RESEARCH ARTICLE

TrypsNetDB: An integrated framework for the functional characterization of trypanosomatid proteins

Vahid H. Gazestani, Chun Wai Yip, Najmeh Nikpour, Natasha Berghuis, Reza Salavati*

Institute of Parasitology, McGill University, Ste. Anne de Bellevue, Quebec, Canada

* reza.salavati@mcgill.ca

Abstract

Trypanosomatid parasites cause serious infections in humans and production losses in livestock. Due to the high divergence from other eukaryotes, such as humans and model organisms, the functional roles of many trypanosomatid proteins cannot be predicted by homology-based methods, rendering a significant portion of their proteins as uncharacterized. Recent technological advances have led to the availability of multiple systematic and genome-wide datasets on trypanosomatid parasites that are informative regarding the biological role(s) of their proteins. Here, we report TrypsNetDB (http://trypsNetDB.org), a web-based resource for the functional annotation of 16 different species/strains of trypanosomatid parasites. The database not only visualizes the network context of the queried protein(s) in an intuitive way but also examines the response of the represented network in more than 50 different biological contexts and its enrichment for various biological terms and pathways, protein sequence signatures, and potential RNA regulatory elements. The interactome core of the database, as of Jan 23, 2017, contains 101,187 interactions among 13,395 trypanosomatid proteins inferred from 97 genome-wide and focused studies on the interactome of these organisms.

Author summary

Methods to predict protein function based on sequences enable the rapid annotation of newly sequenced genomes. However, as most of these methods rely on homology-based approaches, non-conserved proteins in trypanosomatids remain elusive for annotation, rendering approximately half of the sequenced proteins uncharacterized. In this study, we developed a user friendly integrated database, TrypsNetDB, which fills multiple gaps in the field by depositing the current interactome knowledge on trypanosomatid proteins and combining this information with other available resources accompanied by related statistical analyses. The database allows automatic inter-species mapping of available data to better characterize the queried proteins in the species of interest. The database is built on fast and reliable ASP.Net framework and provides **(i)** a significant increase in the genome-wide functional annotation of trypanosomatid proteins, **(ii)** potential novel

targets for therapeutics against trypanosomatids, and **(iii)** a robust methodology that can be adapted for the functional annotation of other non-model organisms.

Introduction

Trypanosomatid parasites cause life-threatening diseases in humans and major production losses in animals. They pose global threats, and various issues are associated with available drugs against trypanosomatids (including tolerability, cost, and resistance), necessitating the identification of novel essential parasitic-specific pathways/genes as potential drug targets [1]. However, as supported by whole genome sequencing data, it is well known that species of the trypanosomatid family, while showing high similarity in proteomes with one another, are highly diverged from other eukaryotes [2–5]. This makes the annotation transfer of nearly half of their proteome by homology-based approaches from model organisms unreliable [3].

During the past decade, several genome-wide and focused studies have been conducted to functionally characterize trypanosomatid proteins. The construction of global and local protein interaction maps has served as one of the main resources for functional annotation by reflecting the molecular context of proteins in a cell [6–15]. Several experimental techniques exist to identify the interacting partners of proteins that differ in selectivity and sensitivity. Therefore, one major challenge in the study of protein interactions is the ability to distinguish between the correctly associated proteins from confounding elements that are present in the results of these experiments. It is also helpful to know the potential interacting proteins that are missing from the results of an experiment based on previously known knowledge of the species of interest or other related trypanosomatid species. Several databases have been developed to represent the experimentally identified or computationally inferred physical and functional protein interactions [16–21]. Such databases greatly help researchers to interrogate cellular processes and gain a systems level view of the protein(s) of choice. Although it is of critical importance for studies on trypanosomatids, only a limited number of databases cover information on protein interactions of these parasites, and such interactions are mostly predicted by transferring the available data from other eukaryotes, missing most parts of the published data on trypanosomatid species [16].

Another major approach for the functional characterization of proteins stems from recent technological advances that have allowed measuring transcriptome, proteome, and transcript half-life changes in response to environmental changes, different life stages, or cell conditions [8, 22–38]. Moreover, it is possible to gain insights on the function of a protein by gathering information on: 1) its annotation from resources such as gene ontology, KEGG pathways, and the BioCyc database [37, 39]; 2) protein characteristics, such as protein sequence motifs, isoelectric points, molecular weights, and the number of transmembrane domains; 3) the essentiality of gene knock-down on cell survival [33]; and 4) the potential cis-regulatory elements present in the 3′-UTR of the gene and the collective response of genes containing that regulatory element to environmental changes [29]. Currently, the TriTrypDB database is a gene-centric framework devoted to the kinetoplastid parasites and provides extensive information on the queried protein ranging from genomic sequence and position to involved biological pathways and captured responses in previously reported studies [38]. However, in many cases, researchers are interested in knowing the collective response of a list of pre-specified proteins along with their interacting partners according to large-scale studies rather than focusing on one protein. Combining interaction data with enrichment analyses of gene ontology, molecular pathways, gene essentiality, and protein sequence features is the key to perceiving the function of proteins.

