"Integrating tDNA Epigenomics and Expression with Codon Usage Unravel an Intricate Connection with Protein Expression Dynamics in *Trypanosoma cruzi* "

Herbert G. S. Silva[1,2], Satoshi Kimura[3,4,5], Pedro L. C. Lima[1,2], David S. Pires[1], Matthew K. Waldor[3,4;] and Julia P. C. da Cunha[1,2*].

1. Laboratório Ciclo Celular, Instituto Butantan, São Paulo, 05585-000, Brazil.

2. Departamento de Microbiologia, Imunologia e Parasitologia, Escola Paulista de Medicina da Universidade Federal de São Paulo, São Paulo, 04023-062, Brazil.

3. Division of Infectious Diseases, Brigham and Women's Hospital, Boston, 02115, USA.

4. Department of Microbiology, Harvard Medical School, Boston, 02115, USA

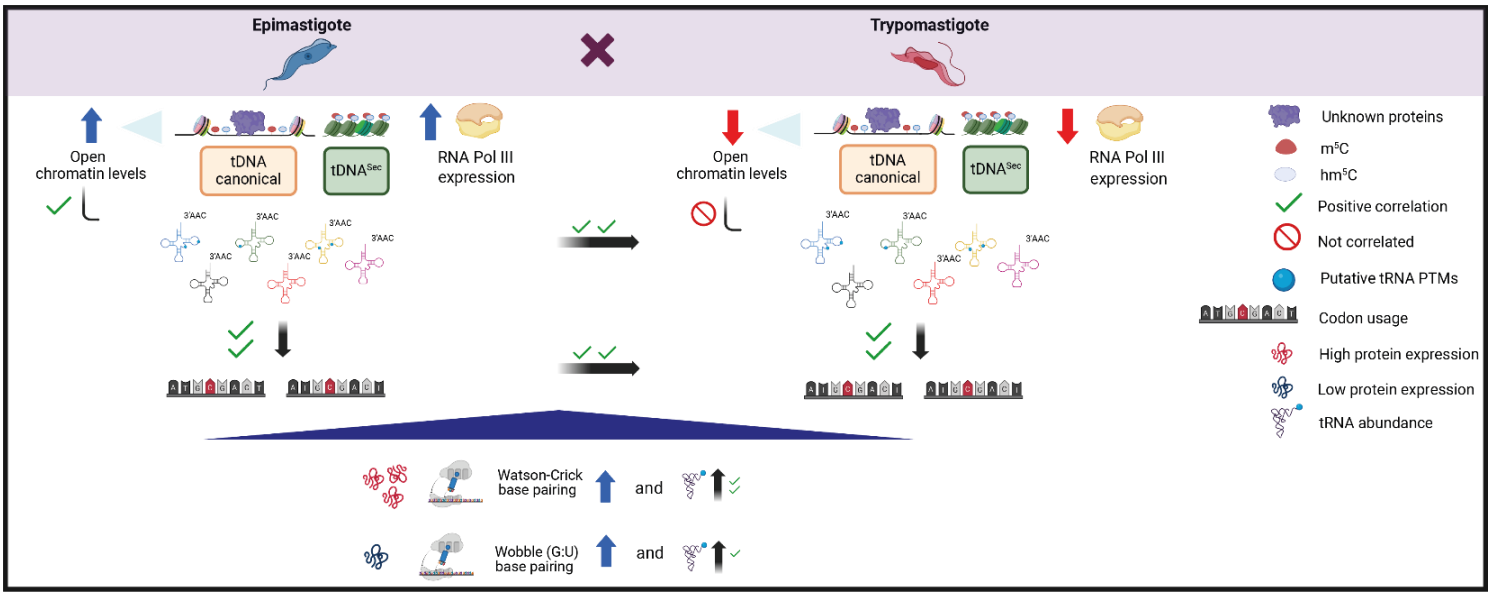
5. Department of Microbiology & Immunology, Cornell University College of Veterinary Medicine, Ithaca, 14853, USA.

#Address correspondence to Julia Pinheiro Chagas da Cunha, julia.cunha@butantan.gov.br

Abstract

Codon usage bias impacts protein expression across all kingdoms of life, including trypanosomatids. These protozoa, such as the *Trypanosoma cruzi*, primarily regulate their protein-coding genes through posttranscriptional mechanisms. Here, we integrated analyses of codon usage with multiple high-throughput sequencing data to investigate how codon usage optimizes the production of surface virulence factors (disruptive compartment), conserved housekeeping proteins (core compartment), and proteins involved in the developmental stages of *T. cruzi*. For the first time in trypanosomatids, tRNA sequencing was employed to reveal coadaptation between codon usage and anticodon availability. Despite notable differences in the proteomes of infective and non-infective forms, they exhibited similar pools of tRNAs and similar codon usage preferences. We observed that open chromatin levels of tRNA genes correlate with tRNA expression in non-infective forms, but not in infective forms, suggesting chromatin states do not control the tRNA pool in the latter. Our analysis also revealed a relationship between anticodon:codon pairing modes and protein abundance. Highly expressed mRNAs favored Watson–Crick base pairing, whereas less expressed mRNAs displayed more wobble base pairing. Overall, our findings suggest that protein expression in *T. cruzi* is influenced by a combination of codon usage bias, tRNA abundance, and anticodon:codon pairing modes.

Graphical abstract



Introduction

The genetic code, defined by triplet codons in messenger RNAs (mRNAs), dictates the amino acid sequence of proteins. Typically, sixty-four codons are present in the genome of most organisms, each encoding one of the twenty common amino acids, except for three stop codons (UGA, UAA, UAG) involved in translation termination (1). Most amino acids are encoded by two to six 'synonymous' codons (1). The biased usage of synonymous codons, where a subset of genes has a different codon bias (nonuniform usage of synonymous codons) from that of the remainder of the genome, has been observed among different organisms (2). Such synonymous codon usage bias has been reported to contribute to the optimization of translation and mRNA stability (2–8) . However, our understanding of the impacts of codon usage bias on translation remains limited in many organisms.

