as a flagellum-associated protein, whereas TcCLB.503903.70 and TcCLB.511693.70 were annotated by both algorithms as a MORN1 protein and a protein associated with actin and the flagellum, respectively. A study by Morriswood et al. [51] reported that MORN-1 acts as a multiprotein complex that could facilitate the entry of proteins into the flagellar pocket. Consistent with this analysis, this gene is in the same TriTrypDB ortholog group as MORN proteins from Leishmania spp., *T. brucei* and other

Table 3 Hub genes with unknown functions and their predicted functions via DARK and FoldSeek. A hyphen indicates when the strategy yielded no results. It is indicated whether the predicted function clearly matches the overrepresented GO terms from each module

Module	Hub genes with unknown function	Annotation	DARK annotation	FoldSeek annotation	Consistent with over-represented GO terms?
M1	TcCLB.510849.30	hypothetical protein, conserved (pseudogene)	-	-	-
	TcCLB.508835.10	hypothetical protein, conserved (pseudogene)	-	-	-
	TcCLB.506113.80	hypothetical protein, conserved (pseudogene)	-	-	-
M5	TcCLB.510275.272	hypothetical protein	Trans-sialidase	-	Yes
	TcCLB.509233.10	hypothetical protein	Mucin-associated surface protein (MASP), putative	-	Yes
M7	TcCLB.508771.50	hypothetical protein, conserved	-	-	-
	TcCLB.509157.220	hypothetical protein, conserved	Putative transmembrane protein	Protein Fatty Acid Export 7	Yes
M9	TcCLB.508731.40	hypothetical protein, conserved	Mitochondrial glycoprotein-like protein	Head domain of the mt-SSU assemblosome from *Trypanosoma brucei*	Yes
	TcCLB.510347.29	hypothetical protein, conserved	HP_Q4D059	HORMA domain containing protein	Yes
	TcCLB.506927.20	hypothetical protein, conserved	Actin interacting protein 1	Actin-interacting protein 1	Yes
M10	TcCLB.503811.45	hypothetical protein, conserved	-	-	-
M11	TcCLB.511623.20	hypothetical protein, conserved	-	Pleckstrin homology- Like domain, family B, member 2	Yes
M12	TcCLB.511127.90	Domain of unknown function (DUF4586), putative	Flagellar associated protein	-	Yes
	TcCLB.503903.70	hypothetical protein, conserved	Morn repeat protein	Crystal structure of *Trypanosoma brucei* Morn 1	Yes
	TcCLB.453917.9	hypothetical protein, conserved	-	-	-
	TcCLB.511693.70	hypothetical protein, conserved	Actin interacting protein 1	Cilia- and flagella- associated protein 52	Yes

trypanosomatids. Interestingly, as observed for M9, all three proteins in the subnetwork have associations with cytoskeleton-related proteins (Supplementary File 1).

Identification of common 3'UTR motifs within modules

In recent years, genomic and bioinformatic tools have enabled the identification of most of the RBPs of trypanosomatids and numerous target sequence elements, which are involved mainly in RNA processing [52–55]. In this context, the characterization of sequence motifs in the 3'UTRs that putatively regulate the joint expression of genes belonging to each module was proposed, using the XSTREME and FIMO programs of MEME-Suite [31]. We analyzed full modules since identical or highly similar cis-elements may be present in both mirror subclusters but yield different regulatory outcomes depending on context (e.g., cofactors, RBP modification, UTR position, or developmental stage [56–59].

To do this, first, the 3'UTRs of the genes in each module were determined using UTRme software [29]. Once the 3'UTRs were defined, the identification of de novo motifs for each set of 3'UTRs of each module was performed using XSTREME. This software reports the number of sites where it finds the motifs in the module's genes, as well as a statistic that allows inferring the overrepresentation compared with the background, but not the total number of genes in which the motif is found. Therefore, the FIMO software was used, taking the motifs obtained by XSTREME as input, to obtain the proportions of these motifs both in the modules and in the backgrounds, thus being able to establish how specific the motifs were to each of the modules. Furthermore, this program detects motif-binding proteins previously identified by Ray et al. (2013) (Supplementary Table 4). Table 4 summarizes the results obtained in terms of the number of motifs identified in each module and the number of reported RBPs.

The results revealed a series of significant motifs, many of which have been previously reported in other organisms as functional and even with described binding proteins. Furthermore, several of these sequences have high specificity and are very rare outside the genes of the

Table 4 Summary of identified motifs and previously reported RBPs that interact with those motifs (Ray et al., 2013) for each module

Module	# RNA motifs	# Reported RBPs (Ray et al., 2013)
M1	5	2
M4	42	24
M5	40	23
M7	1	0
M9	2	0
M10	1	1
M11	3	3
M13	36	24

studied module. Therefore, they could play a key role as modulators of transcript abundance.

To characterize the motifs and associated RBPs obtained in detail in Table 5, BLASTP [33] was employed to identify the orthologous proteins from *T. cruzi* corresponding to the reported RBPs. Subsequently, RPISeq [34] was employed to determine whether the RBP-motif pairs interact using only sequence information, employing random forest (RF) and support vector machine (SVM) classifiers. Next, using RNAComposer server [35], the tertiary structure of the motifs that obtained an interaction score greater than 0.5 in both the RF and SVM classifiers was predicted.

Among the 77 RBP-motif pairs, 32 orthologous proteins (with the possibility of a protein having more than one associated motif) were identified, retaining 65 RBP-motif pairs. Upon running RPIseq, 50 out of 65 pairs met the criterion where at least one of the two classifiers obtained an interaction score (in terms of probabilities) greater than 0.5; 24 pairs fulfilled the criterion where the RF classifier obtained an interaction score greater than 0.5; 15 pairs fulfilled the criterion where the SVM classifier obtained an interaction score greater than 0.5; and 11 pairs fulfilled the criterion where both classifiers obtained an interaction score greater than 0.5 (Supplementary Table 5 and Supplementary Table 6). Table 5 illustrates the RBP-motif pairs where both classifiers achieved interaction scores greater than 0.5. All these pairs belonged to modules M4 and M13, which were enriched in genes that code for surface proteins.