Here, we describe TrypsNetDB, a user friendly, integrated database that fills the aforementioned gaps by not only depositing the current interactome knowledge on trypanosomatid proteins but also combining such information with other available resources accompanied with related statistical analyses. Moreover, the database automatically performs inter-species mapping of the available data and provides information to allow for a better characterization of the queried proteins in the species of interest. Finally, based on the built-in features, the database can help researchers with their interactome related experiments by distinguishing the likely binding partners of a protein from confounding elements identified in their experiments and suggesting other potentially interacting proteins that are missing from the list of queried proteins. Built on powerful ASP.Net framework, the database performance is fast and reliable. TrypsNetDB is freely available at trypsNetDB.org.

Program description and methods

Overall view of the database

The current release of the database is focused on physical protein interaction data that are already published in the trypanosomatid field, supporting 16 trypanosomatid parasites including *T. cruzi* strain CL Brener, *T. cruzi* CL Brener Esmeraldo-like, *T. cruzi* CL Brener Non-Esmeraldo-like, *T. brucei* gambiense DAL972, *T. brucei* Lister strain 427, *T. vivax* Y486, *T. evansi* strain STIB 805, *T. brucei* TREU927, *L. major* strain Friedlin, *L. mexicana* MHOM/GT/ 2001/U1103, *L. infantum* JPCM5, *L. donovani* BPK282A1, *L. braziliensis* MHOM/BR/75/ M2903, *L. braziliensis* MHOM/BR/75/M2904, *L. arabica* strain LEM1108, and *L. enriettii* strain LEM3045. Fig 1 represents the schematic architecture of the database. To systematically extract the protein interaction data, we searched the NCBI PubMed database using the keywords *Trypanosoma* and *Leishmania* and extracted all resultant abstracts. Next, by a manual search, initial positive and negative gold standard sets were constructed by considering 46 and 251 articles, respectively. A multinomial naïve Bayes classifier was used to prioritize 6581 articles that were more likely to contain protein interaction data based on the abstract content with an estimated probability greater than 0.75. By a manual inspection of some articles, the initial positive and negative gold standard set was expanded to 60 and 332 articles, respectively (with extra attention on keeping the diversity of the gold standard sets to reduce the chance of biased predictions). The multinomial naïve Bayes classifier was re-trained and then re-applied to all the extracted abstracts from PubMed using the new gold standard set. A total of 1996 articles that were likely to include interaction data were identified (estimated probability of 0.9). By reviewing the articles of the final list, we could extract protein interaction data from 97 different studies. The interaction data were obtained using a variety of techniques, including affinity purification, immunoprecipitation, yeast two-hybrid (Y2H), fractionation patterns, and other possible experimental techniques. We have only considered the syntenic orthologs reported by TriTrypDB to transfer the inter-species information. Users can query the database based on either tritrypDB IDs (recognizing IDs of recent and older versions of the database) or gene names. Support for the remainder of the trypanosomatid species is scheduled to be added in the coming months. In cases where the gene names match multiple organisms, the user will be asked to select the species of interest from a dropdown box. As shown in Fig 2, querying of protein(s) will redirect the user to the interaction page, which is composed of the following three main elements: information panel, network, and reference section.

The information panel can be used to explore the details of the three sections of annotations, protein descriptions, and the constructed network. The annotation tab contains gene set enrichment analyses of the combined set of queried proteins with the suggested proteins by the database (i.e., proteins with gray background) for enrichment in the following five different
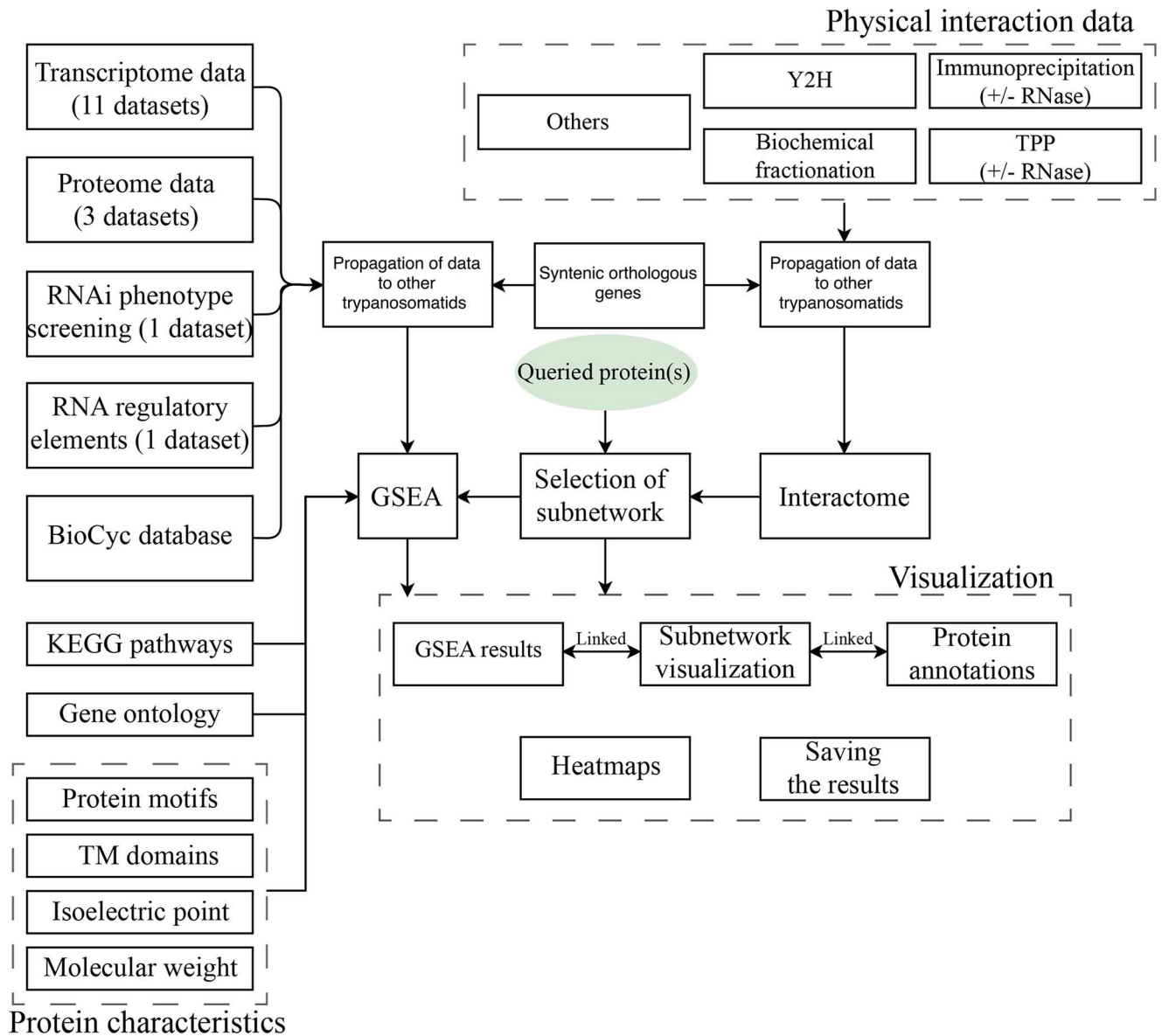
Fig 1. Architecture of TrypsNetDB. The database integrates multiple resources to help functional characterization of trypanosomatid proteins. GSEA: gene set enrichment analysis; TM domains: transmembrane domains; TPP: tagged protein purification.

