
Figures and figure supplements

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

Lucas Inchausti et al.

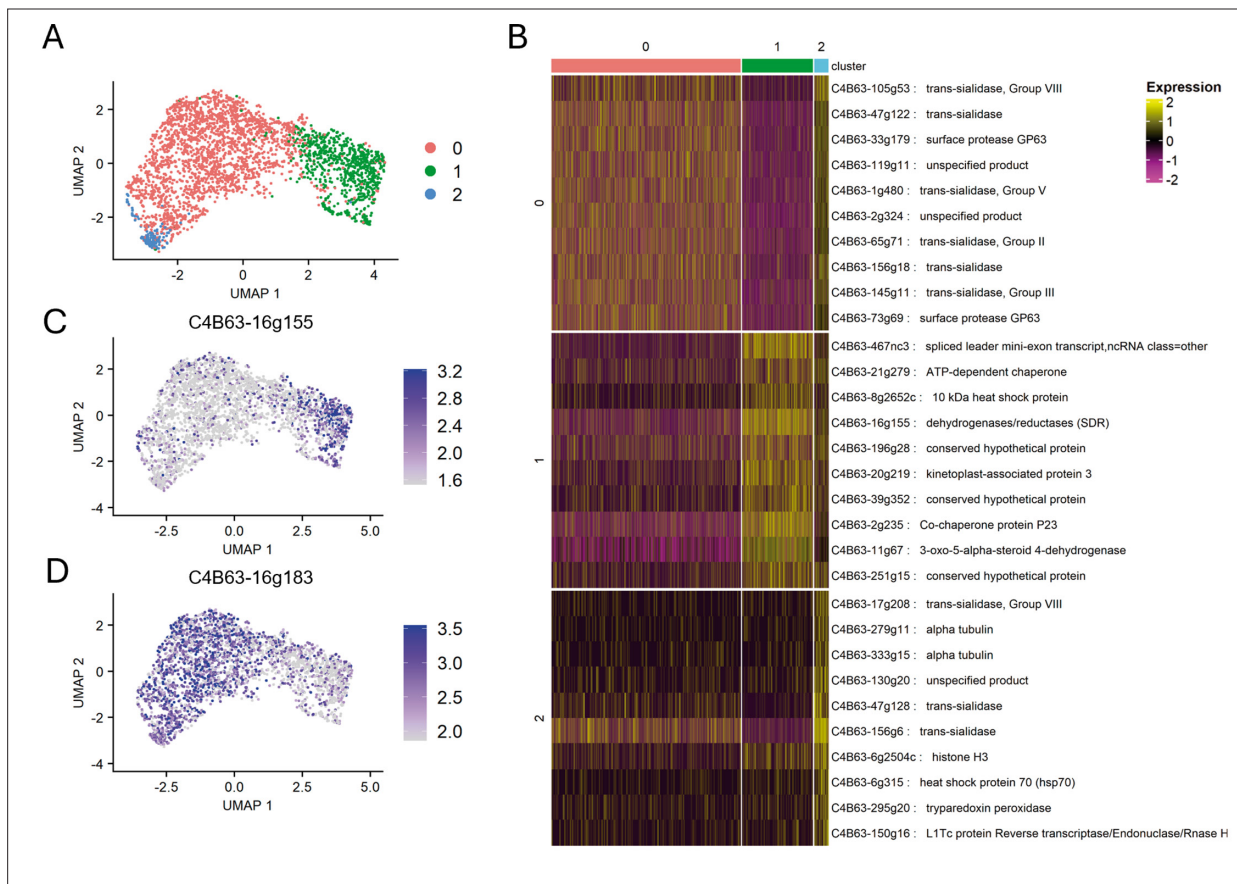


Figure 1. Identification of amastigote and trypomastigote cell populations. **(a)** UMAP colored by detected clusters based on gene expression profiles. **(b)** Heatmap of the top 10 gene markers upregulated in each of the three-cell populations identified ($\log_2FC > 1$ and adjusted p -value < 0.05). **(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.