Although these RBPs have not previously been reported to bind surface-protein mRNAs, we validated their interactions by performing docking simulations for all eleven RNA–protein pairs using HDOCK [36], employing the structures of the RBPs obtained from the AlphaFold database [27] and the corresponding RNA motif (Fig. 4 and Table 6). After the docking simulations, 9 out of the 11 predicted interactions had confidence scores greater than 0.9, whereas the confidence scores from the other two RBP-RNA interactions were greater than 0.8, suggesting that all these molecular associations occur.

To reinforce our findings through an independent methodology, we conducted docking of the eleven candidate RNA–protein interactions using HADDOCK web server [37], guided by β-sheet or Pumilio-domain restraints. Our HADDOCK docking of the previous interactions yielded three highly stable complexes (Fig. 4A, D, K) with favorable scoring metrics and well-defined interfaces. Six additional complexes (Fig. 4B, E, F, G, H, J) displayed moderate stability, characterized by consistent hydrogen-bond networks and acceptable HADDOCK scores (Supplementary Table S7). The two remaining models (Fig. 4C and 4I) were less stable. Overall, nine out of eleven RPIseq/HDOCK predictions are corroborated as viable by HADDOCK, underscoring the robustness of our computational pipeline. Detailed scoring and interface analyses for all models are provided in Supplementary Table S7.

Additionally, we searched for several published RNA elements across the expression modules, providing functional support for at least three of them: the SGPm sequence, previously associated with *T. cruzi* surface protein-coding genes [60], was enriched in M4-B, M5-B, and

Table 5 RBP-motif pairs where both classifiers achieved an interaction score greater than 0.5 in both the RF and SVM classifiers from RPIseq. The RNCMPT id from (Ray et al., 2013), the identified *T. cruzi* ortholog TriTrypDB id and protein name are shown, together with the corresponding module, the motif sequence and the RF and SVM interaction scores

RNCMPT id	*T. cruzi* id	Protein name	Module	Motif sequence	RF score	SVM score
RNCMPT00043	TcCLB.508461.140	TcPABP2	M13	MAAAAAAAAAAAAAA	0.75	0.96
RNCMPT00061	TcCLB.506625.70	TcRBSR1	M13	GAGGGGCCGCGACGA	0.55	0.57
RNCMPT00123	TcCLB.510689.60	TcDRBD7B	M13	UCCUGCCUGGC	0.55	0.66
RNCMPT00133	TcCLB.506399.40	TcDRBD13A	M13	GGUUGAGWGCACCCA	0.65	0.84
RNCMPT00144	TcCLB.508981.20	TcRBP8B	M13	UUGCGGYCAGGGCGG	0.6	0.53
RNCMPT00046	TcCLB.510125.10	PUF6	M4	AUUGUAUAUA	0.65	0.87
RNCMPT00061	TcCLB.506625.70	TcRBSR1	M4	GAGGGGCCGCGACGA	0.55	0.57
RNCMPT00061	TcCLB.506625.70	TcRBSR1	M4	UGCGACGAGG	0.55	0.63
RNCMPT00090	TcCLB.508461.140	TcPABP2	M4	GAGAGAGCC	0.65	0.92
RNCMPT00229	TcCLB.507025.60	TcDRBD5C	M4	UGUUUUGCUUYAYKK	0.7	0.88
RNCMPT00283	TcCLB.506825.10	TcDRBD12B	M4	AGWGUGUGUGCGCUG	0.85	0.9

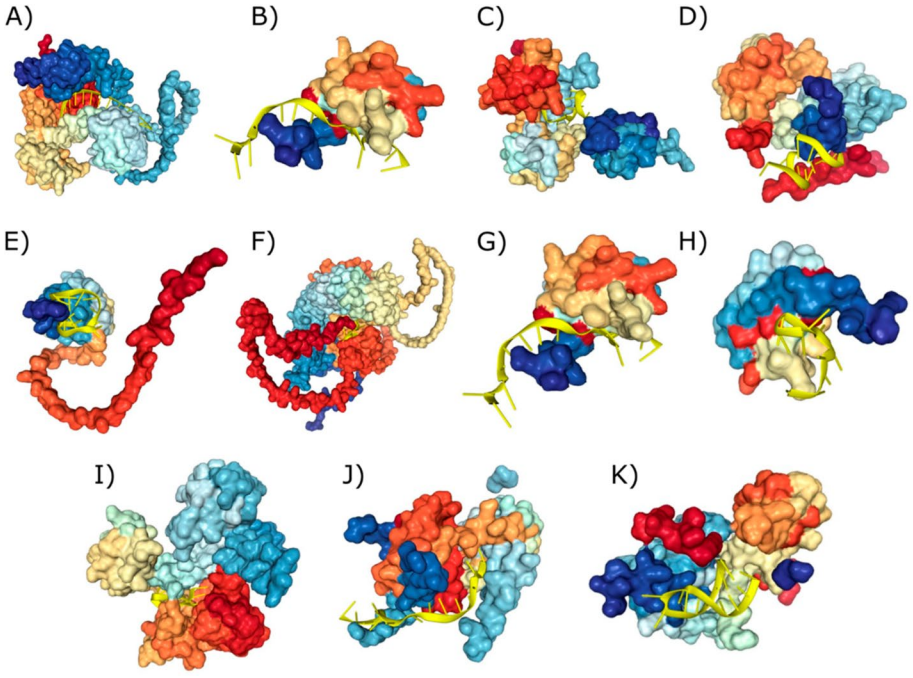

Fig. 4 Docking simulations from the 11 RBP-motif pairs that achieved interaction scores greater than 0.5 in both the RF and SVM classifiers from RPISeq and the proportion of the presence of the motif in the module and in the background (all other modules) are shown. From module M13: **A)** MAAAAAAAAAAAAAA-TcPABP2, **B)**GAGGGGCCGCGACGA-TcRBSR1, **C)**UCCUGCCUGGC-TcDRBD7B, **D)** GGUUGAGWGCACCCTcDRBD13A and **E)** UUGC-GGYCAGGGCGG-TcRBP8B. From module M4, **F)**AUUGUAUAUA-PUF6, **G)**GAGGGGCCGCGACGA-TcRBSR1, **H)**UGCGACGAGG-TcRBSR1, **I)**GAGAGAGCC-TcPABP2, **J)** UGUUUUGCUUYAYKK-TcDRBD5C and **K)** AGWGUGUGUGCUG-TcDRBD12B