categories: 1) GO & Pathway: genes are examined for enrichment in Gene Ontology, KEGG pathway, and BioCyc annotations using hypergeometric tests. Terms with Benjamini-Hochberg corrected p-values less than 0.05 are reported back to the user. Hovering over each term will highlight the proteins that are associated with the term. Clicking on each represented term will show a description of the term, its category, number of proteins in the network associated with that term, and the corresponding corrected p-value. 2) Sequence & Structure: The sequence and structural features of the proteins are examined, such as the protein motifs, isoelectric points, molecular weights, and predicted number of transmembrane domains (statistical test for protein motifs is based on the hypergeometric test, and for the other categories based on Wilcoxon-Mann-Whitney rank sum test). This information can provide a complementary view of the function of the proteins. For example, a group of soluble interacting
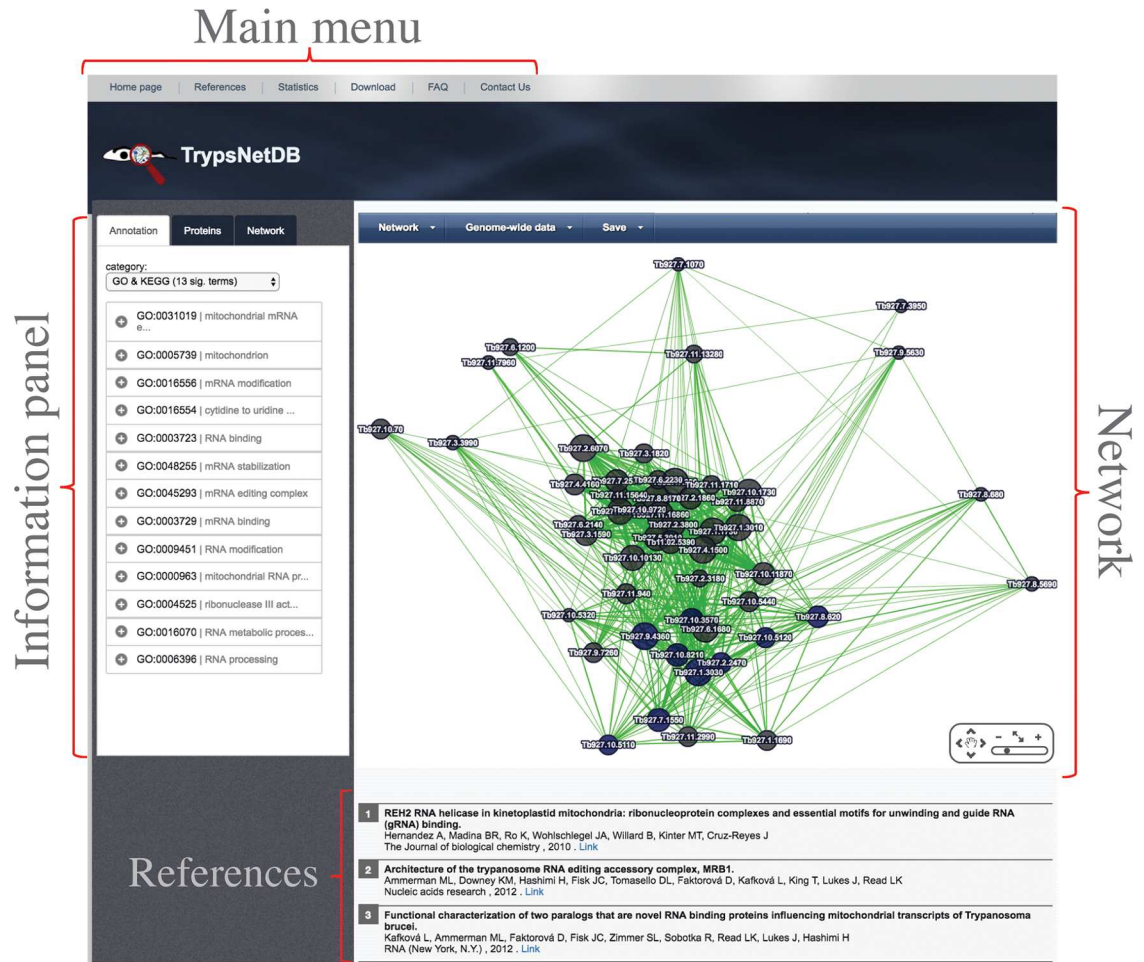
Fig 2. The main result page of TrypsNetDB. The main result page is composed of three elements; that is, information panel, network section, and references section. The information panel contains enrichment analysis results, brief characteristics of the proteins, and the interaction types present in the illustrated network. The network section can be used to explore the interactions among the proteins. Using the menu on the top of the interaction section, users can change the visualization style of the network, access the dynamic heatmaps from various genome-wide data, and save the results. The reference section includes studies from which the interaction data of the illustrated network were extracted with the direct PubMed link.