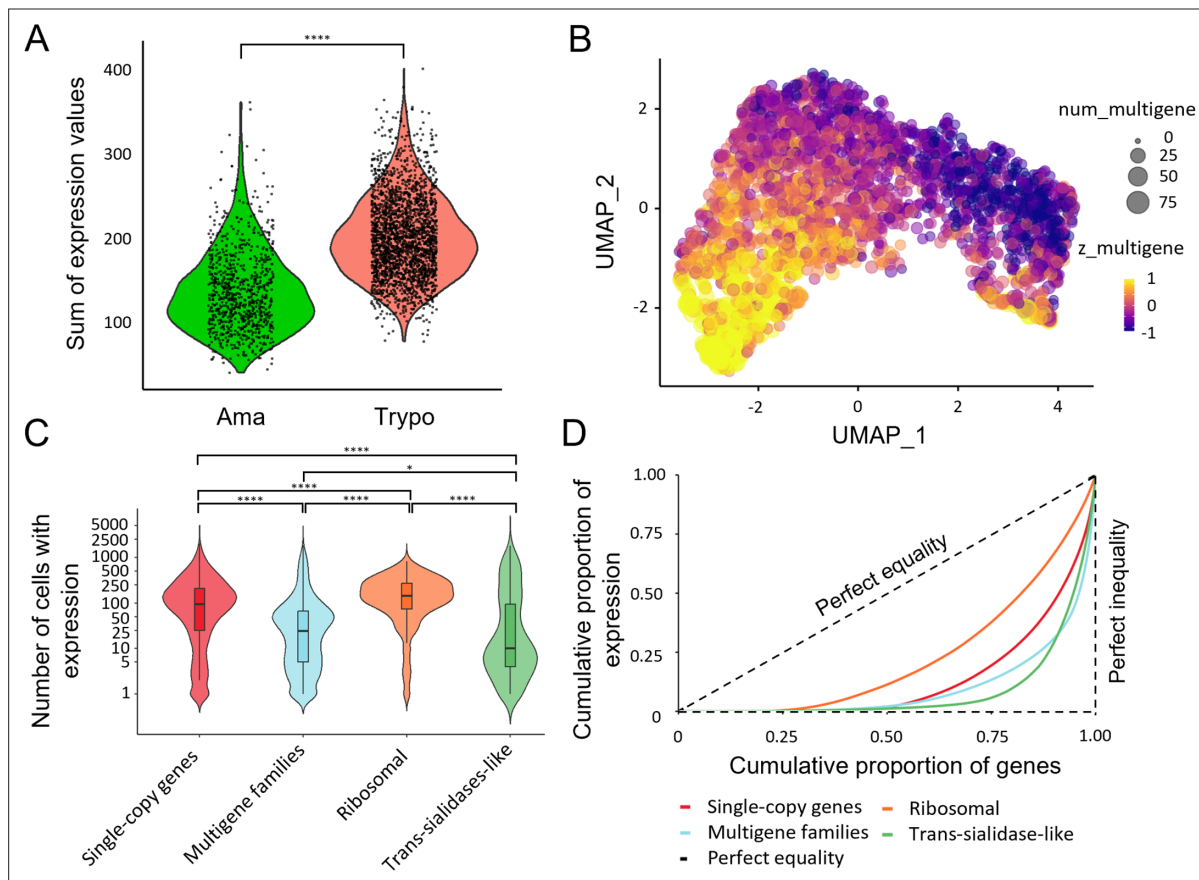


Figure 2. Overview of expression patterns across amastigote and trypomastigote cells. **(a)** Summation of expression levels values from all multigene family genes for each cell from amastigote (Cluster 1) and trypomastigote (Cluster 0) cell populations (**** $p < 0.0001$, $\text{mean}_{\text{Ama}} = 137.2$, $\text{SD}_{\text{Ama}} = 48.7$, $\text{mean}_{\text{Trypo}} = 201.1$, $\text{SD}_{\text{Trypo}} = 48.6$, $\text{FC}_{\text{Trypo/Ama}} = 1.5$). **(b)** UMAP visualization of the expression patterns of multigene family genes; num_multigene indicates the number of multigene family genes detected per cell (genes with >0 UMI counts). Z_multigene reflects the relative expression level of multigene family genes per cell, calculated as the z-score-standardized sum of their UMI counts, such that positive values reflect above-average multigene family expression and negative values reflect below-average levels. **(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 single-copy and multigene family genes, 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 (* $p < 0.05$, **** $p < 0.0001$. See **Supplementary file 3**). **(d)** Lorenz curves showing the cumulative proportion of gene expression relative to the cumulative proportion of genes for subsampled single-copy, multigene family genes, ribosomal protein coding genes, and trans-sialidase-like genes. Genes were ordered by total expression, and the dashed line indicates perfect equality. Curves that deviate further from the diagonal reflect greater inequality, meaning that fewer genes account for most of the expression within each category. Statistically significant differences between groups for **c** and **d** are shown in **Supplementary file 3**.

