Single-cell RNA-seq reveals trans-sialidase-like superfamily gene expression heterogeneity in *Trypanosoma cruzi* populations.

Lucas Inchausti[1,2], Lucía Bilbao[1,2], Vanina A. Campo[3,4], Joaquín Garat[5], José Sotelo-Silveira[5,6], Gabriel Rinaldi[7], Virginia M. Howick[8], María Ana Duhagon[2], Javier G. De Gaudenzi[3,4], Pablo Smircich[1,2]

1 - Laboratorio de Bioinformática, Departamento de Genómica, Instituto de Investigaciones Biológicas Clemente Estable, Montevideo, Uruguay

2 - Sección Genómica Funcional, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay

3 - Instituto de Investigaciones Biotecnológicas, Universidad Nacional de San Martín—Consejo Nacional de Investigaciones Científicas y Técnicas, General San Martín, Prov. de Buenos Aires, Argentina

4 - Escuela de Bio y Nanotecnologías (EByN), Universidad Nacional de San Martín, General San Martín, Prov. de Buenos Aires, Argentina

5 - Departamento de Genómica, Instituto de Investigaciones Biológicas Clemente Estable, Montevideo, Uruguay

6 - Sección Biología Celular, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay

7 - Department of Biology, University of Oxford, Oxford, United Kingdom

8 - Institute of Biodiversity, Animal Health, and Comparative Medicine, University of Glasgow, Glasgow, United Kingdom

Abstract

Trypanosoma cruzi, the causative agent of Chagas disease, presents a major public health challenge in Central and South America, affecting approximately 8 million people and placing millions more at risk. The *T. cruzi* life cycle includes transitions between epimastigote, metacyclic trypomastigote, amastigote, and blood trypomastigote stages, each marked by distinct morphological and molecular adaptations to different hosts and environments. Unlike other trypanosomatids, *T. cruzi* does not employ antigenic variation but instead relies on a diverse array of cell-surface-associated proteins encoded by large multi-copy gene families (multigene families), essential for infectivity and immune evasion.

This study analyzes cell-specific transcriptomes using single-cell RNA sequencing of amastigote and trypomastigote cells to characterize stage-specific surface protein expression during

mammalian infection. Through clustering and identification of cell-specific markers, we assigned cells to distinct parasite developmental forms. Analysis of individual cells revealed that surface protein-coding genes, especially members of the trans-sialidase TcS superfamily (TcS), are expressed with greater heterogeneity than single-copy genes. Additionally, no recurrent combinations of TcS genes were observed between individual cells in the population. Our findings thus reveal transcriptomic heterogeneity within trypomastigote populations where each cell displays unique TcS expression profiles. Focusing on the diversity of surface protein expression, this research aims to deepen our understanding of *T. cruzi* cellular biology and infection strategies.

Introduction

The protozoan parasite *Trypanosoma cruzi* is the etiological agent of Chagas disease, a highly prevalent infectious disease in Central and South America that affects ~8 million people, with several million more at risk of infection [1].

The life cycle of the parasite involves both an invertebrate vector (triatomine bug) and a vertebrate host. The change in environmental conditions triggers differentiation processes in the parasite developing across four main stages. The epimastigote form replicates within the insect and differentiates into metacyclic trypomastigotes in the insect's rectal tract. This latter form does not replicate, specializes in host infection, and can invade various cell types. Once inside the cell, the parasite differentiates into an amastigote, the replicative form within the mammalian host. After several rounds of division, amastigotes differentiate into bloodstream trypomastigotes, which, after cell lysis, can either infect new cells or be ingested by the vector [2]. Parasite developmental stages exhibit significant morphological and molecular differences. *T. cruzi* infection relies on a heterogeneous set of membrane proteins, encoded mainly by large multigene families [3]. Among these are trans-sialidases and trans-sialidase-like proteins (TcS), Mucins, MASPs, GP63, DGFs, and RHS proteins, most of which are involved in infection, tropism, and immune evasion [4].

The trans-sialidase (TcS) superfamily is involved in processes underlying host-parasite interactions [5]. Members that retain enzymatic activity catalyze the transfer of sialyl groups from host glycoconjugates to galactopyranosyl units on the parasite’s surface, an essential activity, as *T. cruzi* is unable to synthesize *de novo* sialic acid [6]. This is the largest superfamily, with over 1,400 members, and is subdivided into eight groups based on amino acid sequence [5]. Few members have been functionally characterized, most expressed primarily in mammalian stages [7]. The TcS group I include proteins that retain TS activity, though members from all groups are involved in host-parasite interactions [8].

In recent years, single-cell RNA sequencing (scRNA-seq) has been employed in protozoa, with reports including *Plasmodium falciparum*, *Trypanosoma brucei*, and *Leishmania major* [9-18].

These studies revealed key aspects of the infection process undetectable with conventional methods, highlighting the relevance of this approach for understanding individual variation in gene expression in single-celled organisms [19]. In *T. brucei*, antigenic variation driven by variant surface glycoproteins (VSGs) has been studied at single-cell resolution to understand the mechanisms that enable subpopulations of this parasite to evade the immune system, as a bet hedging strategy, ensuring parasite survival [14]. In this parasite, scRNA-seq revealed that pre-metacyclic cells express multiple VSG transcripts simultaneously, in contrast to metacyclic forms, which display a protein coat composed of a single VSG type.

While population-level heterogeneity in surface protein expression has been suggested as critical for *T. cruzi* infection, immune evasion, and persistence [20, 21], this has not been studied at the intra-population level, and the underlying mechanisms remain poorly understood. Here, we present a scRNA-seq analysis of *T. cruzi*, to understand the heterogeneity in surface protein expression within trypomastigote populations.

Results and Discussion

Identification of cell populations

To assess the expression of cell-surface protein-coding genes in *Trypanosoma cruzi*, we conducted a 10X Chromium Single Cell 3' assay from a mixed population of amastigotes and cell-derived trypomastigotes, aiming at sequencing the transcriptome of 5000 cells. After low-quality cell filtering and gene expression quantification (see Materials and Methods), we obtained 3192 single-cell transcriptomes with 14321 total genes detected, with a mean of 1088 genes and 2461 UMIs detected per cell. These results were comparable to other scRNA-seq studies done in *Trypanosoma brucei* and recently reported as a preprint in *T. cruzi* using 10X Chromium technology [15, 22, 23]. Cell populations (**Figure 1a**) were defined by identifying cluster-specific gene markers (**Figure 1b, Supplementary File 1**). Two cell clusters were assigned to trypomastigotes and amastigotes (cluster 0 and 1, respectively). Markers gene expression was consistent with previous bulk RNA-seq data from Dm28c [24] (**Figures 1b, 1c and 1d, Supplementary Table 1**). In turn, we hypothesize that cluster 2 reflects amastigote-trypomastigote transitioning parasites, as its gene markers are differentially expressed in bulk RNA-seq data, but some are upregulated in amastigotes and others in trypomastigotes.