doi:10.1371/journal.pntd.0005368.g002

proteins are expected to have a significantly low number of transmembrane domains. Likewise, proteins interacting with RNA and DNA are expected to have high isoelectric points. Similar to the GO and pathway enrichment results, hovering over significantly enriched protein motifs will highlight the associated proteins in the network. 3) Expression patterns: the proteins in the network are examined for their collective transcriptome and proteome responses across 48 distinct samples using Wilcoxon-Mann-Whitney rank sum test. Each sample is color coded with yellow and blue indicating over-expression/enrichment and under-expression/depletion, respectively. Statistically significant terms with p-values less than 0.05 are highlighted by darker colors, while non-significant conditions are semi-transparent. The 48 considered cell states were obtained from genome-wide experiments on *T. brucei*, *T. cruzi*, and *L. infantum* [22–24, 26–28, 30–32, 34–36, 40, 41]. By considering syntenic orthologs (as defined in TriTrypDB), the database automatically propagates the information to other trypanosomatid species. Clicking on each sample will open information on the title of the sample, a

description of the results of the statistical tests, the calculated p-value, the title of the study that published the sample and its PMID with a link to the PubMed abstracts. 4) Gene essentiality: The essentiality of proteins in four different cell conditions of *T. brucei* are examined based by application of hypergeometric tests on the results of a genome-wide phenotyping study [33]. Ortholog mapping is performed for cases in which the queried organism is not *T. brucei*. 5) 3′ regulatory elements: Using a novel approach, we recently predicted 88 cis-regulatory elements that are potentially involved in the developmental regulation of *T. brucei* [29]. Although only a limited number of functional elements have been identified thus far, by a rigorous analysis of results, we showed that 11 predicted motifs strikingly resemble previously identified regulatory elements in trypanosomatids, suggesting the high accuracy of the predictions. This section examines whether the 3′-UTRs of the orthologs in the set of proteins in *T. brucei* are significantly enriched for any of the predicted 88 motifs using hypergeometric tests. In cases where enrichment is found, the motif logo along with the transcriptome and proteome responses of the motif in different cell conditions are reported.

The proteins tab provides brief information (such as transcript and protein length, isoelectric point, molecular weight, etc.) with a link to the TriTrypDB database in a sorted way, starting from the queried proteins and ending with the proteins that were included in the network by the program (these suggested proteins have been highly connected to the queried proteins based on literature derived interactions).

The network tab can be used to explore the contribution of each experimental technique to the construction of the illustrated network. In two cases of tagged affinity purification and immunoprecipitation in which interactions can show indirect associations, the database distinguishes between interactions that are identified based on RNase treatment of the samples from those that are not. Hovering over each technique will highlight the interactions that they support. It is also possible to filter some of the techniques by unchecking the corresponding checkboxes and clicking on the "set filters" button.

The network section, using a dynamically interactive interface, represents the interactions among the proteins with each protein indicated by a circular node. It is possible to zoom in or out of the network and reposition the proteins. Queried proteins and other proteins suggested by the database are shown in blue and gray, respectively. The node size of proteins indicates the number of interactions that they have in the global network with larger nodes representing nodes with a higher number of interactions. Selecting a protein by clicking on it will highlight the first neighbors of that protein and open the corresponding information in the proteins tab of the information panel. Finally, the network option on the top-left part of the network section can be used to automatically rearrange the network for perhaps a better presentation or to show/hide the protein labels, which may prove useful for the visualization of relatively large networks.

The reference section provides the references from which the interactions were extracted. The full source of resources used for the extraction of the interaction data can be accessed by going to the "References" section from top menu or going directly to the trypsNetDB.org/references.aspx webpage.