Synonymous codon usage bias is linked to the availability of transfer RNA (tRNA) (2, 3, 9–11). tRNAs, which are linked to specific amino acids, have distinct anticodon sequences that bind to the triplet codons in mRNA, thereby translating the genetic code into amino acids in the A site of ribosomes (12, 13). The correspondences between tRNA anticodons and mRNA codons are complex. The pairing of the anticodon-codon sequences could be mediated by different interactions, such as Watson–Crick and wobble base pairing. The Watson–Crick interaction demands a precise match of the triplet, following the strict rules of complementary pairing (14, 15). On the other hand, wobble base pairing via inosine modification or G:U interactions allows flexibility at the first position of the anticodon (5'-3') to recognize some ribonucleotides at the third position of the codon (5'-3'). For instance, inosine bases can pair with A, U or C in codon sequences, while G pairs with U in the anticodon:codon interaction context (16–18). Thus, in many cases, a tRNA binds to multiple synonymous codons, and a single codon can be decoded by multiple tRNA species. Such redundancy in codon:anticodon pairing enables balanced and robust codon decoding.

The balance between codon usage patterns and the availability of tRNAs is crucial for efficient translation (2, 3, 9–11). Matching of the supply of tRNAs to the demand for codons is necessary for

rapid and constant rates of translation. A low availability of tRNA leads to reduced rates of translation and ribosome pausing, which leads to reduced protein production and sometimes triggers stress responses (8, 19).

tRNA abundance is regulated through genetic, epigenetic, and posttranscriptional mechanisms, including tRNA modifications (20–24). It has been reported that tRNA pools can be adjusted in response to different environmental cues, enabling rapid and increased synthesis of proteins essential for cellular adaptation to environmental changes (25). Substantial changes in the expression of different tRNA isodecoders are observed in different mouse tissues (26) and during the differentiation of human cells (27), although the overall tRNA anticodon pools (related to the decoding rates) are largely stable. Therefore, profiling the dynamics of tRNA abundance is critical for understanding the impact of codon usage bias on protein synthesis.

Several methods have been used to estimate tRNA abundance profiles. tRNA genomic copy numbers have been widely used since, in many eukaryotes, they correlate with the abundance of their corresponding codons (28–30). Although they are widely used, it is not clear whether they reflect the actual tRNA expression profiles in most organisms or under different conditions. The recent development of tRNA sequencing has enabled direct profiling of tRNA expression, but this technique has been applied only to a limited number of organisms and under a limited number of conditions. Direct profiling of tRNA abundance and its correlation with codon usage bias is crucial for understanding the principles of global translation regulation.

Trypanosomatids are a class of unicellular parasitic protozoa that include pathogens such as *Trypanosoma cruzi* (*T. cruzi*), *Trypanosoma brucei* (*T. brucei*) and *Leishmania major* (*L. major*). These protozoa exhibit polycistronic transcription of mRNA genes, and their gene expression regulation occurs mainly posttranscriptionally (31). *T. cruzi* is the etiological agent of Chagas disease and has four different developmental stages: two infective and nonproliferative stages (metacyclic trypomastigotes - MTs and tissue culture trypomastigotes - TCTs) and two noninfective and proliferative stages

(epimastigotes - EPIs and amastigotes) (32). *T. cruzi* differentiation from EPIs to MTs is followed by changes in gene expression, which is important for the parasite's virulence and survival in its new habitat (33–35).

The *T. cruzi* genome is organized into conserved syntenic segments within trypanosomatids (the core compartment) and a nonsyntenic segment (the disruptive compartment) composed of virulence factor genes associated with multigene families, such as trans-sialidase (TS), mucins, and maspins (MASP) (36). GP63, dispersed gene family 1 (DGF-1), and retrotransposon hot spot (RHS) multigene families can be found in either compartment (36); when one of these families is found in a given compartment, the compartment is referred to as the GpDR compartment. EPIs predominantly translate genes from the core compartment, while MTs have increased production of disruptive proteins (33, 35). These distinct compartments have different GC contents (36), suggesting that distinct codon usage biases are likely associated with parasite development. Codon bias has been proposed as an important mechanism to direct global relative mRNA and protein expression levels in trypanosomatids (6, 37). However, a poor correlation between codon usage frequency and the copy number of tRNA genes was observed in *T. cruzi* (38), suggesting that tRNA abundance is controlled by additional mechanism(s). However, how codon usage and tRNA profiles affect the *T. cruzi* translatome and development is largely unclear.

Here, we combined multiple genome-wide datasets to investigate the relationships among tRNA abundance, codon usage, and gene expression profiles in *T. cruzi*. Codon usage analysis revealed that the two *T. cruzi* genomic compartments, as well as transcripts of higher and lower abundance, exhibit codon bias. We adopted tRNA sequencing in two developmental forms (EPI and MT) to profile tRNA abundance. Despite the distinct tRNA chromatin states in these two forms, their tRNA profiles were very similar, suggesting that another mechanism controls tRNA levels across different developmental forms. We observed that highly translated genes predominantly use Watson–Crick base pairing, while poorly translated genes rely more on wobble pairing. This differential use of anticodon:codon pairing

modes, together with tRNA availability, may represent a previously undescribed layer of gene expression regulation in *T. cruzi*.