Figure 1. (a) UMAP colored by detected clusters based on gene expression profiles, (b) Heatmap of the top 10 gene markers upregulated in each of the 3 cell populations identified (log2FC > 1 and adjusted p-value < 0.05), (c) Expression of a cluster 0 marker gene (C4B63-16g183) on the UMAP, and (d) Expression of a cluster 1 marker gene (C4B63-16g155) on the UMAP.

Expression pattern of surface protein-coding genes

We analyzed differences in gene expression between single-copy genes and multigene families, between and within the identified cell populations. As expected, single-copy gene expression is high in both amastigotes and trypomastigotes and similar on average between both cell types (**Supplementary Figure 1a**), while surface protein-coding genes are more expressed in trypomastigotes [25] (**Figure 2a**).

Within trypomastigotes, heterogenous expression of surface gene families was observed, with high variation in the number of cells in which each surface gene was detected, as well as the total expression level in each cell (**Figure 2b**). Even though expression heterogeneity is also observed for single-copy genes (**Supplementary Figure 1a and 1b**), probably due to the sampling biases that cause gene dropout in 10X Genomics technology, we investigated whether this phenomenon was more pronounced in surface proteins. Therefore, we analyzed differences among single-copy and multigene family genes (together or grouped by multigene family) in trypomastigotes, in terms of the number of cells expressing each of the individual genes of each group (**Figure 2c and Supplementary Figure 1c**). We additionally calculated their Gini index (**Figure 2d and Supplementary Figure 1d**) [26] to assess the extent to which the expression levels of a given gene were evenly distributed across individual cells. Ribosomal protein-coding genes were included as a control group for homogeneous expression.

Interestingly, compared to single-copy genes, especially ribosomal genes, multigene family genes showed a greater dispersion regarding the number of cells in which each gene was detected (**Figure 2c and Supplementary Figure 1c**), and a higher value of its Gini indexes with a lower dispersion (**Figure 2d and Supplementary Figure 1d**). Both observations indicate a higher expression heterogeneity for surface protein expression in the trypomastigote population. Moreover, when different families were analyzed separately, TcS genes showed a particularly high heterogeneity (**Supplementary Figure 1c and 1d**).

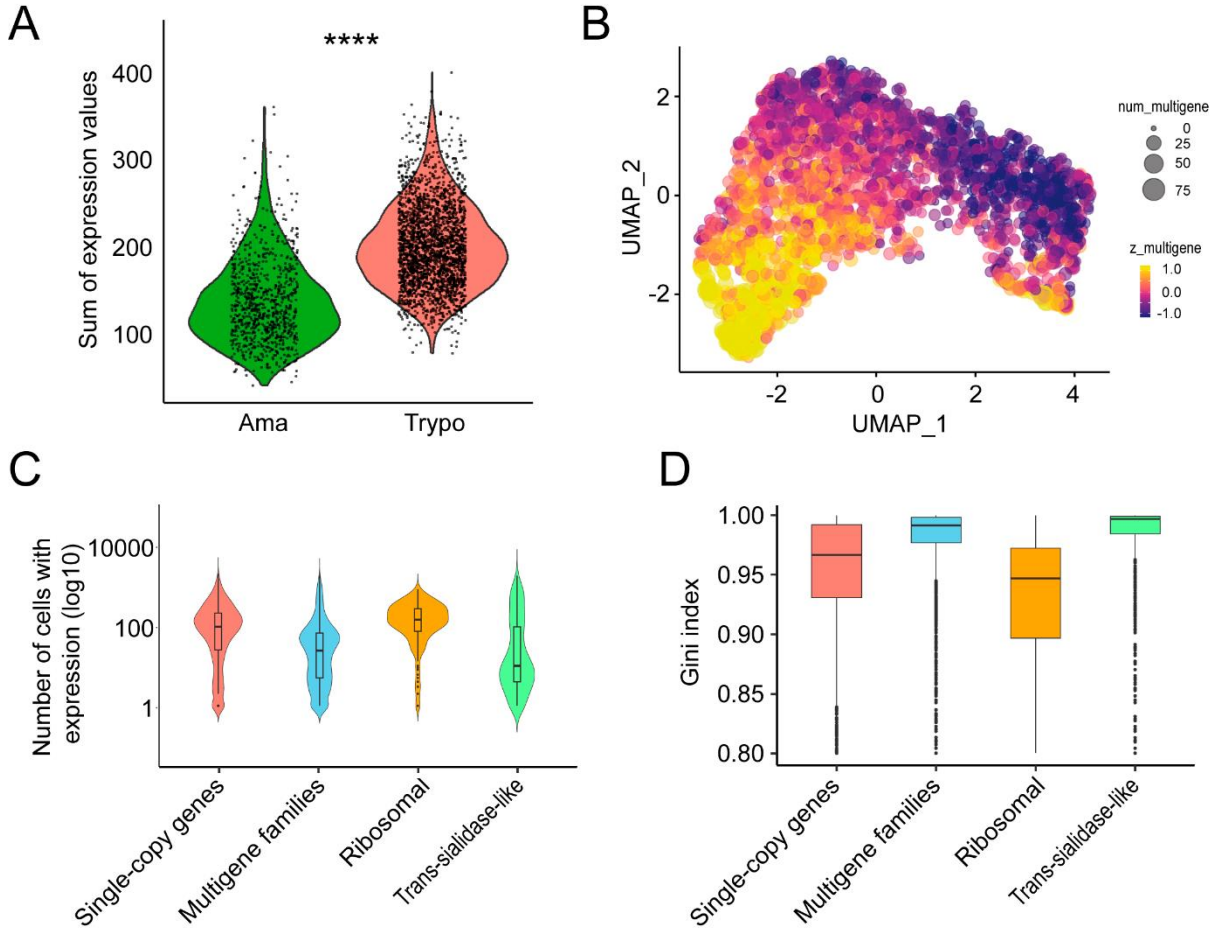


Figure 2. (a) Summatory of expression levels values from all genes belonging to the disruptive compartment for each cell from amastigote (Cluster 1) and trypomastigote (Cluster 0) cell populations, (b) UMAP visualization of the expression patterns of multigene family genes. (c) Violin plots showing the number of cells expressing a specific gene belonging to each group of genes: subsampled single-copy and multigene families, ribosomal genes and trans-sialidases. To avoid biases against size differences between compartments, we generated a subsampled single-copy genes list, randomly selecting an equal number of genes as those from the multigene family's gene set. The expression distribution of the subsampled single-copy genes is similar to the distribution of the entire dataset. (d) Boxplots depict the Gini index, across single-copy and multigene families, ribosomal genes and trans-sialidases. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. For (c) and (d), all comparisons of the multigene families against single-copy and ribosomal genes are statistically significant.