Genome-wide data section

The "Genome-wide data" section on the menu enables users to visualize the genome-wide data available for the queried proteins and their interacting partners suggested by the database. The supported genome-wide data in the current release of the database are categorized in three main groups of fractionation patterns, gene expression patterns, and phenotypic effects, with each containing sub-categories. Users can select one of the main categories (indicated

with a blue background) to represent all related sub-categories at once or directly select the sub-categories.

This part of the database is particularly useful for the validation of results obtained from interactome-related experiments (such as affinity purifications) by helping users to distinguish between direct binding partners of a protein from potentially spurious elements. For example, the fractionation heatmaps can be exploited to assess whether the potentially interacting proteins show similar fractionation patterns (Fig 3). Currently, the database provides fractionation patterns for whole-cell, mitochondrial-enriched, and cytosolic-enriched cell extracts that can be informative for the localization of previously unannotated proteins (i.e., mitochondrial proteins are expected to be identified in the mitochondrial-enriched fractions while depleted in the cytosolic-enriched sample). Fractionation patterns are also informative regarding the nature of the interactions. As described elsewhere [8], glycerol gradient-based fractionation patterns can capture more transient interactions, while ion exchange-based fractionation favors more stable interactions due to the presence of a salt gradient. Finally, physically interacting proteins are expected to be involved in similar biological processes and, hence, show similar expression patterns and degrees of essentiality in each cell state, which can be easily assessed using the corresponding heatmaps.

Saving the results

By going to the save option on the top of the network section, users can save the whole represented network or only the sub-networks that are supported by a specific experimental technique. It is also possible to save the enrichment analysis results and the annotations of genes, such as the description, transcript or protein characteristics (length, weight, isoelectric point, and identified SNPs), and gene ontology. Users can also use the save query list option for later regeneration of the same results.

Implementation

The web application is developed based on the.Net framework 4.5 technology. To improve the performance, the statistical analysis modules (including hypergeometric test, Wilcoxon-Mann-Whitney test, and Benjamini-Hochberg p-value adjustment procedure) were implemented in C# and added as a library to the web application and the performance of the modules has been validated by comparing the results with those of MATLAB 2015b on multiple test sets to ensure accuracy. The network visualization is based on the cytoscape web library, which requires flash player for the representation of the network. All analyses are performed in real-time and a session for each user is ended after 1hr of inactivity. For high-performance, the database is implemented in Microsoft SQL Server 2012.

Conclusions and future directions

Protein interaction maps remains one of the major resources for the functional annotation of proteins. Embedding other lines of information with these maps can help researchers gain insights regarding the molecular contexts of the proteins. Here, we introduce TrypsNetDB, a web tool to consolidate the current knowledge on the interactome of the trypanosomatid parasites and dynamically integrate them with a wealth of available orthogonal information. We are continuously working on expanding the core, literature-derived, protein interaction depository of the database. Future plans also include providing supports for the remaining trypanosomatid parasites and the inclusion of other genome-wide data. TrypsNetDB is an open source effort, and hence, the code and databases are available through the portal. Moreover, the
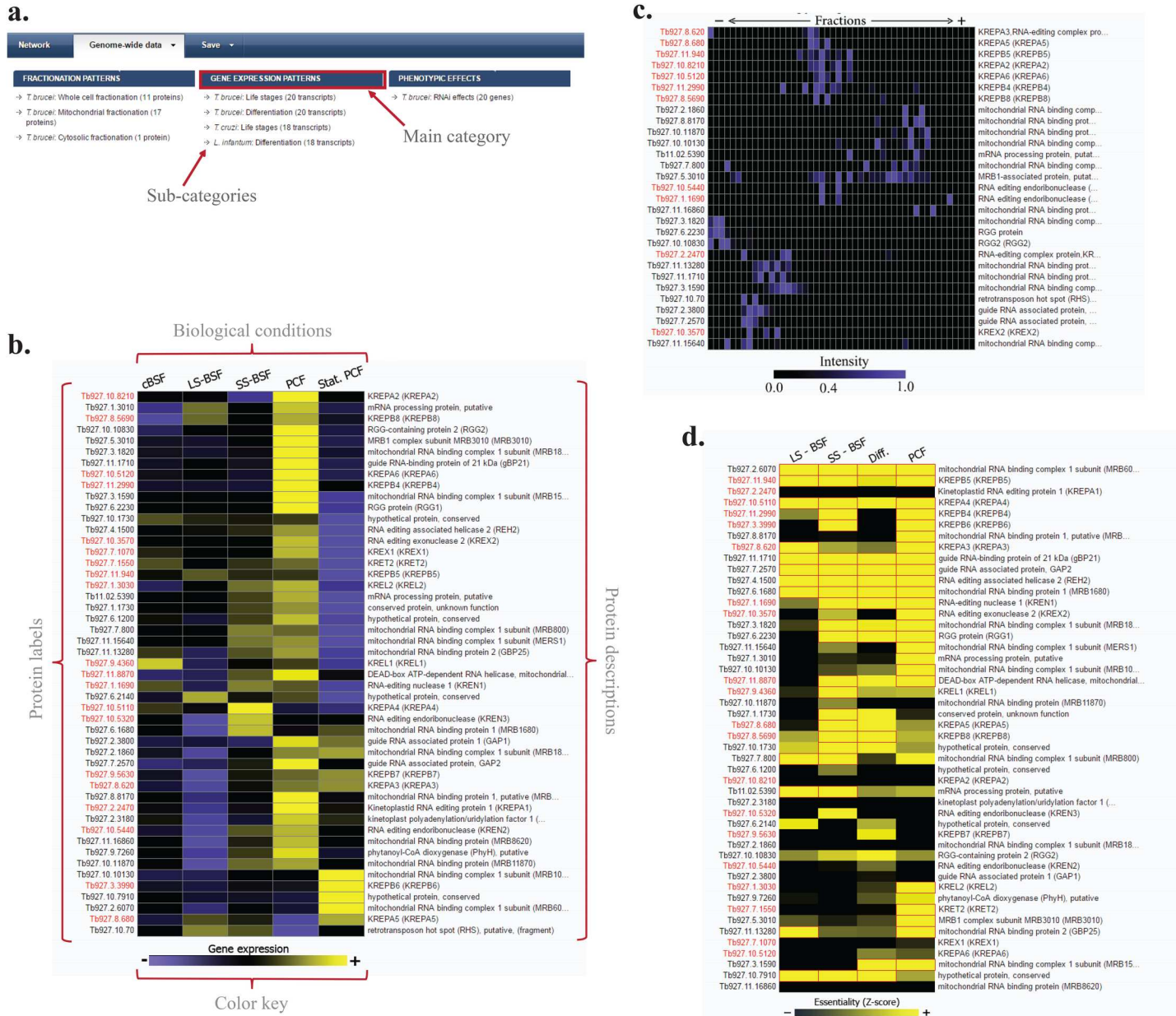
Fig 3. Genome-wide section of TrypsNetDB. (a) Genome-wide sections can be accessed from the menu located on the top of the interaction section. Users can visualize all relevant data by selecting a category or visualize a more specific dataset by choosing the corresponding sub-category. (b) A sample representation of transcriptome heatmap. Queried proteins are highlighted by red labels on the left of the heatmap, while suggested proteins by the database are shown with black labels. (c) A sample representation of a fractionation heatmap. (d) A sample representation of a gene essentiality heatmap. In each life stage (i.e., each column), statistically significant genes are represented by red borders.