Materials and methods

Cell culture: *T. cruzi* (Dm28c strain) EPIs were cultivated at 28 °C in Liver Infusion Tryptose (LIT) medium supplemented with 10% fetal bovine serum (FBS; Vitrocell), 0.4% glucose, 0.1 µM hemin, and 60 mg/mL penicillin G, as described by Camargo *et al*, 1964. The metacyclic trypomastigote forms were obtained following the protocol outlined by Contreras et al., 1985 (39) with some modifications. In brief, epimastigotes in the exponential growth phase ($4x10^6$ parasites/mL) were cultured for 4 days until they reached the stationary phase ($5\text{-}6x10^7$ parasites/mL). Subsequently, the parasites were resuspended at a concentration of $5x10^8$ parasites/mL in Triatomine Artificial Urine (TAU) medium (190 mM NaCl, 17 mM KCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$, and 8 mM phosphate buffer, pH 6.0) and incubated for 2 hours at 28 °C. Then, the parasites were diluted to $5x10^8$/mL in TAU 3AAG medium (containing 10 mM L-proline, 50 mM L-glutamate, 2 mM L-aspartate, and 10 mM glucose) and maintained in a CO_2 incubator at 28 °C. Metacyclic forms of the *T. cruzi* Dm28c strain were harvested from the culture supernatant after 76 hours of incubation and purified using DEAE-cellulose resin (Sc-211213).

tRNA species identification: The *T. cruzi* Dm28c (version 42-2018) genome and its coordinates were downloaded in FASTA and gff formats, respectively, from the TritrypDB platform (https://tritrypdb.org/tritrypdb/app/downloads). To obtain the list of tDNAs and their respective coordinates, the term “tRNA” was used as a filter in the gff file. Identification of selenocysteine tRNA ($tRNA^{Sec}$) was performed using the BLASTp tool v.2.10.0 (40) to search for the best hit via the TriTrypDB platform (https://tritrypdb.org/tritrypdb/app/workspace/blast/new). The query reference was “tRNA selenocysteine” (ID=Tb927.9.2380) from *T. brucei* TREU927. tRNAscan-SE software (http://lowelab.ucsc.edu/tRNAscan-SE/) (41) was used to identify the transcribed sense strands and anticodons of the tDNAs. Afterward, a genome FASTA file containing all identified tDNA sequences was obtained.

tRNA sequencing: Total RNA was extracted using TRIzol (Invitrogen) following the manufacturer's instructions. tRNA purification and sequencing were performed according to Kimura *et al*., 2020 (42) with some modifications. Five micrograms of RNA samples were run on a 10% denaturing gel (TBE-UREA), and bands corresponding to tRNAs (70 to 85 bp) were excised, digested, eluted, and recovered by isopropanol precipitation. Subsequently, 250 ng of purified tRNA was deacetylated with 500 µL of 100 mM Tris-HCl (pH 9.0) solution and incubated for 1 hour at 37 °C. The tRNA was recovered using isopropanol. Afterward, library preparation for sequencing was carried out as described by Kimura et al., 2020. The tRNA samples were sequenced in biological duplicates for EPIs and MTs on the Illumina NextSeq 1000 system (single-end). The removal of adapters (AGATCGGAAG) from the FASTQ files was performed using cutadapt v.3.5 (43) , followed by HEADCROP:2 using the TrimmomaticSE v.0.39 tool (44). The tDNA FASTA file generated in this study (see Supplementary Table 2SA) was used for mapping. The **i.** mapping, **ii.** abundance counting, **iii.** Analysis of differential tRNA expression, counting of 3'CCA on tRNA and misincorporation frequencies between epimastigote and trypomastigote metacyclic forms were performed using the mim-tRNAseq script with the parameters --cluster-id 0.97, --max-mismatches 0.075, --max-multi 4 and --max-multi 4 (45), executed in Python 3.9.18 and R 4.3.1. The tRNA-seq dataset were deposited at PRJNA1124437.

Identification of RNA polymerase III subunits in the *T. cruzi* Dm28c genome: The sequences of the RNA polymerase III subunits (RPC1, RPC2, RPC40, RPC19, RPB6, RPB5, RPB8, RPB10, RPC10, RPC11, RPC17, RPC25, RPC82, RPC53, RPC37, RPC34, and RPC31) of *Saccharomyces cerevisiae* S288c were downloaded from the FungiDB database (https://fungidb.org/fungidb/app). Subsequently, two approaches were employed to identify orthologs in the *T. cruzi* Dm28c strain (version 42–2018) genome. The first approach involved using the BLASTp tool v.2.10.0 (40), which is available on the TriTrypDB platform (https://tritrypdb.org/tritrypdb/app/) to search for alignments with an E-value ≤0.001, followed by confirmation of the protein domain using the platform

(https://www.ebi.ac.uk/interpro/search/sequence/). For the unidentified RPC31 protein homologous using the BLASTp method, probabilistic models were created with the HMMER Algorithm v.3.3 (46) based on orthologous sequences of RPC31 from several eukaryotes. Subsequently, a search for the best hits was conducted within the *T. cruzi* Dm28c strain (version 42–2018) genome using the Ubuntu v.20.04 system.

FAIRE-seq, MNAse-seq, modified bases analysis in EPI and TCT forms: The RPGC values available from FAIRE-seq data performed in the EPI and MT forms were obtained from Lima et al, 2021(47). Raw files from MNase-seq data in the EPI and TCT forms were obtained from the SRA project number PRJNA665060 (48). The reads were processed using Trimmomatic (44) to remove adapter sequences, and the following parameters were applied: HEADCROP:5, SLIDINGWINDOW:15:25, and MINLEN:35. The rRNA reads were masked using bedtools maskfasta v2.31.1. Reads mapping was performed using Bowtie2 (49) to the *T. cruzi* Dm28c genome (version 42–2018) available on TriTrypDB using the following parameters: -D 25, -R 4, -N 0, -L 19, and -i S,1,0.40. SAM files were converted to BAM files, which were subsequently sorted and indexed using SAMtools version 1.12 (50). Read count values for tDNA regions, selenocysteine loci, and shuffled regions equivalent to tRNAs (generated by running Bedtools v2.31.1 with the shuffle option) were obtained using BAMscale Version v1.0 (51). The resulting counts in FPKM were analyzed using libraries such as readr, dplyr, reshape2, ggpubr, and ggplot2 in R Version 4.1.2. Modified bases analysis was performed as described in Lima *et al*, 2024 (52). In short, public Oxford Nanopore sequencing datasets available at Díaz-Viraqué *et al*, 2023 (53) were converted into POD5 files using pod5 library version 0.3.6 (pod5 convert fast5), which were used to call methylation (m) and hydroxymethylation (h) cytosines with the Dorado program, version 0.5.3, followed by read mapping using the minimap2 program (54) against the genome of the Dm28c 2018 strain (version 60). The

analysis of methylated bases was performed using the Modkit program, version 0.2.5 (available at https://github.com/nanoporetech/modkit).