Expression pattern of the TcS superfamily

When we re-clustered the trypomastigote population based solely on trans-sialidase gene expression, we identified two sub-populations: trypomastigote cluster 0 ("trypo_0" composed of 1186 cells), which over-expressed these genes compared to trypomastigote cluster 1 ("trypo_1" composed of 1015 cells) (**Figure 3**).

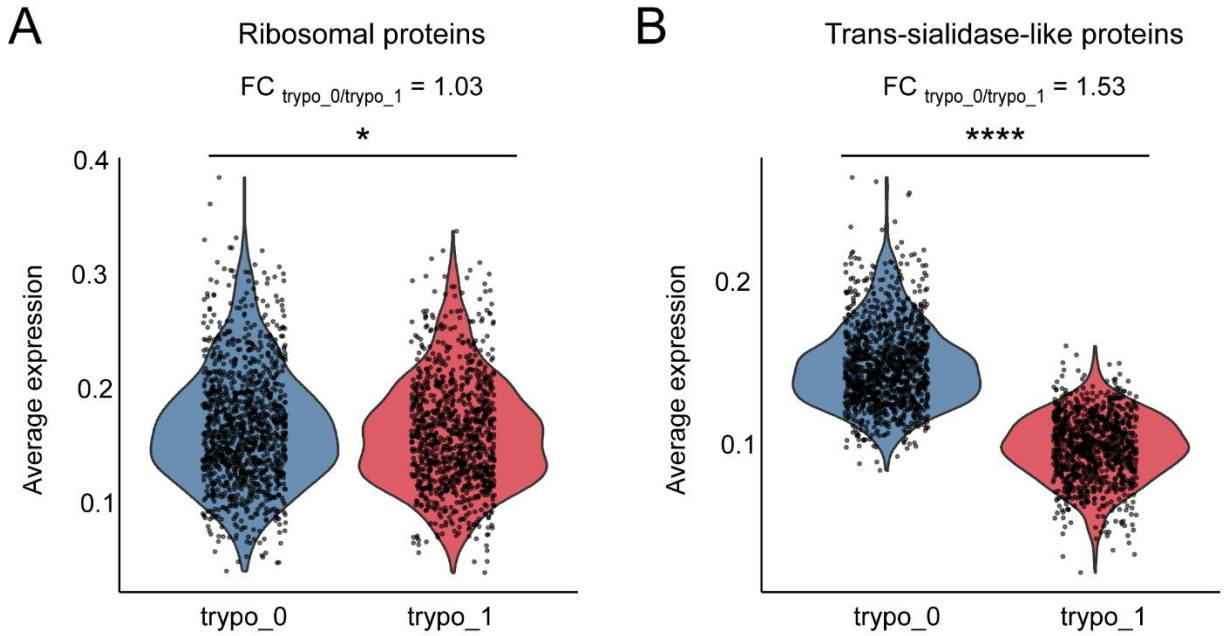


Figure 3. Trypomastigote sub-populations identified based on trans-sialidase expression profiles. (a) violin plot displaying average expression levels of ribosomal protein-coding genes across sub-populations. (b) violin plot showing combined trans-sialidase expression levels for each sub-population. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

The two trypomastigote sub-populations segregated by TcS expression, show only slight differences in the expression of other gene categories, such as ribosomal protein-coding genes (FC = 1.03), transporters (FC = 1.09), polymerase-related genes (FC = 1.15), and phosphatases (FC = 1.10) (Supplementary Figure 2). Although there is a tendency for surface protein genes to be more expressed in cluster trypo_0 (FC = 1.38), trans-sialidases displayed the highest fold change between the two trypomastigote subpopulations (FC = 1.53) (**Supplementary Figure 2)**. It is tempting to speculate that this may reflect different infectivity amongst trypomastigote subpopulations, consistent with reports of "broad" and "slender" forms [27]. Recently, while this manuscript was being prepared, a similar approach by Laidlaw et al. described the phenomenon by clustering trypomastigote cells based on the expression of all genes [23]. When applying this strategy to our data, their observation was reproduced, further demonstrating the consistent reproducibility of the scRNA-seq results across studies.

When all cells are considered, most TcS genes are expressed, but in each individual cell only approximately 40 TcS genes were detected (**Figure 4d upper pannel**). Interestingly, we observed that both trypomastigote subpopulations contain a subgroup of TcS genes that are detected in a large portion of cells (>40%) (**Figure 4a, Supplementary Table 2**), indicating high-level expression at the population level. However, these genes did not account for most of the family´s expression in each individual cell. Even family members detected in only a few cells significantly contribute to the total TcS expression in those cells, being quantified at similar levels as those from the frequently detected group (**Figure 4b, 4c and 4d**). Gene dropouts might favor random patterns of gene family's detection in scRNA-seq experiments, particularly affecting genes with low expression. Since genes detected in a high percentage of cells are not more highly expressed than those detected in only a few cells (**Figure 4b**), we do not believe that this characteristic of scRNA-seq data is responsible for our observation. Actually, at the single-cell level, each detected TcS gene contributes similarly to the overall family expression in those cells, as shown by a small average Gini index for the percentage of the expression of the family in each cell (**Figure 4c and 4d**). This highlights a key limitation of bulk RNA-seq, as it may wrongly indicate that a set of few genes are highly expressed in each cell, while in fact, TcS expression is evenly distributed among all detected family members, regardless of the number of cells expressing each gene. Indeed, most TcS members detected in more than 40% of cells in this work were found in the top 100 expressed TcS in an independent bulk RNA-seq study (**Supplementary Figure 3a**).

When analyzing each trypomastigote subpopulation, no coordinated expression among specific TcS members was observed, as no subclusters of cells were identified based on TcS detection profiles. Even when clustering was restricted to genes detected in more than 40% of cells, no clear subclusters of cells were identified (**Supplementary Figure 3b**).