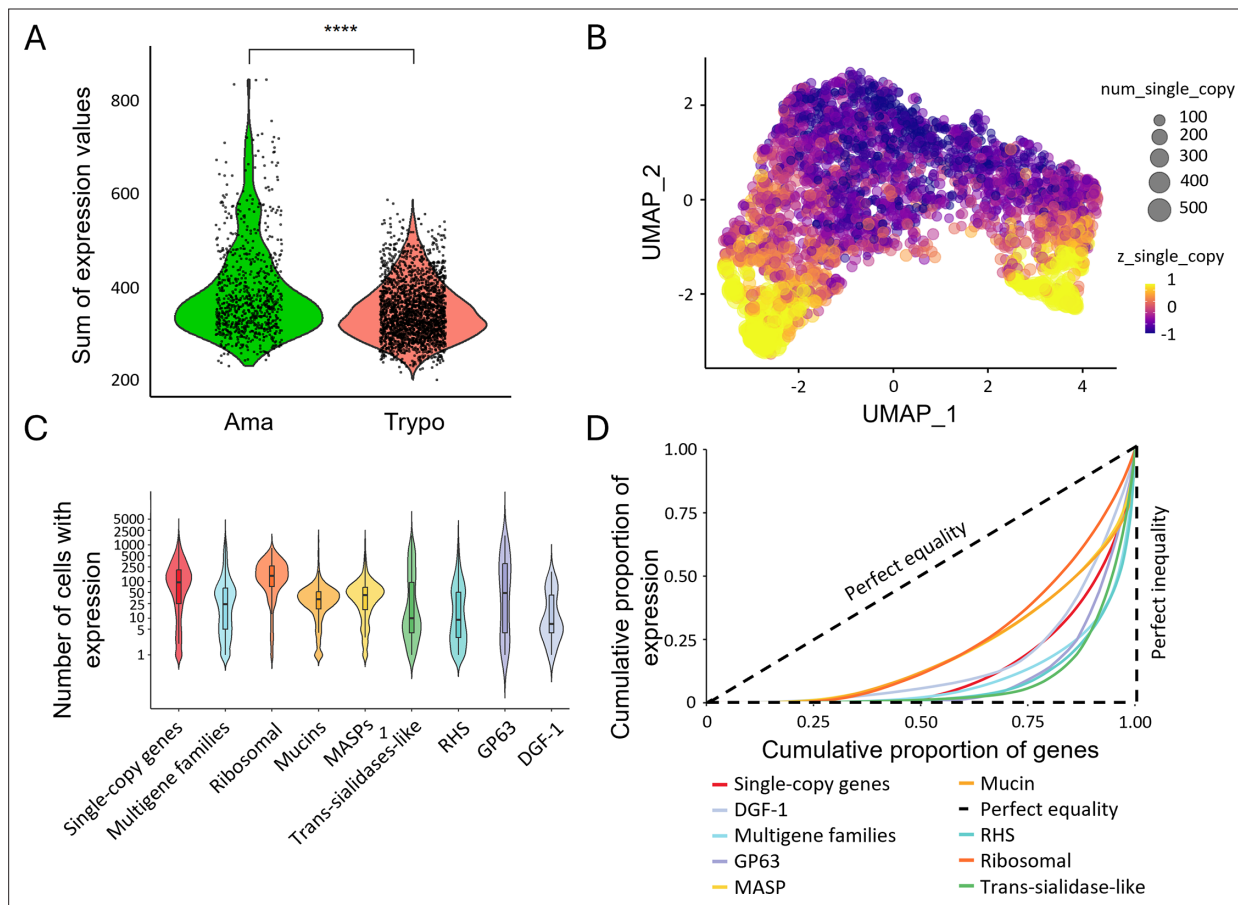


Figure 2—figure supplement 1. Extended overview of expression patterns across amastigote and trypomastigote cells. **(a)** Summatory of expression levels values from subsampled single-copy genes for each cell from amastigote (Cluster 1) and trypomastigote (Cluster 0) cell populations (**** $p < 0.0001$, meanAma = 394.9, SDAma = 101.7, meanTrypo = 353.1, SDTrypo = 62.1, FCAma/Trypo = 1.1)., **(b)** UMAP projection for 2D visualization of core gene expression among cells; num_multigene indicates the number of multigene family genes detected in each cell, whereas z_multigene indicates the expression levels calculated by summing the UMI counts of all multigene family genes in each cell and then standardizing this value using a z-score transformation, such that positive values reflect above-average multigene family expression and negative values reflect below-average levels. **(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)** Lorenz curves showing the cumulative proportion of gene expression relative to the cumulative proportion of genes for subsampled single-copy, multigene family genes (together or grouped by multigene family) and ribosomal protein coding genes. Genes were ordered by total expression, and the dashed line indicates perfect equality. Curves that deviate further from the diagonal reflect greater inequality, meaning that fewer genes account for most of the expression within each category. Statistically significant differences between groups for **c** and **d** are shown in **Supplementary file 3**.