Table 6 Docking simulation output information from the 11 RBP-motif pairs that achieved interaction scores greater than 0.5 in both the RF and SVM classifiers from RPISeq

	A	B	C	D	E	F	G	H	I	J	K
Docking score	319.20	281.17	286.04	305.23	234.37	271.16	281.17	237.71	312.17	289.58	277.33
Confidence score	0.9672	0.9324	0.9382	0.9571	0.8439	0.9186	0.9324	0.8525	0.9624	0.9422	0.9273
Ligand rmsd (Å)	44.55	38.03	21.11	29.62	70.71	22.19	38.03	22.13	14.16	58.87	24.61

M13-B, as expected. These same modules also showed strong coverage of a novel cis-element described in the family of trypomastigote-specific TcTASV transcripts ([61] and unpublished results). In addition, the SMUG-S motif -linked to mucin genes expressed in epimastigotes [62] was detected in M3-B, which contains genes upregulated in that life stage. Besides, module M6-B showed a strong overlap with cluster C7 of differentially expressed *T. cruzi* genes from Li et al.[6], and both groups are enriched for the protein phosphorylation GO term. In the other hand, module M5-B is highly correlated with clusters C5, C32, and C49 identified in a previous report [63], which are enriched in the inositol phosphate signaling pathway, corresponding with the increased expression of surface proteins observed in our analysis. Overall, the presence of known cis-elements, together with these functional overlaps between clusters, suggests that several modules capture post-transcriptional regulatory signals active during the parasite's life cycle.

Discussion

In this study, a weighted GCN was constructed using RNA-seq data from all life cycle stages of *T. cruzi* employing the R package CEMiTool [20]. After network analysis, co-expressed gene modules enriched in GO terms were found. Also, hub genes were identified, which are thought to play key roles in driving molecular processes or serve as representatives of the predominant biological functions within the module. The co-expression modules identified in this study revealed distinct expression profiles, some of them strongly associated with specific *T. cruzi* life cycle stages and biological functions [6, 12].

Analysis of M1 gene expression patterns across the parasite development (Fig. 3 and Supplementary Fig. 1) revealed two coordinated yet distinct subsets: one enriched in genes highly expressed in insect vector stages, and another in genes upregulated in mammalian host stages. This suggests that M1, a center node in the network (Fig. 2B), captures stage-specific expression programs that may include genes with functions in specific

host environments and opens the possibility of identifying context-dependent essential genes offering new candidates for therapeutic targeting. Modules M2 and M10 are associated with epimastigotes and amastigotes replicative stages according to our trait correlation analysis. M10 shows enriched GO terms related to mRNA metabolism, as expected for a replicative and metabolically active *T. cruzi* stage. M2-A shows GO term enrichment related to a spectrum of cellular activities, including post transcriptional regulation of gene expression and cytoskeleton related processes.

Interestingly, modules M7, M8 and M11 show distinct expression profiles between early and late replicative stages (exponentially growing versus more latent epimastigotes), with a similar trend in amastigotes, where expression differs between early and late infection times. Despite these shared dynamics, each module is composed of different types of genes, which explains their separation into distinct modules. Module M7 shows high correlation with the epimastigote stage (Fig. 3) and is significantly enriched for GO terms associated with pyruvate metabolism. This enrichment likely reflects the parasite's reliance on glycosomal pyruvate processing enzymes (such as the pyruvate phosphate dikinase, present in this module) to balance energy and redox homeostasis [64]. The trend exhibited by M7-B genes shows a decrease in expression along the epimastigote growth curve. This phenomenon is also observed in amastigotes where these genes are less expressed at the beginning of the infection experiment. Module M8 (Table 1) is enriched for genes involved in DNA metabolism, consistent with its increased expression in amastigotes, the parasite's replicative stage. Similar to module M7, a difference is observed within the epimastigote and amastigote populations where many of the M8-B genes are also highly expressed in exponentially growing epimastigotes but present a lower expression in quiescent epimastigotes, which likely explains why module M8 does not exhibit a strong overall correlation with the epimastigote stage in our global analysis (Fig. 3). In amastigotes, these DNA-metabolism genes show a marked increase in expression during mid and late infection, presumably reflecting the heightened demand for DNA replication at that time (see profile of M8-B in Supplementary Fig. 1). In turn, M11-A (49 genes) is largely composed of hypothetical proteins (~80%), together with a heterogeneous set that includes some surface-protein coding genes (e.g., trans-sialidases, mucins, gp63), as well as eIF4E, exportins and WAHD-domain proteins. By contrast, M11-B (132 genes) is enriched in mitochondrial-associated genes (including ADP/ATP mitochondrial carriers, ATP synthase subunits and mitochondrial RNA-binding proteins). The expression profile of M11-B peaks in exponentially growing epimastigotes, declines as epimastigotes reach a less metabolically active stage and in amastigotes present a progressive increase during intracellular maturation. The expression profile of mitochondrial-associated genes in M11-B shows that modulation of mitochondrial metabolism accompanies life cycle transitions in *T. cruzi*. Proteomic data support this pattern, showing that mitochondrial proteins (including the mitochondrial ADP/ATP carrier) peak in exponentially growing epimastigotes, consistent with elevated oxidative-phosphorylation capacity during active growth in the insect vector [65]. Taken together with the profiles of M7 and M8, these observations suggest a coherent relationship. M7 is enriched in pyruvate/energy-metabolism genes and can provide substrate fluxes that feed mitochondrial respiration; M11 (particularly M11-B) implements energy production via mitochondrial ATP/ADP carriers and ATP synthase subunits; and M8 is linked to DNA metabolism, reflecting biosynthetic and replicative demand when energy is available. This integrated view is further supported by studies describing stage-dependent reprogramming of oxidative phosphorylation, glycolysis/glycosome activity, and the use of alternative energy sources (e.g., proline, fatty acids) in *T. cruzi* [66–68].