The loci for TcS genes and other gene families of surface proteins are mostly grouped in specific genomic compartments [28] that are regulated epigenetically by the activation or silencing of chromatin folding domains [24]. The TcS that are detected in a high percentage of cells are mostly dispersed throughout the genome (**Supplementary Table 2**). This suggests that their preferential expression is likely not due to colocalization in one or a few ubiquitous activated chromatin-folding domains. These observations raise intriguing questions about the mechanisms governing the selection of TcS members expressed in each cell, warranting further investigation.

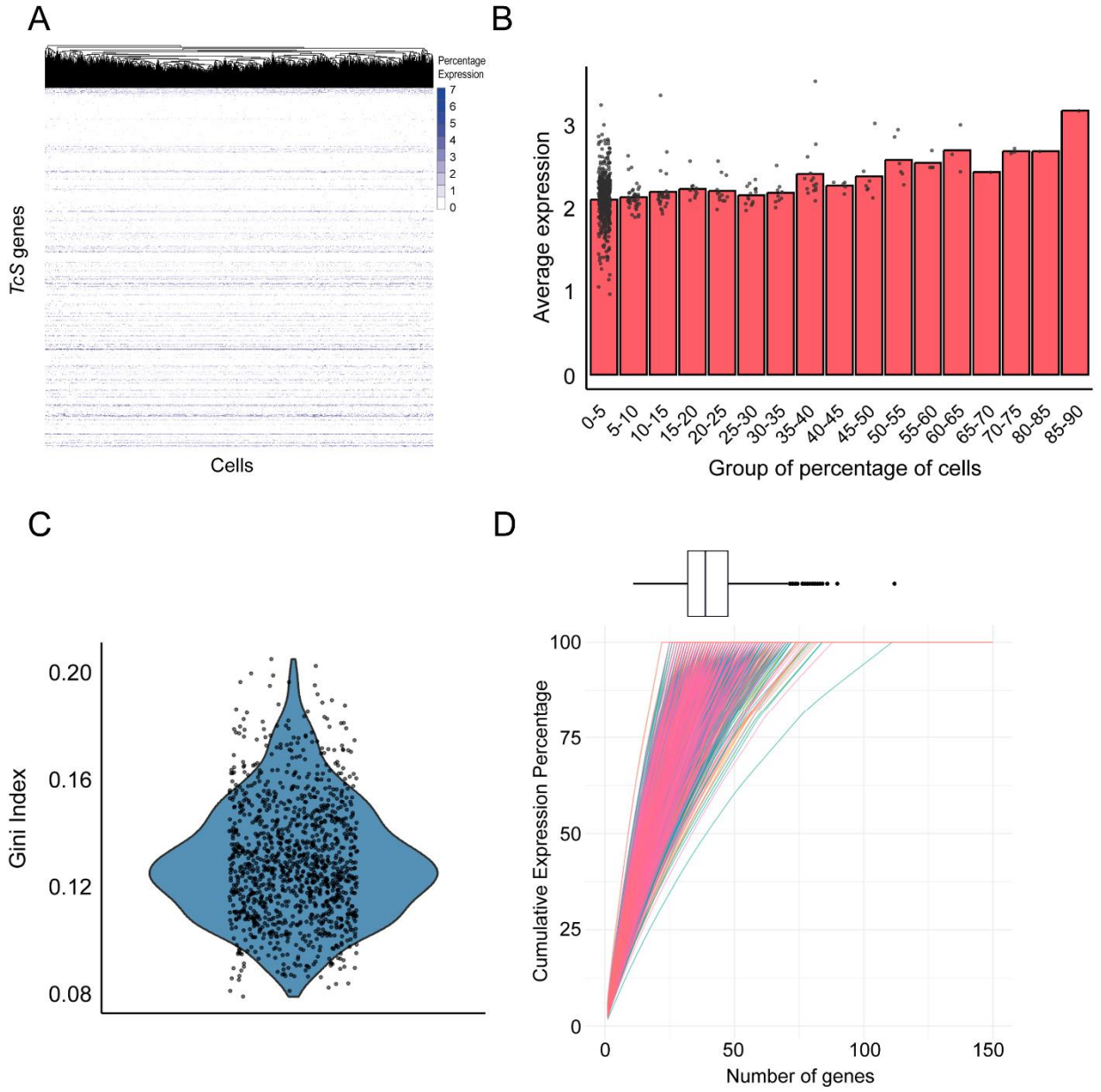


Figure 4. (a) Heatmap displaying the expression of TcS genes in each cell that together account for 75% of total TcS gene expression within cluster trypo 0. Cells are clustered by TcS expression profiles, with colors representing each gene's percentage contribution to the cell's total TcS expression. (b) Average expression of TcS genes grouped by the percentage of cells expressing each gene, (c) Gini index distribution for trypomastigotes cluster 0 cells considering only TcS detected in each cell. (d) Cumulative plot showing the percentage of TcS expression in each cell from cluster trypo_0. Genes are ordered by descending expression level and sequentially added to obtain each value on the Y-axis. The observed linear increase suggests a relatively equal contribution of each gene to the overall expression.

Final remarks

The expression of surface protein-coding genes varied between cell stages. Genes that belong to the trans-sialidase-like superfamily, involved in host-parasite interactions, displayed significant heterogeneity in expression levels within trypomastigotes. This variation in TcS expression suggests that not all trypomastigotes express these genes uniformly. The lack of coordinated expression among TcS members, observed in single-cell and bulk RNA-seq studies [3], suggests a complex regulatory mechanism that enables adaptability in different *T. cruzi* subpopulations. These findings reinforce the importance of TcS variability in driving phenotypic diversity and pathogenicity in *T. cruzi*. Taken together, these findings not only demonstrate the sensitivity of scRNA-seq in delineating parasite life stages and their associated gene expression profiles, but also reveal a complex regulation of surface proteins and TcS family members, with tentative implications for immune evasion and pathogenicity.

Materials and Methods

Trypanosoma cruzi and mammalian cell culture

Epimastigote forms of *Trypanosoma cruzi* strain Dm28c were derived from axenic cultures cultivated in Brain-Heart Infusion medium (BHI, Oxoid) supplemented with 10% heat-inactivated fetal bovine serum (FBS, Capricorn), penicillin (100 units/mL) and streptomycin (100 µg/mL) as described [29]. Cultures were diluted 1/10 with fresh BHI medium every 3 days and maintained at 28 °C.