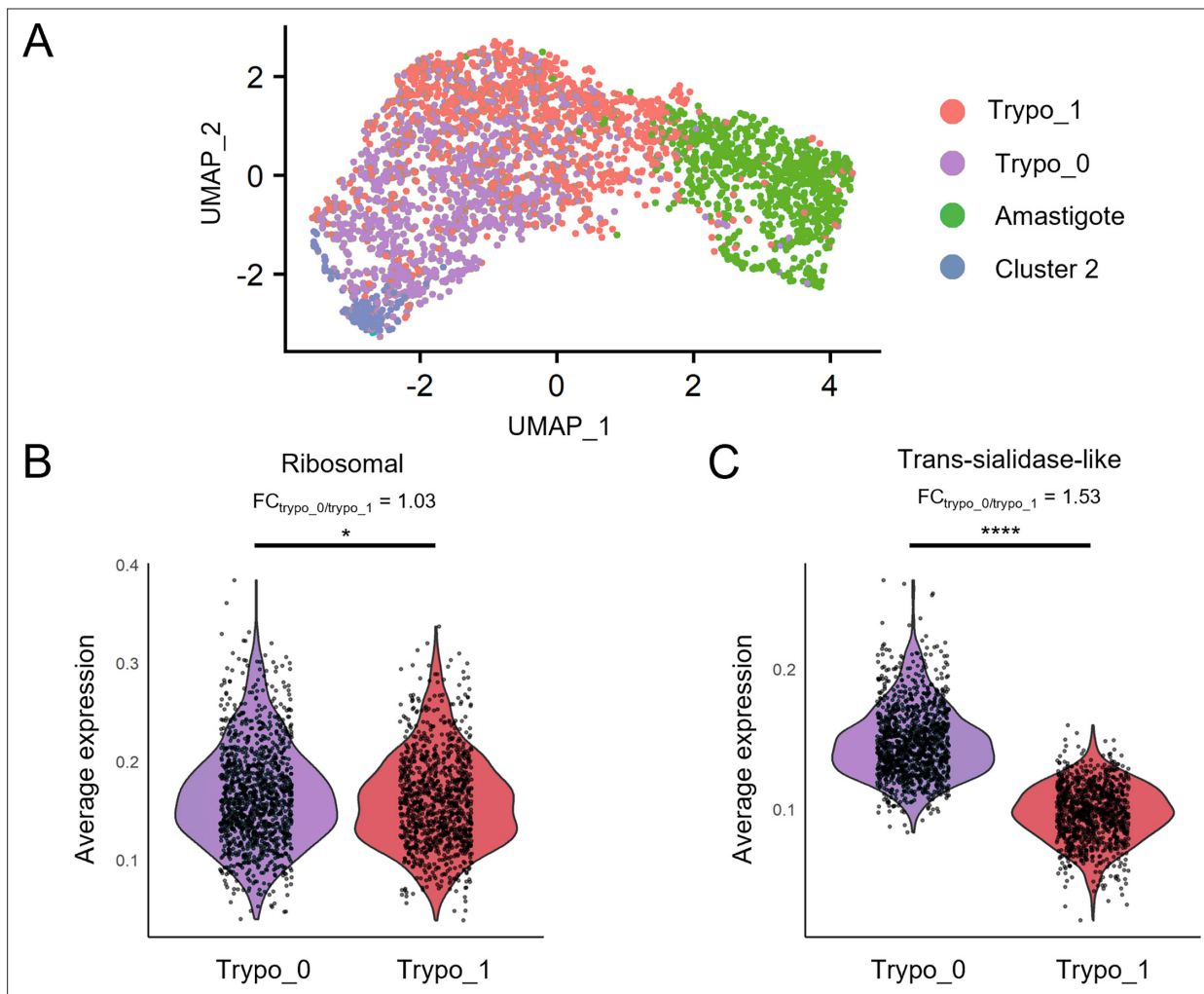


Figure 3. Trypomastigote subpopulations identified based on trans-sialidase expression profiles. (a) UMAP visualization colored by detected clusters based on gene expression profiles, with trypomastigote subpopulations identified. (b) violin plot displaying average expression levels of ribosomal protein-coding genes across sub-populations. (c) violin plot showing combined trans-sialidase expression levels for each sub-population. * $p < 0.05$, **** $p < 0.0001$.