Modules M4, M5, and M13 exhibit strong expression in trypomastigotes, and are enriched in GO terms associated with pathogenesis, supporting their involvement in processes related to infectivity, such as immune evasion or host-parasite interactions. This is consistent with previous reports that show these processes are characteristic of this life stage [40]. However, subclusters may differ in functional focus. Analysis of these groups reveal that M4-B is more diverse with terms like "cell adhesion", M5-B is enriched in "glycosylation", whereas M13-B is clearly a more restricted cluster, showing only enrichment in the pathogenesis term. Interestingly, when analyzing gene composition of these modules, M4-B show an enrichment of genes coding for mucins while M5-B is enriched in trans-sialidases and M13-B in MASPs. Conversely, genes exhibiting an opposite expression are associated with translation related functions with annotations such as "translation", "ribosome" and "elongation". However, only module M4-A shows a statistically significant enrichment for translation-related GO terms (Supplementary Fig. 1 and Supplementary Fig. 3). This reciprocal pattern, surface-protein genes upregulated as translation machinery genes are downregulated, has been noted before in differential expression analyses of *T. cruzi* developmental stages [69]. Our co-expression analysis not only confirms this inverse regulation but also demonstrates that it persists throughout the entire *T. cruzi* life cycle. The discussion presented above indicates that the identified modules capture biologically meaningful transcriptional programs associated with distinct phases of the parasite's life cycle.

Analysis of hub genes revealed that, in most cases, the overrepresented GO terms within the modules and the functions attributed to these hub genes were consistent, confirming that these genes typically play key roles within the modules to which they belong. However, several of these hub genes were annotated as hypothetical proteins. The methods most commonly used for annotating genes and proteins of unknown function rely on sequence alignment and similarity searches, with the aim of finding homologous sequences of genes or proteins already annotated in databases to infer possible molecular and cellular functions [70]. However, in *T. cruzi*, many proteins cannot be annotated using these methods, possibly because of the early evolutionary divergence of the species, leading to significant sequence divergence [71]. Currently, numerous bioinformatics tools have been developed for the functional annotation of genes that use other approaches, such as structural homology search [28, 72], ontology-based annotation [73], and machine learning methods [74]. Among the 16 unannotated hub genes, the potential functions of 10 were inferred using at least one of these methodologies. These predicted functions also correlated with the overrepresented GO terms from the corresponding module and with the function of the closest known proteins in the network. One notable aspect of this study is that it provides information on the connectivity of all genes within each module. The description of hub genes enables prioritization for subsequent experimental functional studies and facilitates the identification of potential drug targets.

In *T. cruzi*, the regulation of gene expression primarily occurs post-transcriptionally, although in recent years, the regulation of gene expression has become evident at the epigenetic level [75–79]. The regulation of mRNA steady-state levels is one of the main post-transcriptional regulatory mechanisms: stabilization or degradation modulates the half-life of mRNAs at different stages of the parasite's life cycle or under different conditions, which is primarily determined by sequences present in the UTRs of mRNAs and RBPs that interact with them, mainly at the 3'UTR [62, 80, 81]. Post-transcriptional regulons in eukaryotic cells coordinate the expression of groups of mRNAs encoding functionally related proteins and trans-acting factors [82]. The existence of RNA regulons allows cells to control protein synthesis from genes that are dispersed throughout the genome but are functionally related. These elements often possess conserved sequence motifs and/or RNA secondary structures, and many of these elements are likely to remain undiscovered [82]. Current evidence suggests that RNA regulons govern gene expression in trypanosomatids [7, 83–85]. Specific sequence motifs within the transcripts play a crucial role in defining the fate of mRNAs, influencing both stability and translation efficiency [62]. Here, identifying regulatory elements within the 3'UTRs in the same module provides insight into how gene expression is coordinated at the post-transcriptional level. Among the 77 motif–RBP pairs that were identified, 11 met further stringent criterion for interaction. After performing docking simulations with HDOCK and HADDOCK, most of these simulated interactions had good confidence scores (Table 6 and Supplementary Table 7). These results clearly support the notion that these RBPs indeed bind to the discovered motifs in co-expressed genes and provide the scientific community with new motifs for experimental evaluation. Interestingly, two interactions from Table 5 correspond with established RNA-binding motifs: PABP2 with poly(A) sequences (Fig. 4A) and PUF6 with AUUGUAUAUA, matching the PUF consensus sequence UGUR, which is recognized by RBPs from the PUF family (Fig. 4F) [86]. Our study revealed that PUF6 from *T. cruzi* can bind to this motif, which is present in genes belonging to module M4 and is enriched in genes coding for surface proteins. It has been previously reported that TcPUF6 promotes the degradation of its associated mRNAs through interactions with RNA degradation complexes, especially those that are upregulated in the infective metacyclic trypomastigote form of the parasite [87]. This result may suggest that the lifetime of mRNAs from surface protein genes belonging to module M4 containing this motif may be regulated by TcPUF6, possibly as a potential mechanism for regulating the repertoire of surface proteins throughout the life cycle of this parasite.

Poly(A)-binding proteins (PABPs) regulate mRNA stability, export, and translation by interacting with the poly(A) tail and the eIF4F complex. In *T. cruzi*, TcPABP1 binds TcUBP1 which targets AU-rich 3′UTRs destabilized in trypomastigotes [88]. PABP1 and PABP2 differ in granule localization, where PABP2 with P-body markers and PABP1 in heat-shock granules [89]. Additionally, PABP2 shows broader RNA and protein interactions but lower poly(A) specificity [90]. In this regard, in our study, TcPABP2 was identified as an interactor of two different 3′UTR motifs belonging to modules M4 and M13, both of which are enriched in surface protein-coding genes, GAGAGAGCC and MAAAAAAAAAAAAAA, respectively. In *T. brucei*, DRBD13 plays a crucial role in regulating transcripts encoding surface coat proteins while also influencing RBP6 mRNA. RBP6 is known to modulate the abundance of transcripts containing AU-rich elements (AREs) [91]. Here, we observed an overrepresentation of the motif GGUUGAGWGCACCC within the 3'UTRs of genes in module M13, and docking simulations revealed that TcDRBD13 could bind to this motif, suggesting a role in regulating surface protein-coding gene expression.

In summary, the characterization of these cis-regulatory elements will enhance our understanding of post-transcriptional gene regulation and offer potential targets against key regulators of gene expression in *T. cruzi*.