doi:10.1371/journal.pntd.0005368.g003

interaction and fractionation data can be directly downloaded from the web interface using the provided links.

Author contributions

Conceptualization: VHG RS.

Data curation: VHG.

Formal analysis: VHG CWY NN NB.

Funding acquisition: RS.

Investigation: VHG CWY NN NB.

Methodology: VHG CWY NN NB.

Project administration: VHG RS.

Resources: VHG CWY NN NB.

Software: VHG.

Supervision: VHG RS.

Visualization: VHG.

Writing – original draft: VHG.

Writing – review & editing: VHG RS.

References

1. Stich A, Ponte-Sucre A, Holzgrabe U. Do we need new drugs against human African trypanosomiasis? Lancet Infect Dis. 2013; 13(9):733–4. doi: 10.1016/S1473-3099(13)70191-9 PMID: 23969207

2. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, et al. The genome of the African trypanosome Trypanosoma brucei. Science. 2005; 309(5733):416–22. doi: 10.1126/science.1112642 PMID: 16020726

3. El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, et al. Comparative genomics of trypanosomatid parasitic protozoa. Science. 2005; 309(5733):404–9. doi: 10.1126/science.1112181 PMID: 16020724

4. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, et al. The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease. Science. 2005; 309(5733):409–15. doi: 10.1126/science.1112631 PMID: 16020725

5. Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, et al. The genome of the kinetoplastid parasite, Leishmania major. Science. 2005; 309(5733):436–42. PubMed Central PMCID: PMCPMC1470643. doi: 10.1126/science.1112680 PMID: 16020728

6. Cestari I, Kalidas S, Monnerat S, Anupama A, Phillips MA, Stuart K. A multiple aminoacyl-tRNA synthetase complex that enhances tRNA-aminoacylation in African trypanosomes. Mol Cell Biol. 2013; 33 (24):4872–88. PubMed Central PMCID: PMCPMC3889560. doi: 10.1128/MCB.00711-13 PMID: 24126051

7. Freire ER, Malvezzi AM, Vashisht AA, Zuberek J, Saada EA, Langousis G, et al. Trypanosoma brucei translation initiation factor homolog EIF4E6 forms a tripartite cytosolic complex with EIF4G5 and a capping enzyme homolog. Eukaryot Cell. 2014; 13(7):896–908. PubMed Central PMCID: PMCPMC4135740. doi: 10.1128/EC.00071-14 PMID: 24839125

8. Gazestani VH, Nikpour N, Mehta V, Najafabadi HS, Moshiri H, Jardim A, et al. A Protein Complex Map of Trypanosoma brucei. PLoS Negl Trop Dis. 2016; 10(3):e0004533. PubMed Central PMCID: PMCPMC4798371. doi: 10.1371/journal.pntd.0004533 PMID: 26991453

9. Ammerman ML, Downey KM, Hashimi H, Fisk JC, Tomasello DL, Faktorova D, et al. Architecture of the trypanosome RNA editing accessory complex, MRB1. Nucleic Acids Res. 2012; 40(12):5637–50. PubMed Central PMCID: PMCPMC3384329. doi: 10.1093/nar/gks211 PMID: 22396527