Ribo-seq data analysis of the *T. cruzi* Dm28c genome: Ribo-seq data for EPI and MT forms of *T. cruzi* were obtained from the Sequence Read Archive (RCA) repository, project ID: PRJNA260933, provided by (33). The FASTQ files were filtered using the TrimmomaticSE v.0.39 tool (44) with parameters –HEADCROP:1 and MAXINFO:24:0:0.05 and mapped to the *T. cruzi* Dm28c (version 42–2018) genome using Kallisto's algorithm to obtain the transcripts per million (TPM) values with the following settings: -b 100–l 50–s 20 (55). The DESeq tool v.1.42 (56) was used to perform differential expression analysis of the translatome between the EPI and MT forms.

Codon usage analysis in the *T. cruzi* Dm28c genome: The aa.usage.py script available at https://github.com/rhondene/Codon-Usage-in-Python was used to compute the **i.** relative synonymous codon usage (RSCU), **ii.** codon relative adaptive weights, **iii**. The absolute codon frequencies and amino acid frequencies are shown in FASTA files. The identified genes for the core, disruptive and GpDR compartments were obtained from Lima *et al*, 2021 (47). All upregulated CDSs between EPIs (1,752 genes) and MTs (361 genes) were generated from Ribo-seq data, excluding pseudogenes. The top 100 and bottom 100 genes (reference set) were obtained from the Ribo-seq data containing the CDSs with 100 genes with higher or lower TPM values in the EPI and MT forms. The reference set was selected based on genes with a log2-fold change of ≤ 2 (EPIs vs. MTs) and an adjusted p value of ≤ 0.05, excluding pseudogenes and genes present in multiple copies. The tRNA abundances, represented as the proportion of mapped reads, were determined for EPIs and MTs. To refine these measurements, a renormalization process was applied to address the potential for a single tRNA to recognize multiple codons. This was achieved by dividing the abundance of each tRNA by the number of its corresponding codons and then assigning these adjusted abundance values to each codon. For

codons recognized by multiple tRNA species, the total abundance was calculated by summing the adjusted abundances of all tRNA species recognizing that codon.

Results

Core and disruptive genomic regions have differences in codon usage.

The core and disruptive genomic compartments differ in their GC content, gene class diversity, evolutionary history, and their protein-coding genes are differentially expressed in *T. cruzi* life forms (36). EPIs predominantly translate core genes, while MTs exhibit elevated expression of disruptive transcripts (33, 35). This pronounced shift in their respective proteomes raised the question of whether these two genomic compartments, core and disruptive, display different patterns of relative frequencies of amino acids and codons in their coding regions. We observed that the codon frequencies of the core genes were different from those of the disruptive and GpDR genes (Figure 1A). Genes from the core and disruptive regions had high correlation levels when amino acid (R=0.81) or codon absolute frequencies (R=0.75) were analyzed (Figure 1B-C and Table Supplementary 1SA). Ala and Gly had the highest amino acid frequencies for the core and disruptive compartments, respectively. Codons for Glu were among the highest for both compartments.

To obtain further insights into codon bias, we compared the relative synonymous codon usage (RSCU) values for core and disruptive CDSs. RSCU indicates the preference for specific codons that encode the same amino acid or translation termination codon (5, 57). Higher RSCU values indicate optimal codons, which are associated with more abundant corresponding tRNA species for those amino acid-encoding codons (9). The GUG and UGA codons exhibited the highest RSCU values for the core and disruptive CDSs, respectively (Figure 1D). Interestingly, UGA is one of the three stop codons, and in accordance, the relative codon adaptive weights for stop codons differ greatly among genomic compartments. Strikingly, UGA stop codons represented 94% of stop codons in disruptive genes but only 47% and 61% of stop codons in the core and GpDR compartments, respectively (Figure 1E). Among disruptive genes, we detected that TS and MASP genes use UGA as a stop codon for almost 100% of their genes (Figure 1F). In addition, we detected five distinct optimal codons for the amino acids Glu, Gly, Pro, Ser and Thr in the core and disruptive genes (Figure 1G and Table Supplementary 1SA). Moreover, when analyzing codon relative adaptive weights compared to their respective optimal counterparts, an evident bias for Arg (CGA), Cys (UGU), and Tyr (UAU) was found for the core CDSs, while Gln (CAA), Glu (GAA), Lys (AAA) and Phe (UUC) were more prevalent in the disruptive CDSs (Figure 1H and Table Supplementary 1SA). Collectively, although the core and disruptive genomic regions have similar frequency of amino acid and codon usage (Figure 1 B and C), these two genetic regions demonstrated different preferences in synonymous codon usage and optimal codons.