Myoblast rat cell line H9c2 (ATCC CRL-1446) was maintained in hgDMEM medium (Gibco) supplemented with 10% heat-inactivated fetal bovine serum, penicillin (100 units/mL) and streptomycin (100 µg/mL) at 37 C in a humidified 5% CO2 incubator. Confluent cells were washed with 1X phosphate-buffered saline (1X PBS), incubated for 5 min with trypsin-EDTA (Gibco), diluted with culture medium and re-plated for maintenance.

Mycoplasma contamination in cell lines was regularly monitored using MycoAlert® Mycoplasma Detection Kit (Lonza), following the manufacturer's protocol.

Isolation and purification of cellular trypomastigotes and intracellular amastigotes

Late stationary phase epimastigotes of Dm28c strain were used to infect H9c2 cells for a primary infection. Six days post-infection, cellular trypomastigotes were obtained from the supernatant and were used to infect 50% confluent H9c2 cells at a 10:1 rate. Twenty-four hours post-infection the cell culture was washed twice with PBS 1X and maintained with fresh hgDMEM at 37 C in a humidified 5% CO2 incubator.

For amastigote purification, H9c2 cells grown in monolayers and infected for 48 hours were washed with 1X PBS and incubated with trypsin-EDTA for 5 minutes at 37°C. The trypsinization was stopped by adding an equal volume of hgDMEM with 10% FBS. The cell

suspension was repeatedly passed through a 27-gauge needle attached to a 30-mL syringe until complete cell disruption was confirmed under the microscope. The supernatant, containing free amastigotes, was collected and centrifuged at 500xg for 10 minutes at 4°C to remove large host-cell debris. The resulting supernatant was then centrifuged at 4,000xg for 10 minutes at 4°C, and the amastigote-containing pellet was washed twice in chilled 1X DPBS (Dulbecco's Phosphate-Buffered Saline, No Calcium, No Magnesium) and resuspended in 1X DPBS at 200 µL per $1x10^6$ cells, ready for the fixation step.

Cellular trypomastigotes derived from infected H9c2 cells and present in the cell supernatant fraction were collected and centrifuged at 500xg for 10 minutes at 4°C to remove large host-cell debris. The washing and resuspension steps in DPBS were performed for amastigotes as described above.

scRNA-seq library preparation and sequencing

Cell fixation was performed using the Methanol Fixation Protocol for Single-Cell RNA Sequencing [30] as recommended by 10X Genomics technical support. Briefly, chilled 100% Methanol (for HPLC, ≥99.9%, Millipore) was added drop by drop ($1x10^6$ cells in 800 µl) and incubated at -20°C for 30 min. For rehydration, fixed cells were first equilibrated at 4°C and then centrifuged at 4,000xg for 5 min at 4°C, the supernatant was discarded and Wash-Resuspension Buffer (3X SSC in Nuclease-free Water, 0.04% UltraPure Bovine Serum Albumin, 1mM DTT, and 0.2 U/ml RNase Inhibitor) was added to the pellet. Cell debris and large clumps were eliminated by passing the sample through a 40 µm Flowmi Cell Strainer.

10X Genomics library preparation was performed at the service provided by the Instituto de Biología y Medicina Experimental (IBYME, Argentina). The library was sequenced by a service provider (Macrogen, Korea) in a HiSeq2500 equipment (two lanes), generating approx. 880 million reads. Raw data is available in the SRA (https://www.ncbi.nlm.nih.gov/sra/) under BioProject PRJNA1200704.

Transcript quantification

T. cruzi 2018 Dm28c genome [28] (release 62, TriTrypDB) and *T. cruzi* Dm28c maxi circle kDNA sequence [31] were combined and used as reference. To improve the proportion of reads assigned to genes, 11,362 3′UTR regions of the coding sequences (CDS) were annotated using peaks2UTR [32]. Gene expression quantification was performed by pseudoalignment using kallisto bustools [33], with the options --filter to remove potential noise from environmental RNA and --em to apply the Expectation-Maximization (EM) algorithm. This algorithm outperforms other mapping software in handling multimapping reads, enabling more accurate quantification of multigene families [33].

scRNA-seq data processing and analysis

Count matrices from two technical replicates obtained by sequencing the same library were merged: the common barcodes across both datasets were retained, and the count matrices were

combined to generate a unified dataset. Subsequently, the following metrics were calculated: nUMI, nGene, and mitoRatio, as well as the log10 of genes per UMI (log10GenesperUMI). A filtering criterion was applied, retaining cells with nUMI > 1200, nGene > 100, log10GenesperUMI > 0.8, and mitoRatio < 0.1. Additionally, ribosomal genes were excluded from subsequent analyses. Data normalization and scaling were performed using the Seurat R version 5 package [34], employing NormalizeData and ScaleData functions.

FindNeighbors function was used to construct a k-nearest neighbors graph of cells using 10 principal components (PCs) and Louvain algorithm was employed for cell clustering using FindClusters function.

To define marker genes, the FindAllMarkers function from Seurat package was used, selecting those with an adjusted p-value < 0.05 and a log2 fold change (log2FC) > 1. To validate these markers, bulk RNA-seq data of Dm28c was incorporated (NCBI BioProject ID PRJNA850400 [24]). Transcripts were quantified using Kallisto [35] (with -b 100 option to perform 100 bootstraps), followed by differential expression analysis conducted with Sleuth [36]. Genes with an adjusted p-value < 0.05 and |log2FC| > 0.25 were filtered for further analysis. Finally, the gene IDs of the Seurat-defined markers were cross-referenced with the IDs of the differentially expressed genes obtained from the bulk RNA-seq analysis to corroborate stage-specific gene expression. For 2D visualization of cell clusters and gene markers expression profiles across cells, UMAP projection was employed [37].

Gene IDs corresponding to multigene families (trans-sialidase and trans-sialidase-like, MASP, Mucins, GP63, RHS and DGF) were obtained by text searches using the current genome annotation on the “description” field. Single and low copy number genes were defined as those that did not belong to the latter gene list. For simplicity “single-copy” will be used to refer to these genes throughout the manuscript.

Data processing was conducted using R version 4.2.0. Statistical analyses were performed using the Wilcoxon rank-sum test and p-values < 0.05 were considered statistically significant.

To study expression inequality, we applied the Gini index, a metric originally developed in the field of economics [26]. The Gini index serves as a measure of expression heterogeneity, and we employed it in two distinct contexts: first, to evaluate the degree of inequality in the distribution of a gene's expression levels across individual cells (**Figure 2d and Supplementary Figure 1d**); and second, to assess the extent to which a given cell expresses individual genes at varying rates (**Figure 4c**).