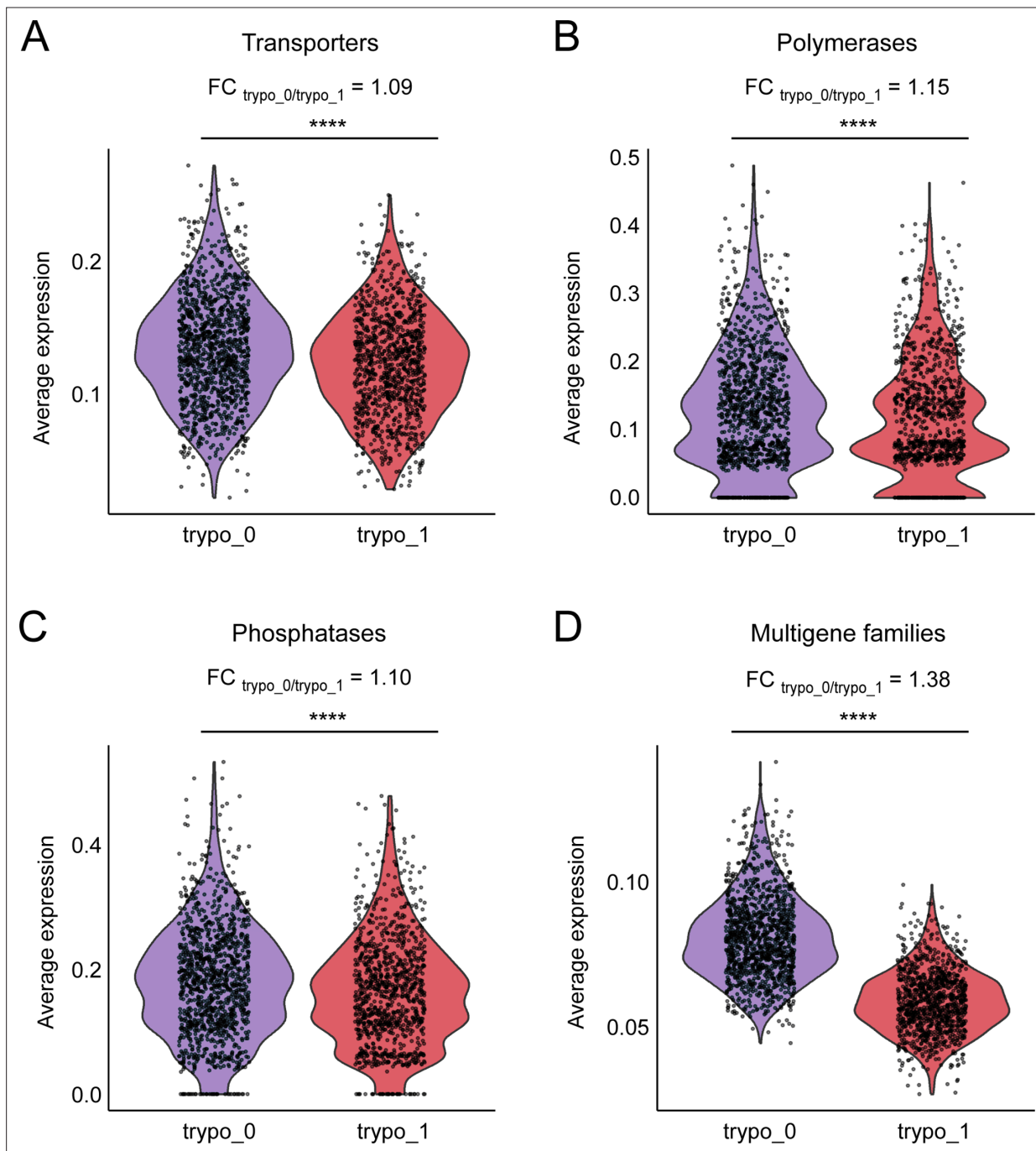


Figure 3—figure supplement 1. Trypomastigote sub-clusters identified based on trans-sialidase expression profiles. Violin plots displaying average expression levels across sub-clusters and associated fold changes ($FC_{\text{trypo_0/trypo_1}}$) of (a) transporters coding genes, (b) DNA and RNA polymerase-associated protein coding genes, (c) phosphatases coding genes and (d) multigene family genes. **** $p < 0.0001$.