Concluding remarks

Post-transcriptional regulatory networks in trypanosomatids have been widely studied, revealing that RBPs can exert control over multiple mRNAs, forming interconnected regulatory circuits [84]. In this context, gene co-expression analyses are valuable tools for studying changes in the expression levels of functionally interacting gene clusters. In this study, we aimed to contribute to the understanding of gene expression regulation in this parasite by constructing a GCN using RNA-seq data from the four stages of its life cycle. Our results suggest a correspondence between the enriched functions of the co-expressed gene modules, the identified annotated hub gene functions and established aspects of *T. cruzi* biology. Additionally, there was correspondence with the new annotations obtained in this study from hub genes previously annotated as hypothetical proteins. Additionally, our analysis enabled the prioritization of genes based on their network connectivity, including several that lack functional annotation. Finally, the identification of cis-regulatory elements in 3'UTRs and the potential binding of RBPs significantly advances our understanding of post-transcriptional gene regulation in *T. cruzi*.

This study not only elucidates the complex regulatory networks governing gene expression but also highlights potential targets for therapeutic intervention against key regulators of gene expression from this parasite.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-12095-7.

Supplementary material 1. Figure S1. Gene expression patterns of genes belonging to identified gene co-expression modules (M1-M13) across the four developmental stages: epimastigotes (E, green), metacyclic trypomastigotes (M, blue), amastigotes (A, pink), and bloodstream trypomastigotes (T, purple). Each module was partitioned according to the sign of the correlation between individual genes and the corresponding module eigengene. This approach allowed us to separate genes that increase or decrease together across the ***T. cruzi*** life cycle and to highlight shared expression trends within modules. Each line represents the expression of a single gene across different samples, with the black line indicating the average expression trend. Figure S2. Co-expression graph of the top 5 most connected genes (hub genes) from each module with functional enrichment. In the graph each node represents a gene, and each line connects two genes. Figure S3. Functional classification of genes in modules M4, M5, and M13, clustered into two subsets (A and B) according to the sign of their correlation with the module eigengene. Bars represent the proportion of genes in each functional category. Supplementary File 1. Subnetworks centered on selected hub genes. Each subnetwork shows the hub gene and its nine most strongly connected neighbors based on co-expression strength.

Acknowledgements
We acknowledge all members of the Sección Genómica Funcional at Facultad de Ciencias, Universidad de la República, and the Departamento de Genómica at Instituto de Investigaciones Biológicas Clemente Estable for constant support and discussion. We also thank Programa de Desarrollo de las Ciencias Básicas (PEDECIBA) and ICGEB for their training support and valued collaboration. JGDG is a member of the Research Career of CONICET.

Authors' contributions
PS, LI and AM: conceived the idea and conceptualized the study. LI and PS: acquired funding. LPD, PS, MAD and BG: epimastigote RNA-seq experiments. LI: data analysis. JDG and LI: RBP-motif analysis. LI, PS and JDG: wrote the initial version of the manuscript. All authors: interpretation of results and discussion. All authors read and approved the final version of the manuscript.

Funding
This work received financial support from Programa de Desarrollo de las Ciencias Básicas (PEDECIBA), Comisión Académica de Posgrado (CAP) and Comisión Sectorial de Investigación Científica (CSIC, grant number C249-347) from Uruguay.

Data availability
Publicly available datasets were analyzed in this study. This data can be found in the National Center for Biotechnology (NCBI) Sequence Read Archive (SRA) under projects PRJNA915394, PRJNA251582 and PRJNA251583; and in the European Nucleotide Archive (ENA) under project PRJEB33521. The code used to generate the results is available at GitHub (github.com/bioinfo-iibce/tcruzi-coexpression-network).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Laboratorio de Bioinformática, Departamento de Genómica, Instituto de Investigaciones Biológicas Clemente Estable, Montevideo 11600, Uruguay
[2]Sección Genómica Funcional, Facultad de Ciencias, Universidad de La República, Montevideo 11400, Uruguay
[3]Instituto de Computación, Facultad de Ingeniería, Universidad de La República, Montevideo 11300, Uruguay
[4]Departamento de Genómica, Instituto de Investigaciones Biológicas Clemente Estable, Montevideo 11600, Uruguay
[5]Departamento de Genética, Facultad de Medicina, Universidad de La República, Montevideo 11800, Uruguay
[6]Departamento de Biología Celular y Molecular, Facultad de Ciencias, Sección Biología Celular, Universidad de La República, Montevideo 11400, Uruguay
[7]Instituto de Investigaciones Biotecnológicas, Universidad Nacional de San Martín - Consejo Nacional de Investigaciones Científicas y Técnicas, General San Martín, Prov. de Buenos Aires 1650, Argentina
[8]Escuela de Bio y Nanotecnologías (EByN), Universidad Nacional de San Martín, General San Martín, Prov. de Buenos Aires 1650, Argentina

Received: 10 April 2025 / Accepted: 4 September 2025
Published online: 14 October 2025

References

1. Buscaglia CA, Di Noia JM. *Trypanosoma cruzi* clonal diversity and the epidemiology of Chagas' disease. Microbes Infect. 2003;5(5):419–27.
2. Rassi A, Rassi A, Marin-Neto JA. Chagas disease. Lancet. 2010;375(9723):1388–402.