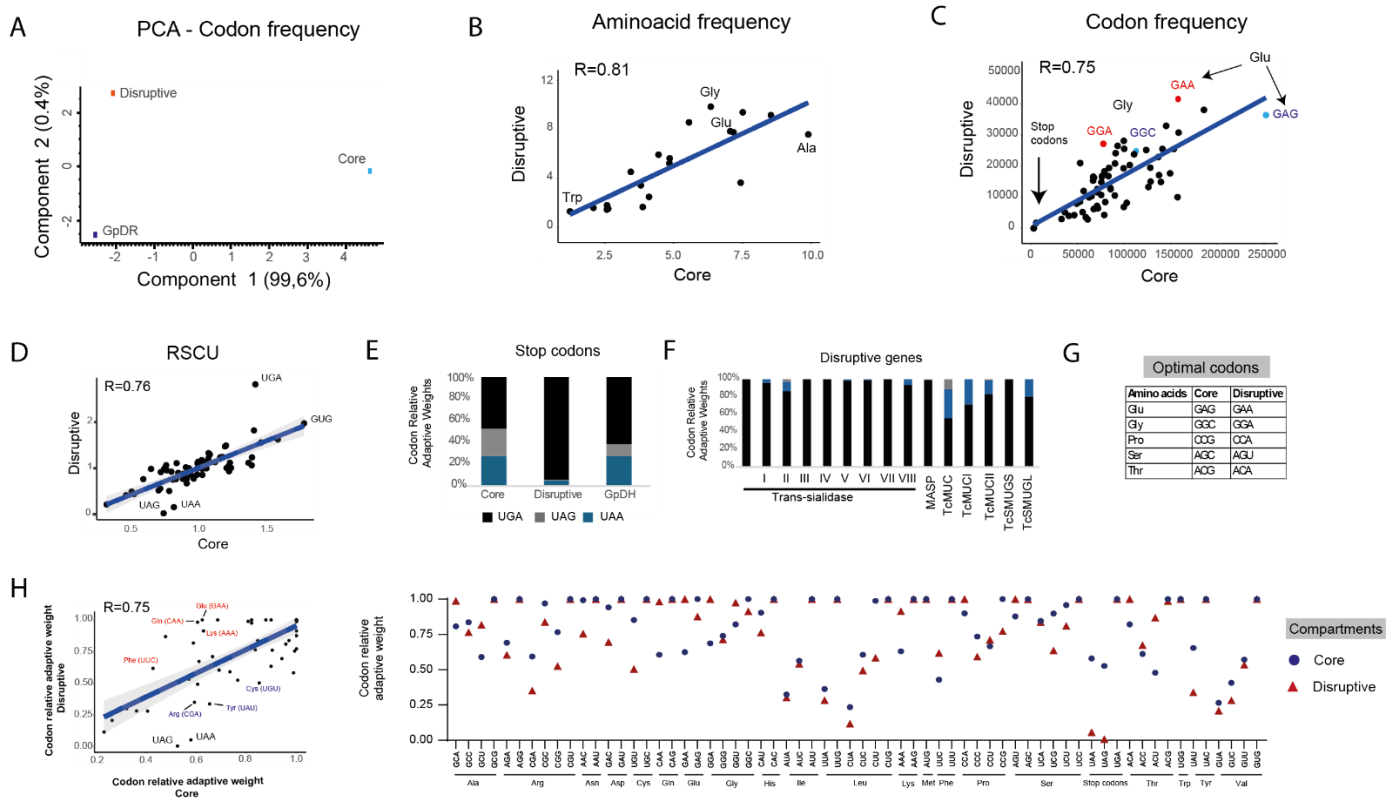


Figure 1. Identification of codon usage bias in the *T. cruzi* genome. **A)** Principal component analysis (PCA) of codon absolute frequencies for the core, disruptive and GpDR compartments. Scatter plot for either **B)** amino acid or **C)** codon usage frequencies in core and disruptive nucleotide sequences. Pearson's correlations are shown above each graph. **D)** Scatter plot based on Pearson's correlation of RSCU values for core and disruptive genomic compartments. **E)** Percentage of codon relative adaptive weights for stop codons (UGA, UAG and UAA) in the core, disruptive and GpDR compartments. **F)** Enrichment of the stop codons on the class of disruptive genes: trans-sialidases, MASP, and mucins. **G)** Table displaying the distinct optimal codons between the core and disruptive compartments. **H)** Relative adaptive weights of codons compared to those of optimal codons from core and disruptive compartments. The highlighted codons are at least 20% different between the indicated conditions. The Pearson's correlation between both compartments is depicted on the left side, with the respective values for each codon presented on the right side. Gray shading, 95% confidence interval.

Characterization of tRNA isotypes and anticodon:codon pairing modes

In the *T. cruzi* hybrid strain CL Brener, the genome contains forty-six anticodons, allowing for the translation of sixty-two codons representing the 20 canonical amino acids and one noncanonical amino acid, selenocysteine (Sec) (38). The nonhybrid *T cruzi* strain Dm28c displays a lower tDNA copy number, with a total of 105 tDNAs compared to CL Brener, which has 120 copies (38, 58). Our study

focused on the *T. cruzi* Dm28c strain, which, similar to the CL Brener strain, has poorly characterized anticodon:codon pairing modes. Thus, we used the tRNAscan-SE tool to characterize the isoacceptors and isodecoders and subsequently established their respective base pairings with the sixty-two codons present in the *T. cruzi* Dm28c genome. We identified forty-six isoacceptors (anticodons) and eight isodecoders in the *T. cruzi* Dm28c strain (Figure 2A) that specifically carry one of twenty canonical amino acids found in nature, as well as one tRNA that carries the noncanonical amino acid selenocysteine. The majority of tRNA species are present as a single- or two-gene copies, as depicted in Figure 2A.

In addition, we evaluated whether tRNA anticodons containing adenosine at the first position (5'-3') could undergo modification to inosine (I) in the EPI and MT forms as a strategy to complement codons with missing anticodon matches. The I modification provides wobble base pairing flexibility between synonymous codons. For instance, $tRNA^{Ala(CGA)}$ modified to $tRNA^{Ala(CGI)}$ recognizes the codons ending (5'-3') with U, C and A as Ala (GC**U**), Ala (GC**C**), and Ala (GC**A**), expanding its recognition beyond Ala (GCU). The detection of I can be achieved using tRNA-seq data, as this modification induces high levels of nucleotide misincorporation (90% to 100%) at modified base positions during reverse transcriptase (RT) from RNA to cDNA, thereby generating an RT signature (42, 59) . Thus, we sequenced the total tRNAs from both the EPI and MT forms in duplicate, and the reads were mapped onto a FASTA file genome composed of fifty-five tRNA sequences (Figure Supplementary S1A). The sequencing reads exhibited high quality, coverage and reproducibility between the duplicates (Figure Supplementary S1B-D). We found that all tRNA anticodons beginning (5'-3') with adenosine in *T. cruzi* are modified by inosine, including $tRNA^{Ser1\ (AGA)}$, $tRNA^{Val1\ (AAC)}$, $tRNA^{Ala1\ (AGC)}$, $tRNA^{Leu1\ (AAG)}$, $tRNA^{Arg2\ (ACG)}$, $tRNA^{Pro1\ (AGG)}$, $tRNA^{Ile1\ (AAT)}$, and $tRNA^{Thr1\ (AGT)}$. These modifications remain stable when comparing the EPI and MT forms; they had no detectable changes in their levels across both life forms (Figure 2B).