10. Aphasizheva I, Maslov D, Wang X, Huang L, Aphasizhev R. Pentatricopeptide repeat proteins stimulate mRNA adenylation/uridylation to activate mitochondrial translation in trypanosomes. Mol Cell. 2011; 42 (1):106–17. PubMed Central PMCID: PMCPMC3073060. doi: 10.1016/j.molcel.2011.02.021 PMID: 21474072

11. Hernandez A, Madina BR, Ro K, Wohlschlegel JA, Willard B, Kinter MT, et al. REH2 RNA helicase in kinetoplastid mitochondria: ribonucleoprotein complexes and essential motifs for unwinding and guide RNA (gRNA) binding. J Biol Chem. 2010; 285(2):1220–8. PubMed Central PMCID: PMCPMC2801250. doi: 10.1074/jbc.M109.051862 PMID: 19850921

12. Ridlon L, Skodova I, Pan S, Lukes J, Maslov DA. The importance of the 45 S ribosomal small subunit-related complex for mitochondrial translation in Trypanosoma brucei. J Biol Chem. 2013; 288 (46):32963–78. PubMed Central PMCID: PMCPMC3829147. doi: 10.1074/jbc.M113.501874 PMID: 24089529

13. Acestor N, Zikova A, Dalley RA, Anupama A, Panigrahi AK, Stuart KD. Trypanosoma brucei mitochondrial respiratome: composition and organization in procyclic form. Mol Cell Proteomics. 2011; 10(9): M110 006908. PubMed Central PMCID: PMCPMC3186196.

14. Li Z, Wang CC. Functional characterization of the 11 non-ATPase subunit proteins in the trypanosome 19 S proteasomal regulatory complex. J Biol Chem. 2002; 277(45):42686–93. doi: 10.1074/jbc.M207183200 PMID: 12213827

15. Luz Ambrosio D, Lee JH, Panigrahi AK, Nguyen TN, Cicarelli RM, Gunzl A. Spliceosomal proteomics in Trypanosoma brucei reveal new RNA splicing factors. Eukaryot Cell. 2009; 8(7):990–1000. PubMed Central PMCID: PMCPMC2708463. doi: 10.1128/EC.00075-09 PMID: 19429779

16. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013; 41 (Database issue):D808–15. PubMed Central PMCID: PMCPMC3531103. doi: 10.1093/nar/gks1094 PMID: 23203871

17. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, et al. The IntAct molecular interaction database in 2012. Nucleic Acids Res. 2012; 40(Database issue):D841–6. PubMed Central PMCID: PMCPMC3245075. doi: 10.1093/nar/gkr1088 PMID: 22121220

18. Zuberi K, Franz M, Rodriguez H, Montojo J, Lopes CT, Bader GD, et al. GeneMANIA prediction server 2013 update. Nucleic Acids Res. 2013; 41(Web Server issue):W115–22. PubMed Central PMCID: PMCPMC3692113. doi: 10.1093/nar/gkt533 PMID: 23794635

19. Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D, et al. The BioGRID interaction database: 2015 update. Nucleic Acids Res. 2015; 43(Database issue):D470–8. PubMed Central PMCID: PMCPMC4383984. doi: 10.1093/nar/gku1204 PMID: 25428363

20. Bader GD, Betel D, Hogue CW. BIND: the Biomolecular Interaction Network Database. Nucleic Acids Res. 2003; 31(1):248–50. PubMed Central PMCID: PMCPMC165503. PMID: 12519993

21. Ruepp A, Waegele B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, et al. CORUM: the comprehensive resource of mammalian protein complexes—2009. Nucleic Acids Res. 2010; 38(Database issue):D497–501. PubMed Central PMCID: PMCPMC2808912. doi: 10.1093/nar/gkp914 PMID: 19884131

22. Jensen BC, Sivam D, Kifer CT, Myler PJ, Parsons M. Widespread variation in transcript abundance within and across developmental stages of Trypanosoma brucei. BMC Genomics. 2009; 10:482. PubMed Central PMCID: PMCPMC2771046. doi: 10.1186/1471-2164-10-482 PMID: 19840382

23. Kabani S, Fenn K, Ross A, Ivens A, Smith TK, Ghazal P, et al. Genome-wide expression profiling of in vivo-derived bloodstream parasite stages and dynamic analysis of mRNA alterations during synchronous differentiation in Trypanosoma brucei. BMC Genomics. 2009; 10:427. PubMed Central PMCID: PMCPMC2753553. doi: 10.1186/1471-2164-10-427 PMID: 19747379

24. 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; 10:495. PubMed Central PMCID: PMCPMC2772864. doi: 10.1186/1471-2164-10-495 PMID: 19857263

25. Urbaniak MD, Martin DM, Ferguson MA. Global quantitative SILAC phosphoproteomics reveals differential phosphorylation is widespread between the procyclic and bloodstream form lifecycle stages of Trypanosoma brucei. J Proteome Res. 2013; 12(5):2233–44. PubMed Central PMCID: PMCPMC3646404. doi: 10.1021/pr400086y PMID: 23485197