3. Martín-Escolano J, Marín C, Rosales MJ, Tsaousis AD, Medina-Carmona E, Martín-Escolano R. An updated view of the *Trypanosoma cruzi* life cycle: intervention points for an effective treatment. ACS Infect Dis. 2022;8(6):1107–15.
4. Smircich P, Pérez-Díaz L, Hernández F, Duhagon MA, Garat B. Transcriptomic analysis of the adaptation to prolonged starvation of the insect-dwelling *Trypanosoma cruzi* epimastigotes. Front Cell Infect Microbiol. 2023;6(13):374.
5. Cruz-Saavedra L, Muñoz M, Patiño LH, Vallejo GA, Guhl F, David Ramírez J. Slight temperature changes cause rapid transcriptomic responses in *Trypanosoma cruzi* metacyclic trypomastigotes. Parasit Vectors. 2020. https://doi.org/10.1186/s13071-020-04125-y.
6. Li Y, Shah-Simpson S, Okrah K, Belew AT, Choi J, Caradonna KL, et al. Transcriptome Remodeling in *Trypanosoma cruzi* and Human Cells during Intracellular Infection. PLoS Pathog. 2016;12(4):1–30.
7. Queiroz R, Benz C, Fellenberg K, Hoheisel JD, Clayton C. Transcriptome analysis of differentiating trypanosomes reveals the existence of multiple post-transcriptional regulons. BMC Genomics. 2009;10(1):495.
8. Chávez S, Eastman G, Smircich P, Becco LL, Oliveira-Rizzo C, Fort R, et al. Transcriptome-wide analysis of the *Trypanosoma cruzi* proliferative cycle identifies the periodically expressed mrnas and their multiple levels of control. PLoS ONE. 2017;12(11):e0188441.
9. dos Santos CMB, Ludwig A, Kessler RL, Rampazzo R de CP, Inoue AH, Krieger MA, et al. *Trypanosoma cruzi* transcriptome during axenic epimastigote growth curve. Mem Inst Oswaldo Cruz. 2018 May 1;113(5):1–13.
10. Dillon LAL, Suresh R, Okrah K, Corrada Bravo H, Mosser DM, El-Sayed NM. Simultaneous transcriptional profiling of *Leishmania major* and its murine macrophage host cell reveals insights into host-pathogen interactions. BMC Genomics. 2015;16(1):1–15.
11. Minning TA, Weatherly DB, Atwood J, Orlando R, Tarleton RL. The steady-state transcriptome of the four major life-cycle stages of *Trypanosoma cruzi*. BMC Genomics [Internet]. 2009 [cited 2020 Nov 6];10. Available from: https://pubmed.ncbi.nlm.nih.gov/19664227/.
12. Berná L, Chiribao ML, Greif G, Rodriguez M, Alvarez-Valin F, Robello C. Transcriptomic analysis reveals metabolic switches and surface remodeling as key processes for stage transition in *Trypanosoma cruzi*. PeerJ. 2017;2017(3):e3017.
13. Jensen BC, Ramasamy G, Vasconcelos EJR, Ingolia NT, Myler PJ, Parsons M. Extensive stage-regulation of translation revealed by ribosome profiling of *Trypanosoma brucei*. BMC Genomics. 2014;15(1):1–21.
14. Rastrojo A, Carrasco-Ramiro F, Martín D, Crespillo A, Reguera RM, Aguado B, et al. The transcriptome of *Leishmania major*in the axenic promastigote stage: transcript annotation and relative expression levels by RNA-seq. BMC Genomics. 2013;14(1):223.
15. Mwangi KW, Macharia RW, Bargul JL. Gene co-expression network analysis of *Trypanosoma brucei* in tsetse fly vector. Parasit Vectors. 2021;14(1):1–11.
16. Martinez-Hernandez JE, Aliaga-Tobar V, González-Rosales C, Monte-Neto R, Martin AJM, Maracaja-Coutinho V. Comparative and systems analyses of Leishmania spp. non-coding RNAs through developmental stages. PLoS Negl Trop Dis. 2025 May 28;19(5):e0013108.
17. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010;
18. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 2016 345. 2016;34(5):525–7.
19. Love MI, Anders S, Huber W. Analyzing RNA-seq data with DESeq2. 2018; Available from: https://doi.org/10.1186/s13059-014-0550-8. Cited 2022 Sep 8.
20. Russo PST, Ferreira GR, Cardozo LE, Bürger MC, Arias-Carrasco R, Maruyama SR, et al. CEMiTool: A Bioconductor package for performing comprehensive modular co-expression analyses. BMC Bioinformatics. 2018 Feb 20;19(1).
21. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. 2008; Available from: http://www.biomedcentral.com/1471-2105/9/559. Cited 2020 Aug 26.
22. Jordan IK, Mariño-Ramírez L, Wolf YI, Koonin EV. Conservation and coevolution in the scale-free human gene coexpression network. Mol Biol Evol. 2004;21(11):2058–70.
23. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for R. Bioinformatics. 2008;24(5):719–20.
24. Csárdi G, Nepusz T, Müller K, Horvát S, Traag V, Zanini F, et al. igraph for R: R interface of the igraph library for graph theory and network analysis [Internet]. Zenodo; 2025 [cited 2025 Mar 19]. Available from: https://zenodo.org/records/14736815
25. An Implementation of Grammar of Graphics for Graphs and Networks.. Available from: https://ggraph.data-imaginist.com/. Cited 2025 Mar 19
26. Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res. 2009;38(SUPPL.1):D457.
27. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nat 2021 5967873. 2021;596(7873):583–9.
28. van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, et al. Fast and accurate protein structure search with Foldseek. Nat Biotechnol. 2024;42(2):243–6.
29. Radío S, Fort RS, Garat B, Sotelo-Silveira J, Smircich P. UTRme: a scoring-based tool to annotate untranslated regions in Trypanosomatid genomes. Front Genet. 2018;18(9):671.
30. Grant CE, Bailey TL. XSTREME: Comprehensive motif analysis of biological sequence datasets. bioRxiv. 2021 Sep 3;2021.09.02.458722.
31. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME Suite: Tools for motif discovery and searching. Nucleic Acids Res. 2009;37(SUPPL.2):W202–8.
32. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics. 2011;27(7):1017–8.
33. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
34. Muppirala UK, Honavar VG, Dobbs D. Predicting RNA-protein interactions using only sequence information. BMC Bioinformatics. 2011;12(1):489.
35. Biesiada M, Purzycka KJ, Szachniuk M, Blazewicz J, Adamiak RW. Automated RNA 3D Structure Prediction with RNAComposer. In: Turner DH, Mathews DH, editors. RNA Structure Determination: Methods and Protocols [Internet]. New York: Springer; 2016 [cited 2024 Apr 10]. p. 199–215. Available from: https://doi.org/10.1007/978-1-4939-6433-8_13.
36. Yan Y, Zhang D, Zhou P, Li B, Huang SY. HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. Nucleic Acids Res. 2017;45(W1):W365–73.
37. Honorato RV, Trellet ME, Jiménez-García B, Schaarschmidt JJ, Giulini M, Reys V, et al. The HADDOCK2.4 web server for integrative modeling of biomolecular complexes. Nat Protoc. 2024;19(11):3219–41.
38. Albert R. Scale-free networks in cell biology. J Cell Sci. 2005;118(21):4947–57.
39. Barabási AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet 2004 52. 2004;5(2):101–13.
40. Oliveira C, Holetz FB, Alves LR, Ávila AR. Modulation of virulence factors during *Trypanosoma cruzi* differentiation. Pathogens. 2023;12(1):32.
41. Abrahim M, Machado E, Alvarez-Valín F, de Miranda AB, Catanho M. Uncovering Pseudogenes and Intergenic Protein-coding Sequences in TriTryps' Genomes. Genome Biol Evol [Internet]. 2022 [cited 2023 Jun 2];14(10). Available from: https://academic.oup.com/gbe/article/14/10/evac142/6754225.
42. Pink RC, Carter DRF. Pseudogenes as regulators of biological function. Essays Biochem. 2013;54(1):103–12.
43. Jia B, Jia X, Kim KH, Jeon CO. Integrative view of 2-oxoglutarate/Fe(II)-dependent oxygenase diversity and functions in bacteria. Biochim Biophys Acta-Gen Subj. 2017;1861(2):323–34.
44. Marbaix AY, Noël G, Detroux AM, Vertommen D, Van Schaftingen E, Linster CL. Extremely conserved ATP- or ADP-dependent enzymatic system for nicotinamide nucleotide. J Biol Chem. 2011;286(48):41246–52.
45. Yu Hsuan Teng elicia, Wang Y, Luen Tang B. The syntaxins. Genome Biol. 2001;2(11). Available from: http://genomebiology.com/2001/2/11/reviews/3012
46. Zhang J, Wang H, Imhof S, Zhou X, Liao S, Atanasov I, et al. Structure of the trypanosome paraflagellar rod and insights into non-planar motility of eukaryotic cells. Cell Discov. 2021. https://doi.org/10.1038/s41421-021-00281-2.
47. Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, Söding J. HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinformatics. 2019;20(1):473.
48. Muniyappa K, Kshirsagar R, Ghodke I. The HORMA domain: an evolutionarily conserved domain discovered in chromatin-associated proteins, has unanticipated diverse functions. Gene. 2014;545(2):194–7.
49. Chao T, Zhou X, Cao B, Liao P, Liu H, Chen Y, et al. Pleckstrin homology domain-containing protein PHLDB3 supports cancer growth via a negative feedback loop involving p53. Nat Commun 2016 71. 2016;7(1):1–12.
50. Ferguson DJP, Sahoo N, Pinches RA, Bumstead JM, Tomley FM, Gubbels MJ. MORN1 has a conserved role in asexual and sexual development across the Apicomplexa. Eukaryot Cell. 2008;7(4):698–711.