Based on the variations of anticodons in tRNA genes, the all anticodon:codon base pairing possibilities in *T. cruzi* (**i.** Watson–Crick only, **ii.** wobble (G:U) only, **iii.** wobble (inosine) only, **iv.** Watson–Crick and wobble (G:U) and **v.** Watson–Crick and wobble (Inosine)) are illustrated in Figure 2C, in which seventeen codons display only Watson–Crick pairing with their corresponding tRNA anticodon, eight display either wobble (G:U) or wobble (inosine) pairing, thirteen display Watson–Crick and wobble (G:U) pairing and sixteen codons are recognized by Watson–Crick and wobble (inosine) (Figure 2D and Table Supplementary 2A and B).

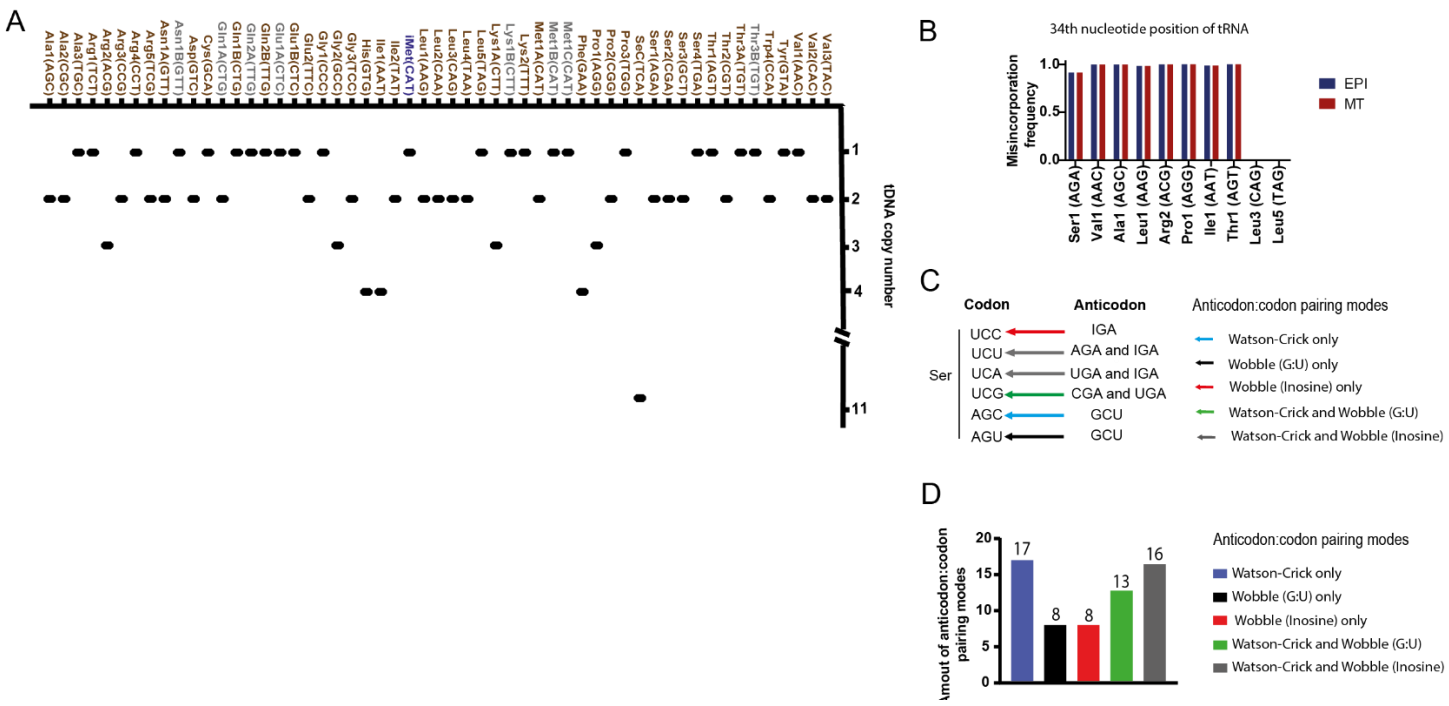


Figure 2. Identification of tRNA isotypes and anticodon:codon base pairing modes in the *T. cruzi* (Dm28c) genome. **A)** tDNA copy number for all tRNA species in the *T. cruzi* genome. The tRNA isoacceptors and isodecoders are represented by brown and gray, respectively. $tRNA^{imet(CAU)}$ is represented in blue. **B)** RT signature (misincorporation frequency) for inosine modification at the 34 nucleotide position of tRNAs in the EPI and MT forms obtained from tRNA-seq data. $tRNA^{Leu3\ (CAG)}$ and $tRNA^{Leu5\ (TAG)}$ were used as negative controls for the absence of this modification. **C)** Illustration of the anticodon:codon base pairing modes found in the *T. cruzi* genome. **D)** Number of anticodon:codon pairing modes (Watson–Crick only, **ii.** wobble (G:U) only, **iii.** wobble (inosine) only, **iv.** Watson–Crick and wobble (G:U) and **v.** Watson–Crick and wobble (Inosine)).

The relative abundance of tRNAs is similar in the EPI and MT forms, despite differences in their biochemical modifications.