Author Contributions

LI: Data Curation, Formal Analysis, Investigation, Methodology, Visualization, Writing-Original Draft Preparation, Writing-Review & Editing; LB: Methodology, Investigation, Writing-Review

& Editing; VAC: Methodology, Investigation, Writing-Review & Editing; JG: Methodology, Formal Analysis, Writing-Review & Editing; GR: Methodology, Writing-Review & Editing; VMH: Methodology, Writing-Review & Editing; MAD: Methodology, Resources, Writing-Review & Editing; JSS: Methodology, Resources, Writing-Review & Editing; JDG: Conceptualization, Methodology, Visualization, Writing-Original Draft Preparation, Writing-Review & Editing; PS: Conceptualization, Funding Acquisition, Methodology, Formal Analysis, Project Administration, Resources, Supervision, Validation, Writing-Original Draft Preparation, Writing-Review & Editing

Conflict of interest

The authors declare that they have no conflicts of interest

Financial Support

This project was supported by: CSIC, Universidad de la República, grant number: I+D-2020-505 awarded to PS; LI, LB, JG, MD, JSS and PS received financial support from PEDECIBA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

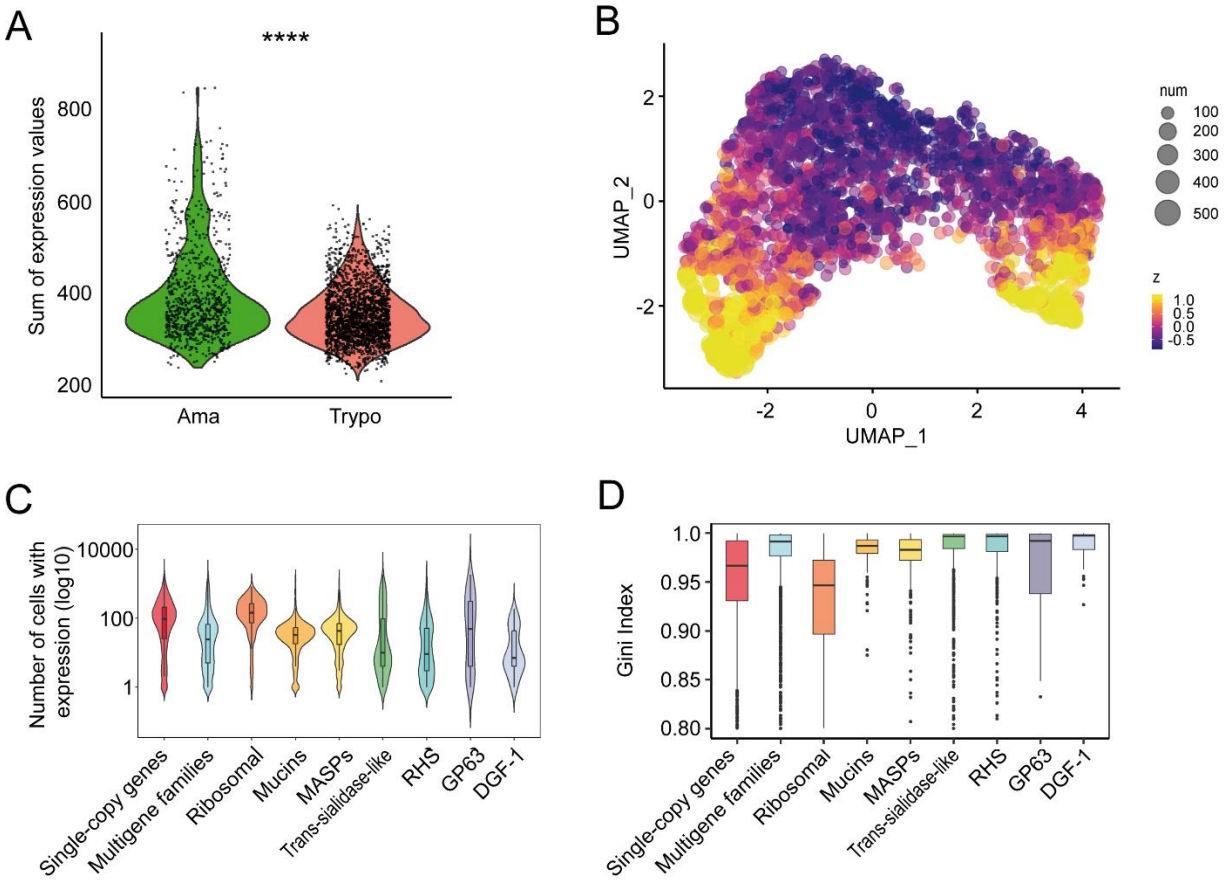
References

1. Tarleton, R.L., *Chagas Disease: A Solvable Problem, Ignored.* Trends Mol Med, 2016. **22**(10): p. 835-838.
2. Martin-Escolano, J., et al., *An Updated View of the Trypanosoma cruzi Life Cycle: Intervention Points for an Effective Treatment.* ACS Infect Dis, 2022. **8**(6): p. 1107-1115.
3. Belew, A.T., et al., *Comparative transcriptome profiling of virulent and non-virulent Trypanosoma cruzi underlines the role of surface proteins during infection.* PLoS Pathog, 2017. **13**(12): p. e1006767.
4. Pech-Canul, A.C., V. Monteon, and R.L. Solis-Oviedo, *A Brief View of the Surface Membrane Proteins from Trypanosoma cruzi.* J Parasitol Res, 2017. **2017**: p. 3751403.
5. Freitas, L.M., et al., *Genomic analyses, gene expression and antigenic profile of the trans-sialidase superfamily of Trypanosoma cruzi reveal an undetected level of complexity.* PLoS One, 2011. **6**(10): p. e25914.
6. Burle-Caldas, G.A., et al., *Disruption of Active Trans-Sialidase Genes Impairs Egress from Mammalian Host Cells and Generates Highly Attenuated Trypanosoma cruzi Parasites.* mBio, 2022: p. e0347821.
7. Berna, L., et al., *Transcriptomic analysis reveals metabolic switches and surface remodeling as key processes for stage transition in Trypanosoma cruzi.* PeerJ, 2017. **5**: p. e3017.