51. Morriswood B, Schmidt K. A MORN repeat protein facilitates protein entry into the flagellar pocket of *Trypanosoma brucei*. Eukaryot Cell. 2015;14(11):1081–93.
52. Benz C, Nilsson D, Andersson B, Clayton C, Guilbride DL. Messenger RNA processing sites in *Trypanosoma brucei*. Mol Biochem Parasitol. 2005;143(2):125–34.
53. Duhagon MA, Dallagiovanna B, Garat B. Unusual features of Poly[dT-dG]·[dC-dA] stretches in CDS-flanking regions of *Trypanosoma cruzi* genome. Biochem Biophys Res Commun. 2001;287(1):98–103.
54. Pastro L, Smircich P, Pérez-Díaz L, Duhagon MA, Garat B. Implication of CA repeated tracts on post-transcriptional regulation in *Trypanosoma cruzi*. Exp Parasitol. 2013;134(4):511–8.
55. Smith M, Blanchette M, Papadopoulou B. Improving the prediction of mRNA extremities in the parasitic protozoan *Leishmania*. BMC Bioinformatics. 2008;9(1):1–12 (**Mar 20;**).
56. Uyhazi KE, Yang Y, Liu N, Qi H, Huang XA, Mak W, et al. Pumilio proteins utilize distinct regulatory mechanisms to achieve complementary functions required for pluripotency and embryogenesis. Proc Natl Acad Sci U S A. 2020;117(14):7851–62.
57. Young LE, Sanduja S, Bemis-Standoli K, Pena EA, Price RL, Dixon DA. The mRNA binding proteins HuR and Tristetraprolin regulate cyclo-oxygenase 2 expression during colon carcinogenesis. Gastroenterology. 2009;136(5):1669–79.
58. Anant S, Houchen CW. HuR and TTP: two RNA binding proteins that deliver message from the 3′ end. Gastroenterology. 2009;136(5):1495–8.
59. Degese MS, Tanos T, Naipauer J, Gingerich T, Chiappe D, Echeverria P, et al. An interplay between the p38 MAPK pathway and AUBPs regulates c-fos mRNA stability during mitogenic stimulation. Biochem J. 2015;467(1):77–90.
60. Sabalette KB, Romaniuk MA, Noé G, Cassola A, Campo VA, De Gaudenzi JG. The RNA-binding protein TcUBP1 up-regulates an RNA regulon for a cell surface–associated *Trypanosoma cruzi* glycoprotein and promotes parasite infectivity. J Biol Chem. 2019;294(26):10349–64.
61. Bernabó G, Levy G, Ziliani M, Caeiro LD, Sánchez DO, Tekiel V. Tctasv-c, a protein family in *Trypanosoma cruzi* that is predominantly trypomastigote-stage specific and secreted to the medium. PLoS ONE. 2013;8(7):e71192.
62. Di Noia JM, D'Orso I, Sánchez DO, Frasch ACC. AU-rich elements in the 3′-untranslated region of a new mucin-type gene family of *Trypanosoma cruzi* confers mRNA instability and modulates translation efficiency. J Biol Chem. 2000;275(14):10218–27.
63. De Gaudenzi JG, Carmona SJ, Ağuero F, Frasch AC. Genome-wide analysis of 3′-untranslated regions supports the existence of post-transcriptional regulons controlling gene expression in trypanosomes. PeerJ [Internet]. 2013 [cited 2022 Aug 18];1(1). Available from: https://pubmed.ncbi.nlm.nih.gov/23904995/
64. Acosta H, Dubourdieu M, Quiñones W, Cáceres A, Bringaud F, Concepción JL. Pyruvate phosphate dikinase and pyrophosphate metabolism in the glycosome of *Trypanosoma cruzi* epimastigotes. Comp Biochem Physiol B Biochem Mol Biol. 2004;138(4):347–56.
65. Avila CC, Mule SN, Rosa-Fernandes L, Viner R, Barisón MJ, Costa-Martins AG, et al. Proteome-wide analysis of *Trypanosoma cruzi* exponential and stationary growth phases reveals a subcellular compartment-specific regulation. Genes. 2018;9(8):413.
66. Quiñones W, Acosta H, Gonçalves CS, Motta MCM, Gualdrón-López M, Michels PAM. Structure, Properties, and Function of Glycosomes in *Trypanosoma cruzi*. Front Cell Infect Microbiol. 2020;10. Available from: https://www.frontiersin.org/journals/cellular-and-infection-microbiology/articles/ Cited 2025 Sep 1. https://doi.org/10.3389/fcimb.2020.00025/full
67. Souza ROO, Damasceno FS, Marsiccobetre S, Biran M, Murata G, Curi R, et al. Fatty acid oxidation participates in resistance to nutrient-depleted environments in the insect stages of Trypanosoma cruzi. PLOS Pathog. 2021 abr;17(4):e1009495.
68. Alencar MB, Marsiccobetre S, Mengarda AC, Ballari MS, Silber AM. Energy metabolism in *Trypanosoma cruzi*: the validated and putative bioenergetic and carbon sources. mBio. 2025;16(6):e02215-24.
69. Smircich P, Eastman G, Bispo S, Duhagon MA, Guerra-Slompo EP, Garat B, et al. Ribosome profiling reveals translation control as a key mechanism generating differential gene expression in *Trypanosoma cruzi*. BMC Genomics. 2015;16(1):1–14.