The global quantification of tRNA species through sequencing has significant challenges, primarily due to the presence of numerous nucleotide modifications in tRNA molecules. These modifications can lead to either nucleotide misincorporation or incomplete conversion of RNA to cDNA during library construction, leading to challenges in mapping and quantification (45). Additionally, the presence of multiple isodecoders, which contain variations in tRNA sequences outside of the anticodon, makes accurate mapping difficult. To circumvent these challenges, we used a highly processive reverse transcriptase for efficient library construction and the mim-tRNAseq analysis pipeline (42, 45). The EPI and MT forms of *T. cruzi* displayed a very strong positive correlation (Pearson's R =0.91) related to

their tRNA abundance (Figure 3A). Three (**i**.e., Asn1A (GTT)/Asn1B (GTT); **ii**. Lys1A (CTT)/Lys1B (CTT); and **iii**. Gln2A (TTG)/Gln2B (TTG) (Figure 3B and Supplementary Table 2C) of the eight isodecoders could not be definitively mapped to their isoacceptors due to the high similarity between these sequences and the lack of unique mapped reads.

EPIs and MTs express different sets of genes. In particular, EPIs mainly express genes encoded in the core compartment, whereas MTs exhibit increased expression of genes encoded in the disruptive compartment (33, 35). Thus, we hypothesized that the tRNA availability in EPIs and MTs fit the codon usage of the core and disruptive compartments, respectively. We detected greater expression of tRNA Glu1B (CTC), Glu1A (CTC), and Thr2 (CGT) in EPIs than in MTs (p-adjusted <0.05) (Figure 3C and Supplementary Table 2D), which corresponds to two (Glu (GAG) and Thr (ACG)) of the five optimal codons in the core compartment (Figure 1G). In contrast, tRNAs corresponding to any of the five optimal codons from the disruptive compartment were not increased in MTs (Figure 1G). These data suggest that there is some, albeit limited, correspondence in the tRNA expression profile in EPIs to fit the core compartment, but the expression profile in MTs does not fit the disruptive compartment. Thus, tRNA abundance undergoes only minor modulation during the differentiation process from EPIs to MTs, despite significant differences in protein expression observed between these forms.

Furthermore, we found that tRNAs with anticodons exhibiting wobble base pairing are not more abundant than those relying solely on Watson–Crick base pairing. This suggests the lack of a regulatory mechanism to increase the abundance of wobble tRNAs to meet the demand for multiple codons (Figure 3D and Supplementary Table 2E).

In addition, since some tRNA modification causes misincorporation of wrong bases during RT, misincorporation frequency can be used to probe the frequency of tRNA modification. Thus, although the abundances of tRNAs in EPIs and MTs are similar, we identified a marked reduction of putative post-transcriptional modifications and maturation in several tRNAs in MT compared to EPI by analyzing misincorporation and 3’CCA rates, respectively (Figure 3E and F), suggesting that the changes of modification states in two life forms can modulate the tRNA functions, thereby contributing to the adaptation to the codon usage in MT.