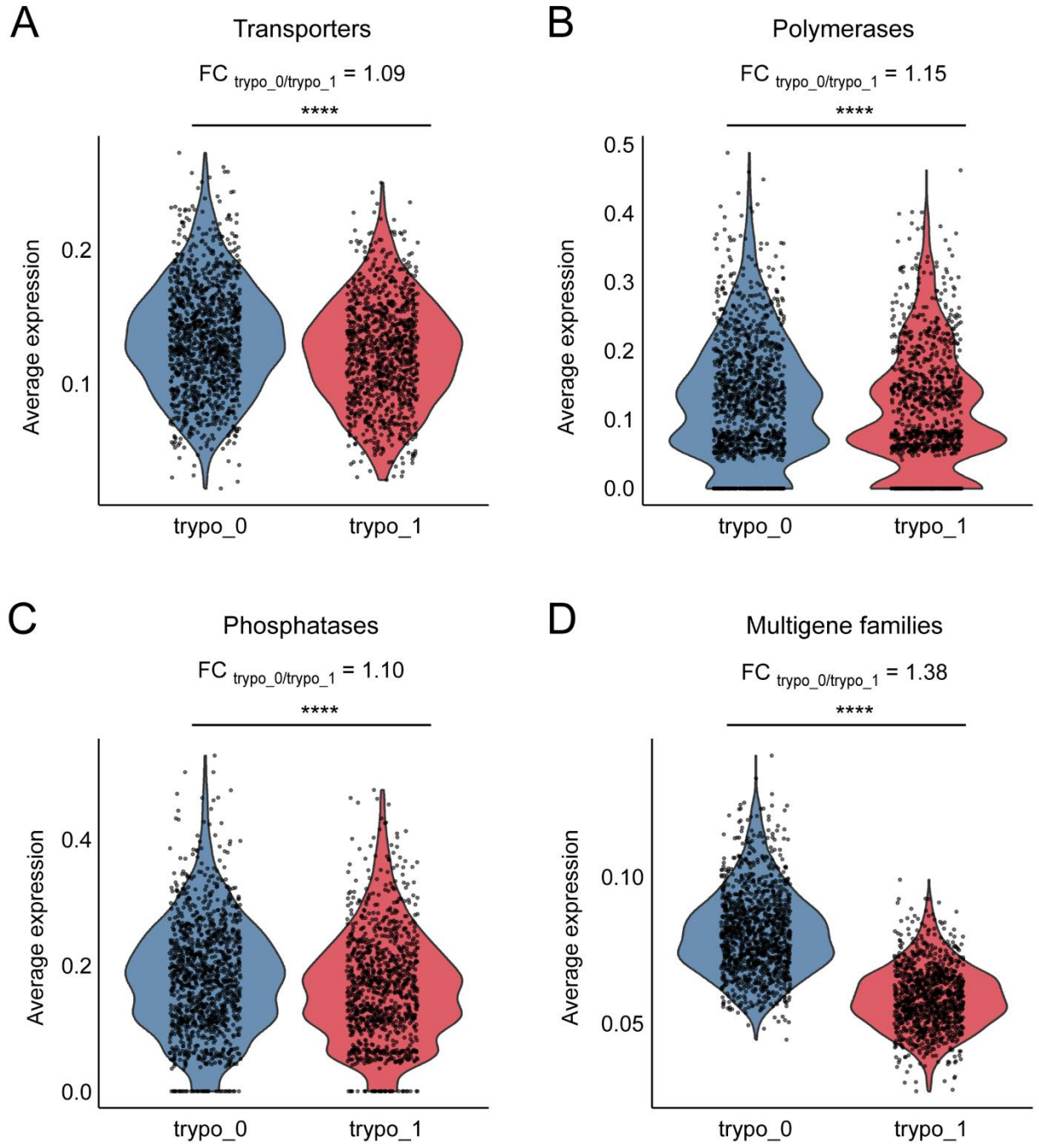
8. Tonelli, R.R., et al., *Role of the gp85/trans-sialidases in Trypanosoma cruzi tissue tropism: preferential binding of a conserved peptide motif to the vasculature in vivo.* PLoS Negl Trop Dis, 2010. **4**(11): p. e864.
9. Walzer, K.A., et al., *Single-Cell Analysis Reveals Distinct Gene Expression and Heterogeneity in Male and Female Plasmodium falciparum Gametocytes.* mSphere, 2018. **3**(2).
10. Ngara, M., et al., *Exploring parasite heterogeneity using single-cell RNA-seq reveals a gene signature among sexual stage Plasmodium falciparum parasites.* Exp Cell Res, 2018. **371**(1): p. 130-138.
11. Reid, A.J., et al., *Single-cell RNA-seq reveals hidden transcriptional variation in malaria parasites.* Elife, 2018. **7**.
12. Howick, V.M., et al., *The Malaria Cell Atlas: Single parasite transcriptomes across the complete Plasmodium life cycle.* Science, 2019. **365**(6455).
13. Vigneron, A., et al., *Single-cell RNA sequencing of Trypanosoma brucei from tsetse salivary glands unveils metacyclogenesis and identifies potential transmission blocking antigens.* Proc Natl Acad Sci U S A, 2020. **117**(5): p. 2613-2621.
14. Hutchinson, S., et al., *The establishment of variant surface glycoprotein monoallelic expression revealed by single-cell RNA-seq of Trypanosoma brucei in the tsetse fly salivary glands.* PLoS Pathog, 2021. **17**(9): p. e1009904.
15. Briggs, E.M., et al., *Single-cell transcriptomic analysis of bloodstream Trypanosoma brucei reconstructs cell cycle progression and developmental quorum sensing.* Nat Commun, 2021. **12**(1): p. 5268.
16. Howick, V.M., et al., *Single-cell transcriptomics reveals expression profiles of Trypanosoma brucei sexual stages.* PLoS Pathog, 2022. **18**(3): p. e1010346.
17. Quintana, J.F., et al., *Single cell and spatial transcriptomic analyses reveal microglia-plasma cell crosstalk in the brain during Trypanosoma brucei infection.* Nat Commun, 2022. **13**(1): p. 5752.
18. Catta-Preta, C.M.C., et al., *Single-cell atlas of Leishmania development in sandflies reveals the heterogeneity of transmitted parasites and their role in infection.* Proc Natl Acad Sci U S A, 2024. **121**(52): p. e2406776121.
19. Abuchery, B.E., J.A. Black, and M.S. da Silva, *Single-cell transcriptomics reveals hidden information in trypanosomatids.* Trends Parasitol, 2022. **38**(1): p. 4-6.
20. Seco-Hidalgo, V., L.M. De Pablos, and A. Osuna, *Transcriptional and phenotypical heterogeneity of Trypanosoma cruzi cell populations.* Open Biol, 2015. **5**(12): p. 150190.
21. Luzak, V., et al., *Cell-to-Cell Heterogeneity in Trypanosomes.* Annu Rev Microbiol, 2021. **75**: p. 107-128.
22. Bard, J.E., et al., *Life stage-specific poly(A) site selection regulated by Trypanosoma brucei DRBD18.* Proc Natl Acad Sci U S A, 2024. **121**(29): p. e2403188121.
23. Laidlaw, R., et al., *The Trypanosoma cruzi cell atlas; a single-cell resource for understanding parasite population heterogeneity and differentiation.* bioRxiv, 2024: p. 2024.10.01.616042.
24. Diaz-Viraque, F., et al., *Genome-wide chromatin interaction map for Trypanosoma cruzi.* Nat Microbiol, 2023. **8**(11): p. 2103-2114.
25. Li, Y., et al., *Transcriptome Remodeling in Trypanosoma cruzi and Human Cells during Intracellular Infection.* PLoS Pathog, 2016. **12**(4): p. e1005511.