26. Gunasekera K, Wuthrich D, Braga-Lagache S, Heller M, Ochsenreiter T. Proteome remodelling during development from blood to insect-form Trypanosoma brucei quantified by SILAC and mass spectrometry. BMC Genomics. 2012; 13:556. PubMed Central PMCID: PMCPMC3545838. doi: 10.1186/1471-2164-13-556 PMID: 23067041

27. Urbaniak MD, Guther ML, Ferguson MA. Comparative SILAC proteomic analysis of Trypanosoma brucei bloodstream and procyclic lifecycle stages. PLoS One. 2012; 7(5):e36619. PubMed Central PMCID: PMCPMC3344917. doi: 10.1371/journal.pone.0036619 PMID: 22574199

28. Kramer S, Queiroz R, Ellis L, Hoheisel JD, Clayton C, Carrington M. The RNA helicase DHH1 is central to the correct expression of many developmentally regulated mRNAs in trypanosomes. J Cell Sci. 2010; 123(Pt 5):699–711. PubMed Central PMCID: PMCPMC2823576. doi: 10.1242/jcs.058511 PMID: 20124414

29. Gazestani VH, Salavati R. Deciphering RNA Regulatory Elements Involved in the Developmental and Environmental Gene Regulation of Trypanosoma brucei. PLoS One. 2015; 10(11):e0142342. PubMed Central PMCID: PMCPMC4631447. doi: 10.1371/journal.pone.0142342 PMID: 26529602

30. Fadda A, Ryten M, Droll D, Rojas F, Farber V, Haanstra JR, et al. Transcriptome-wide analysis of trypanosome mRNA decay reveals complex degradation kinetics and suggests a role for co-transcriptional degradation in determining mRNA levels. Mol Microbiol. 2014; 94(2):307–26. PubMed Central PMCID: PMCPMC4285177. doi: 10.1111/mmi.12764 PMID: 25145465

31. Rochette A, Raymond F, Corbeil J, Ouellette M, Papadopoulou B. Whole-genome comparative RNA expression profiling of axenic and intracellular amastigote forms of Leishmania infantum. Mol Biochem Parasitol. 2009; 165(1):32–47. doi: 10.1016/j.molbiopara.2008.12.012 PMID: 19393160

32. Veitch NJ, Johnson PC, Trivedi U, Terry S, Wildridge D, MacLeod A. Digital gene expression analysis of two life cycle stages of the human-infective parasite, Trypanosoma brucei gambiense reveals differentially expressed clusters of co-regulated genes. BMC Genomics. 2010; 11:124. PubMed Central PMCID: PMCPMC2837033. doi: 10.1186/1471-2164-11-124 PMID: 20175885

33. Alsford S, Turner DJ, Obado SO, Sanchez-Flores A, Glover L, Berriman M, et al. High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. Genome Res. 2011; 21(6):915–24. PubMed Central PMCID: PMCPMC3106324. doi: 10.1101/gr.115089.110 PMID: 21363968

34. Haanstra JR, Kerkhoven EJ, van Tuijl A, Blits M, Wurst M, van Nuland R, et al. A domino effect in drug action: from metabolic assault towards parasite differentiation. Mol Microbiol. 2011; 79(1):94–108. doi: 10.1111/j.1365-2958.2010.07435.x PMID: 21166896

35. Minning TA, Weatherly DB, Atwood J 3rd, Orlando R, Tarleton RL. The steady-state transcriptome of the four major life-cycle stages of Trypanosoma cruzi. BMC Genomics. 2009; 10:370. PubMed Central PMCID: PMCPMC2907688. doi: 10.1186/1471-2164-10-370 PMID: 19664227

36. Nilsson D, Gunasekera K, Mani J, Osteras M, Farinelli L, Baerlocher L, et al. Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of Trypanosoma brucei. PLoS Pathog. 2010; 6(8):e1001037. PubMed Central PMCID: PMCPMC2916883. doi: 10.1371/journal.ppat.1001037 PMID: 20700444

37. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 2014; 42(Database issue):D199–205. PubMed Central PMCID: PMCPMC3965122. doi: 10.1093/nar/gkt1076 PMID: 24214961

38. 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. 2010; 38(Database issue):D457–62. PubMed Central PMCID: PMCPMC2808979. doi: 10.1093/nar/gkp851 PMID: 19843604

39. Gene Ontology C. Gene Ontology Consortium: going forward. Nucleic Acids Res. 2015; 43(Database issue):D1049–56. PubMed Central PMCID: PMCPMC4383973. doi: 10.1093/nar/gku1179 PMID: 25428369

40. Butter F, Bucerius F, Michel M, Cicova Z, Mann M, Janzen CJ. Comparative proteomics of two life cycle stages of stable isotope-labeled Trypanosoma brucei reveals novel components of the parasite's host adaptation machinery. Mol Cell Proteomics. 2013; 12(1):172–9. PubMed Central PMCID: PMCPMC3536898. doi: 10.1074/mcp.M112.019224 PMID: 23090971

41. Utter CJ, Garcia SA, Milone J, Bellofatto V. PolyA-specific ribonuclease (PARN-1) function in stage-specific mRNA turnover in Trypanosoma brucei. Eukaryot Cell. 2011; 10(9):1230–40. PubMed Central PMCID: PMCPMC3187051. doi: 10.1128/EC.05097-11 PMID: 21743004