70. Stein L. Genome annotation: from sequence to biology. Nat Rev Genet. 2001;2(7):493–503.
71. Salavati R, Najafabadi HS. Sequence-based functional annotation: what if most of the genes are unique to a genome? Trends Parasitol. 2010;26(5):225–9.
72. Krissinel E, Henrick K. Multiple Alignment of Protein Structures in Three Dimensions. Lect Notes Comput Sci Subser Lect Notes Artif Intell Lect Notes Bioinforma. 2005;3695 LNBI:67–78.
73. Törönen P, Medlar A, Holm L. PANNZER2: a rapid functional annotation web server. Nucleic Acids Res. 2018;46(W1):W84–8.
74. Kulmanov M, Khan MA, Hoehndorf R. DeepGO: predicting protein functions from sequence and interactions using a deep ontology-aware classifier. Bioinformatics. 2018;34(4):660–8.
75. Díaz-Viraqué F, Chiribao ML, Libisch MG, Robello C. Genome-wide chromatin interaction map for *Trypanosoma cruzi*. Nat Microbiol. 2023;8(11):2103–14.
76. Lima ARJ, Araujo CB de, Bispo S, Patané J, Silber AM, Elias MC, et al. Nucleosome landscape reflects phenotypic differences in *Trypanosoma cruzi* life forms. PLOS Pathog. 2021 ene;17(1):e1009272.
77. Lima ARJ, Silva HG de S, Poubel S, Rosón JN, de Lima LPO, Costa-Silva HM, et al. Open chromatin analysis in *Trypanosoma cruzi* life forms highlights critical differences in genomic compartments and developmental regulation at tDNA loci. Epigenetics Chromatin. 2022;15(1):22.
78. Beati P, Stepñicka MM, Larrea SCV, Smircich P, Alonso GD, Ocampo J. Improving genome-wide mapping of nucleosomes in *Trypanosome cruzi*. PLoS ONE. 2023;18(11):e0293809.
79. Ocampo J, Carena S, López M del R, Vela VS, Zambrano Siri RT, Balestra SA, et al. Trypanosomatid histones: the building blocks of the epigenetic code of highly divergent eukaryotes. Biochem J. 2025;482(06):325–40.
80. Coughlin BC, Teixeira SMR, Kirchhoff LV, Donelson JE. Amastin mRNA abundance in *Trypanosoma cruzi* is controlled by a 3′- untranslated region position-dependent cis-element and an untranslated region-binding protein. J Biol Chem. 2000;275(16):12051–60.
81. Vanhamme L, Pays E. Control of gene expression in trypanosomes. Microbiol Rev. 1995;59(2):223–40.
82. Keene JD. RNA regulons: coordination of post-transcriptional events. Nat Rev Genet 2007 87. 2007;8(7):533–43.
83. Noé G, De Gaudenzi JG, Frasch AC. Functionally related transcripts have common RNA motifs for specific RNA-binding proteins in trypanosomes. BMC Mol Biol. 2008;9(1):1–19.
84. Ouellette M, Papadopoulou B. Coordinated gene expression by post-transcriptional regulons in African trypanosomes. J Biol. 2009;8(11):1–4.
85. De Gaudenzi JG, Carmona SJ, Ağuero F, Frasch AC. Genome-wide analysis of 3′-untranslated regions supports the existence of post-transcriptional regulons controlling gene expression in trypanosomes. PeerJ [Internet]. 2013 [cited 2023 Mar 14];1(1). Available from: https://pubmed.ncbi.nlm.nih.gov/23904995/
86. Wickens M, Bernstein DS, Kimble J, Parker R. A PUF family portrait: 3′UTR regulation as a way of life. Trends Genet. 2002;18(3):150–7.
87. Dallagiovanna B, Correa A, Probst CM, Holetz F, Smircich P, de Aguiar AM, et al. Functional genomic characterization of mRNAs associated with TcPUF6, a Pumilio-like protein from *Trypanosoma cruzi*. J Biol Chem. 2008;283(13):8266–73.
88. D'Orso I, Frasch ACC. TcUbp-1, an mrna destabilizing factor from trypanosomes, homodimerizes and interacts with novel au-rich element- and poly(a)-binding proteins forming a ribonucleoprotein complex. J Biol Chem. 2002;277(52):50520–8 (**Dec 27;**).
89. Kramer S, Bannerman-Chukualim B, Ellis L, Boulden EA, Kelly S, Field MC, et al. Differential Localization of the Two *T. brucei* Poly(A) Binding Proteins to the Nucleus and RNP Granules Suggests Binding to Distinct mRNA Pools. PLOS ONE. 2013 ene;8(1):e54004.
90. Zoltner M, Krienitz N, Field MC, Kramer S. Comparative proteomics of the two *T. brucei* PABPs suggests that PABP2 controls bulk mRNA. PLoS Negl Trop Dis. 2018;12(7):e0006679.
91. Jha BA, Gazestani VH, Yip CW, Salavati R. The DRBD13 RNA binding protein is involved in the insect-stage differentiation process of *Trypanosoma brucei*. FEBS Lett. 2015;589(15):1966–74.

Publisher's Note