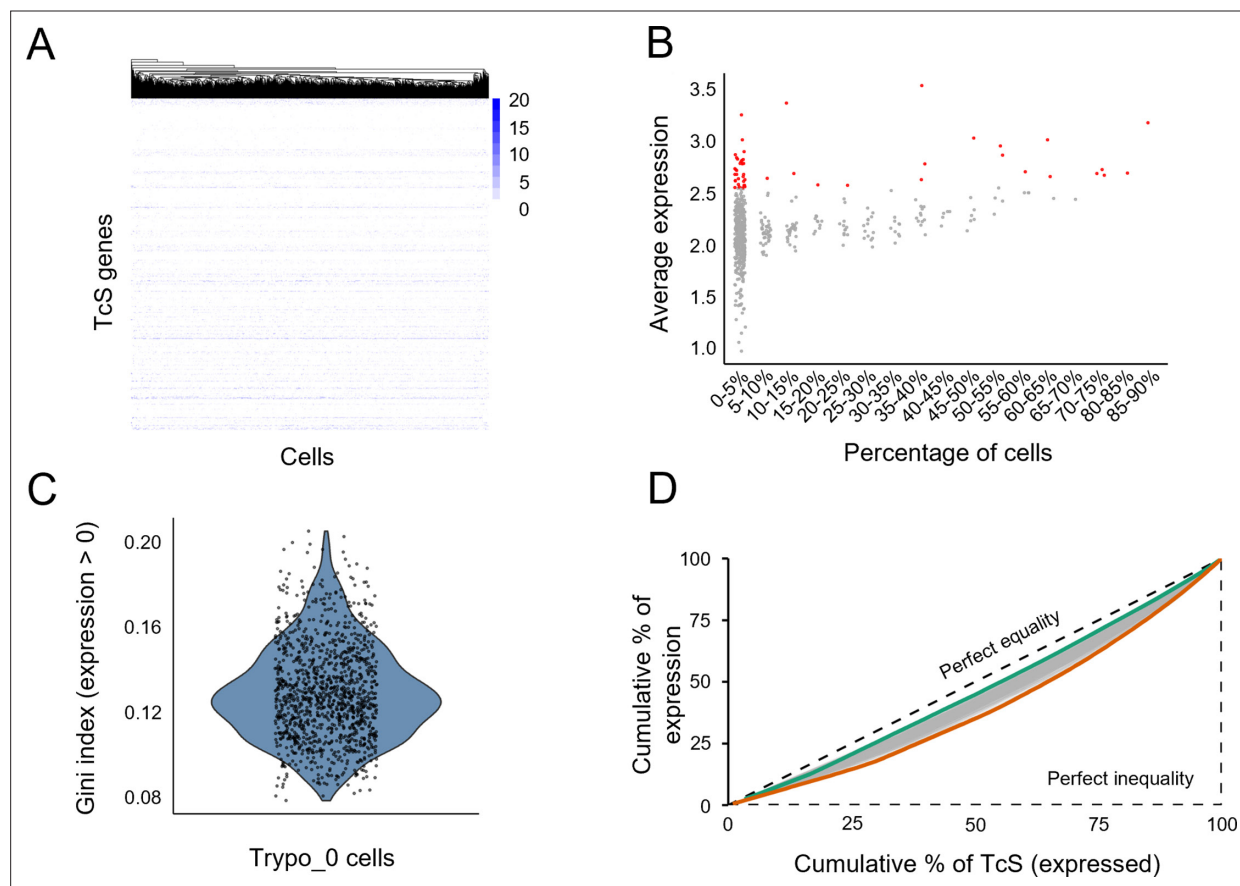


Figure 4. Overview of TcS gene expression patterns in Trypo_0 cells. (a) Heatmap displaying the expression of TcS genes in each cell that together account for 75% of total TcS gene expression within cluster Trypo_0. Cells are clustered by TcS expression profiles, with colors representing each gene's percentage contribution to the cell's total TcS expression. (b) Average expression of TcS genes grouped by the percentage of cells expressing each gene. In red are highlighted the top 50 TcS with highest average expression. (c) Gini index distribution for trypomastigotes cluster 0 (Trypo_0) cells considering only TcS detected in each cell[1] (d) Lorenz curves showing, for each cell in cluster Trypo_0, the cumulative proportion of total TcS expression as a function of the cumulative proportion of detected TcS genes. Genes were ordered