26. Atkinson, A., *On the measurement of inequality.* Journal of Economic Theory, 1970. **2**(3): p. 244-263.
27. Schmatz, D.M., R.C. Boltz, and P.K. Murray, *Trypanosoma cruzi: separation of broad and slender trypomastigotes using a continuous hypaque gradient.* Parasitology, 1983. **87 (Pt 2)**: p. 219-27.
28. Berna, L., et al., *Expanding an expanded genome: long-read sequencing of Trypanosoma cruzi.* Microb Genom, 2018.
29. Smircich, P., et al., *Transcriptomic analysis of the adaptation to prolonged starvation of the insect-dwelling Trypanosoma cruzi epimastigotes.* Front Cell Infect Microbiol, 2023. **13**: p. 1138456.
30. Gutierrez-Franco, A., et al., *Methanol fixation is the method of choice for droplet-based single-cell transcriptomics of neural cells.* Commun Biol, 2023. **6**(1): p. 522.
31. Berna, L., et al., *Maxicircle architecture and evolutionary insights into Trypanosoma cruzi complex.* PLoS Negl Trop Dis, 2021. **15**(8): p. e0009719.
32. Haese-Hill, W., K. Crouch, and T.D. Otto, *peaks2utr: a robust Python tool for the annotation of 3' UTRs.* Bioinformatics, 2023. **39**(3).
33. Sullivan, D.K., et al., *kallisto, bustools and kb-python for quantifying bulk, single-cell and single-nucleus RNA-seq.* Nat Protoc, 2024.
34. Hao, Y., et al., *Dictionary learning for integrative, multimodal and scalable single-cell analysis.* Nat Biotechnol, 2024. **42**(2): p. 293-304.
35. Bray, N.L., et al., *Near-optimal probabilistic RNA-seq quantification.* Nat Biotechnol, 2016. **34**(5): p. 525-7.
36. Pimentel, H., et al., *Differential analysis of RNA-seq incorporating quantification uncertainty.* Nat Methods, 2017. **14**(7): p. 687-690.
37. McInnes, L., et al., *UMAP: Uniform Manifold Approximation and Projection.* Journal of Open Source Software, 2018. **3**(29).

Supplementary Figures

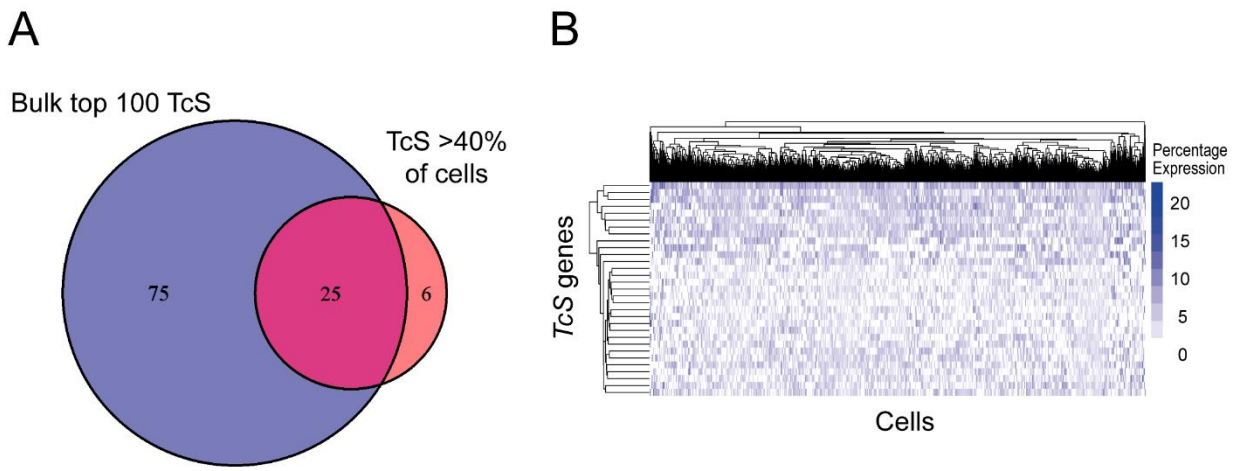


Supplementary Figure 1. Gene expression in amastigote and trypomastigote cell populations. To minimize biases related to size differences between compartments, we generated a subsampled single-copy genes list, by randomly selecting an equal number of genes to match those from the multigene family's gene set. (a) For each gene set, we calculated the summatory expression of those detected genes for each cell (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$), (b) UMAP projection for 2D visualization of core gene expression among cells, (c) Violin plots showing the number of cells expressing a specific gene belonging to each group of genes: subsampled single-copy and multigene families, ribosomal genes and different multigene families, and (d) Gini-indexes for subsampled single-copy genes, multigene family genes (together or grouped by multigene families) and ribosomal protein-coding genes as a constitutive expression control group, in trypomastigotes. All comparisons among multigene families and single-copy and ribosomal genes yield statistically significant differences.



Supplementary Figure 2.

Trypomastigote sub-clusters identified based on trans-sialidase expression profiles. Violin plots displaying average expression levels across sub-clusters and associated fold changes ($FC_{trypo_0/trypo_1}$) of (a) transporters coding genes, (b) DNA and RNA polymerase-associated protein coding genes, (c) phosphatases coding genes and (d) multigene family genes. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.



Supplementary Figure 3. (a) Venn diagram showing the overlap between the top 100 most expressed TcS from bulk RNA-seq data and TcS expressed in more than 40% of cells from cluster trypo_0, and (b) Heatmap displaying the expression of TcS genes in each cell that together account for 75% of total TcS gene expression and are expressed in more than 40% of cells within cluster trypo_0. Cells are clustered by TcS expression profiles, with colors representing each gene's percentage contribution to the cell's expression.