by total expression, and the dashed line indicates perfect equality (i.e. all detected TcS genes contribute equally to the total TcS expression of a given cell). Green and orange curves correspond to cells with higher and lower expression equality, respectively.

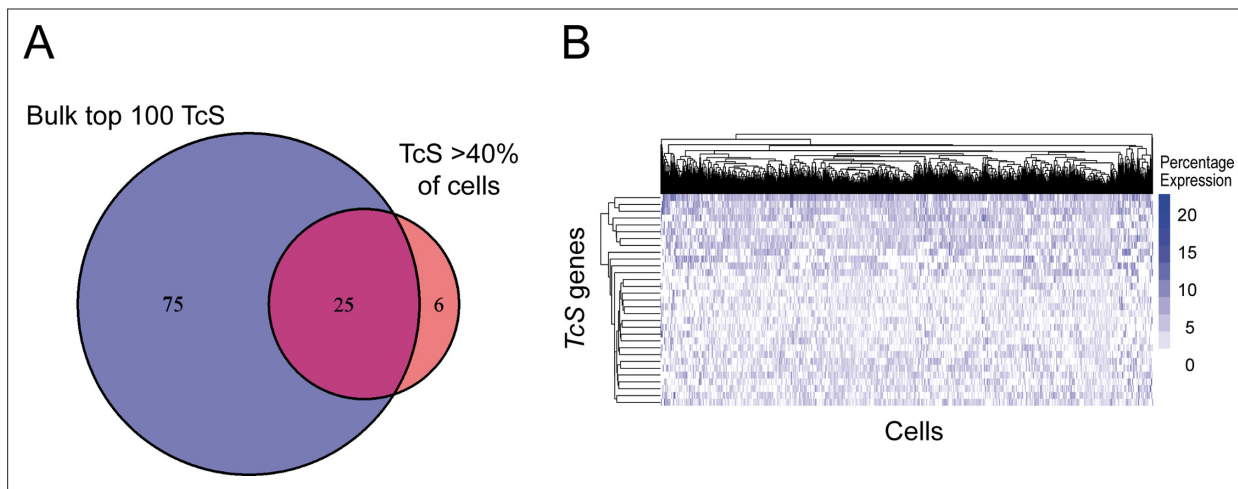


Figure 4—figure supplement 1. Overview of gene expression of high frequency TcS. (a) Venn diagram showing the overlap between the top 100 most expressed TcS from bulk RNA-seq data and TcS expressed in more than 40% of cells from cluster tryo_0, and (b) Heatmap displaying the expression of TcS genes in each cell that together account for 75% of total TcS gene expression and are expressed in more than 40% of cells within cluster tryo_0. Cells are clustered by TcS expression profiles, with colors representing each gene's percentage contribution to the cell's expression.

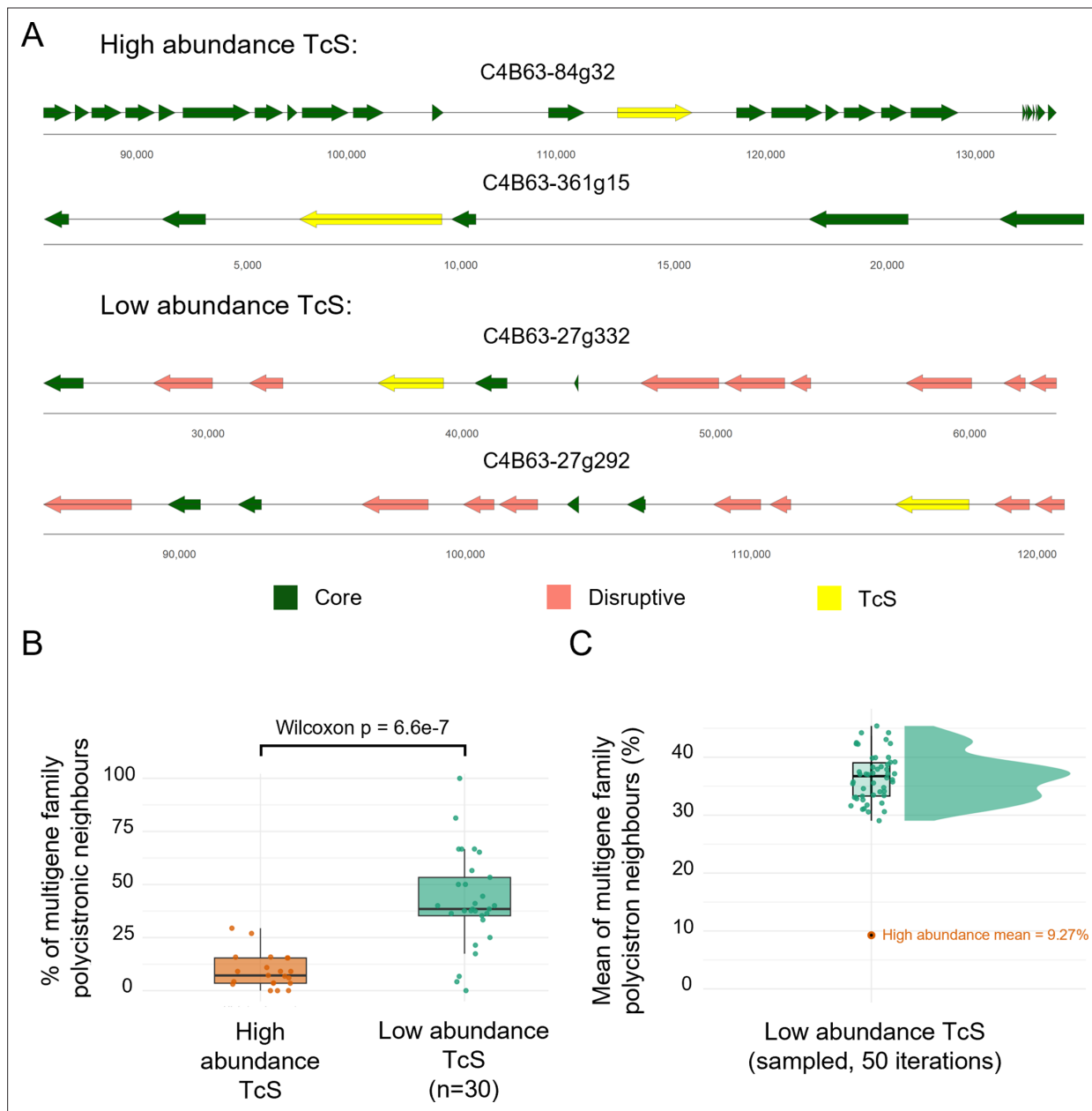


Figure 5. Genomic context and neighborhood composition of frequently detected versus lowly detected TcS loci. **(a)** Representative genomic loci of frequently detected (top, high abundance) and lowly detected (bottom, low abundance) TcS genes. Genes are shown as arrows, colored according to genomic compartment: core (dark green), disruptive (salmon), and TcS genes under analysis (yellow). Chromosomal coordinates are indicated below each locus. **(b)** Comparison of the percentage of multigene-family neighbors within polycistronic transcription units containing frequently detected and lowly detected TcS genes. Lowly detected TcS genes were subsampled to $n=30$. Wilcoxon rank-sum test: $p=6.6 \times 10^{-7}$. **(c)** Mean percentage of multigene-family neighbors in polycistrons calculated from 50 random subsets ($n=30$) of lowly detected TcS genes (mean = 37.76%). The orange dot indicates the corresponding mean for frequently detected TcS genes (9.27%).