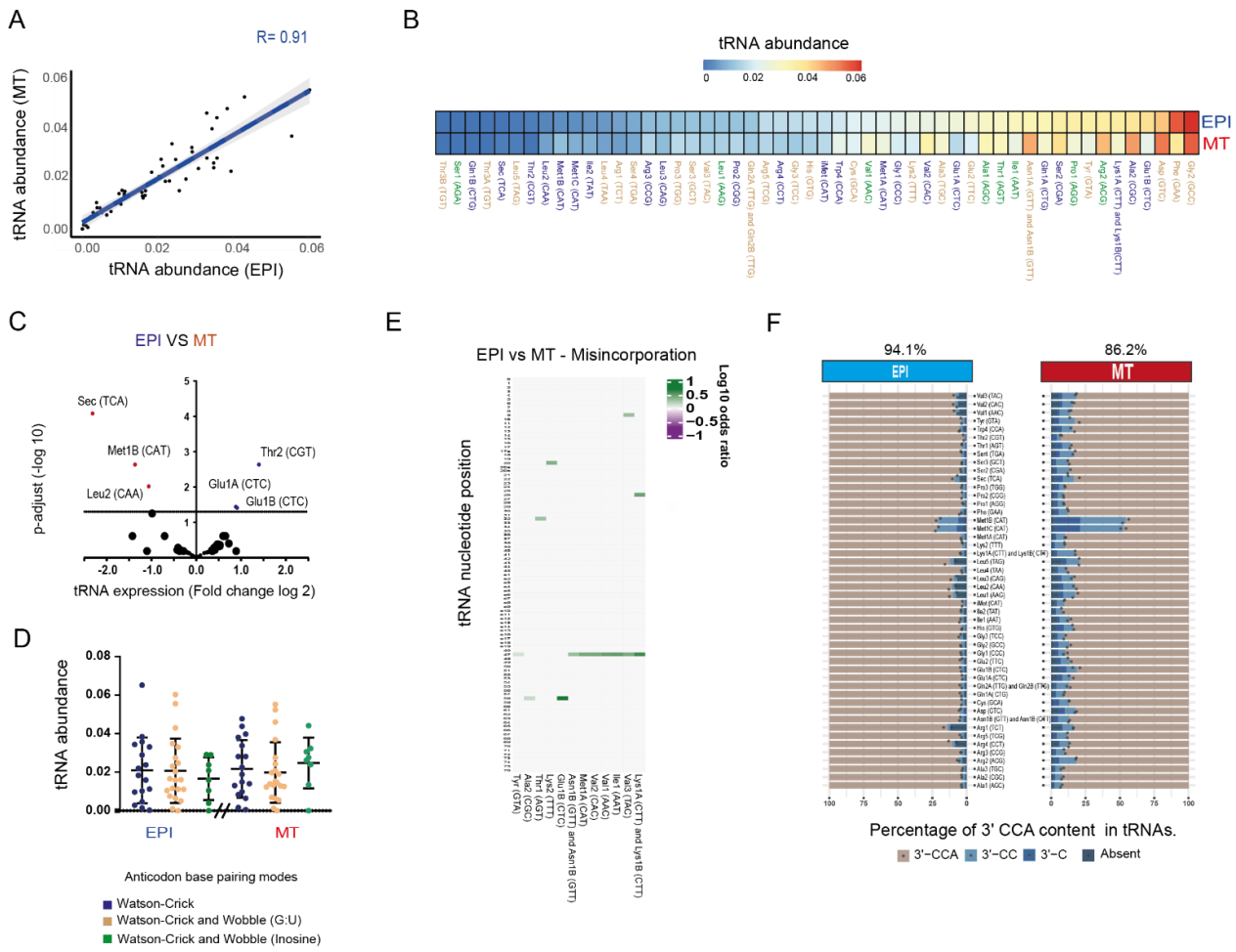


Figure 3. Abundance of tRNA transcripts in EPIs and MTs. **A)** Scatter plot showing Pearson's correlation between tRNA abundance (proportion of mapped reads) in the EPI and MT forms. Gray shading, 95% confidence interval. **B)** Heatmap with tRNA abundance levels normalized by the proportion of mapped reads on the tDNA genome from *T. cruzi*. tRNA anticodons that exhibit only Watson–Crick base pairing with synonymous codons are highlighted in blue. The tRNA anticodons containing either Watson–Crick or wobble base pairing are represented by orange, while those interactions of either Watson–Crick or inosine are illustrated in green. **C)** Volcano plot of fold changes of tRNA expression (Log2) in EPIs compared to MTs. The significance of tRNA differential expression was determined using the DESeq2 algorithm considering a p-adjust value of ≤ 0.05. **D)** Abundance of tRNA species grouped according to their type of base interaction (i. Watson–Crick, ii. Watson–Crick and wobble (G:U), and iii. Watson–Crick and wobble (inosine) with their corresponding codons. Statistic significant tests (p value ≤ 0.05) were performed with the Wilcoxon–Mann–Whitney test (for p values: * = 0.05; ns = not significant). E) Ratio of putative posttranscriptional modification signatures differently expressed (fold change log2≤ 0.25 between modification signatures) in the tRNA nucleotides positions from the EPI and MT forms. Percentage of 3'CCA, 3'CC or 3'C sequences at tRNA molecules in the

EPI and MT forms. A Q-test was performed, and a p value of ≤ 0.01 was considered statistically significant for tRNA maturation and misincorporation ratios across different life forms of *T. cruzi*.

The chromatin profile of tDNAs correlates with the abundance of their transcripts in the epimastigote form

The tDNA copy number exhibits a strong positive correlation with transcript abundance in some eukaryotes (45). Notably, we found only a weak correlation (Spearman's $r = \approx 0.36$) between tRNA transcript abundance and copy number in both the EPI and MT forms, indicating that the tRNA abundance profiles in *T. cruzi* are not solely defined by gene copy numbers (Figure 4A).

We demonstrated that tDNAs from EPI forms exhibited greater levels of open chromatin than did those from their infective forms (47), suggesting a chromatin regulatory role. Here, we investigated whether tDNA chromatin profiles are associated with tRNA abundance. To this end, we used Nanopore, FAIRE-seq and MNase-seq datasets from different *T. cruzi* life forms (47, 48, 53). The Nanopore datasets included information on DNA modifications, such as 5-methylcytosine (m^5C) and 5-hydroxymethylcytosine (hm^5C), which are associated with increased and decreased gene expression, respectively (60, 61). We found that tDNAs are not localized in regions enriched in m^5C and hm^5C modifications compared to the shuffled control (Figure 4B). In addition, the abundance of these DNA modifications was not associated with different tRNA expression levels in either the EPI or infective form (Supplementary Figure S2B).

MNase-seq assesses DNA regions associated with nucleosomes, providing information on nucleosome occupancy and positioning at any given locus (62). We observed lower nucleosome occupancy levels in all canonical tDNAs in both the EPI and infective forms (TCTs), indicating that tDNAs tend to be localized in nucleosome-depleted regions (NDRs) (Figure 4C). However, tDNA loci encoding $tRNA^{Sec}$ showed a different trend. The levels of nucleosome occupancy at the eleven $tDNAs^{Sec}$ did not differ from those of the shuffled control, suggesting that the tDNA Sec loci are localized in nucleosome-enriched regions or closed chromatin.

Next, we used the FAIRE-seq dataset, which detects open chromatin, also known as transcriptionally active chromatin regions (63, 64). We obtained the abundance values of the twenty single-copy tRNA species for comparisons between their expression levels and chromatin status, mitigating the bias likely introduced by multicopy tRNA genes. The tRNAs were classified into three groups (high-, medium- and low-abundance) based on the tRNA abundance obtained by tRNA-seq (Supplementary Figure S2A). We found that tDNAs with high and medium abundances exhibit more open chromatin compared to those with low expression in the EPI forms, suggesting that chromatin

status serves as an important regulator of tRNA expression levels. However, the open chromatin profiles of tDNA did not exhibit a distinct pattern based on the levels of tRNA abundance in *T. cruzi* infective forms, suggesting the presence of different mechanisms regulating tRNA abundance in the infective form of *T. cruzi* (Figure 4D and Supplementary Table 3). Evaluation of nucleosome occupancy levels (MNAse-seq data) retrieved similar conclusions (Figure 4D). Notably, there is no current evidence that tRNAs are transcribed in nonproliferative/infective forms. To gain further insight into this issue, we evaluated the expression of the subunits of RNAP III, which are dedicated to tRNA transcription, in the EPI and MT forms. We detected a downregulation of multiple components of RNAP III in the MT form compared to the EPI form (Figure 4E and Supplementary Table 2F). This observation suggests a lower tRNA transcription activity in *T. cruzi* infective forms, which could account for the similar profile levels of tRNAs from EPIs to MTs.

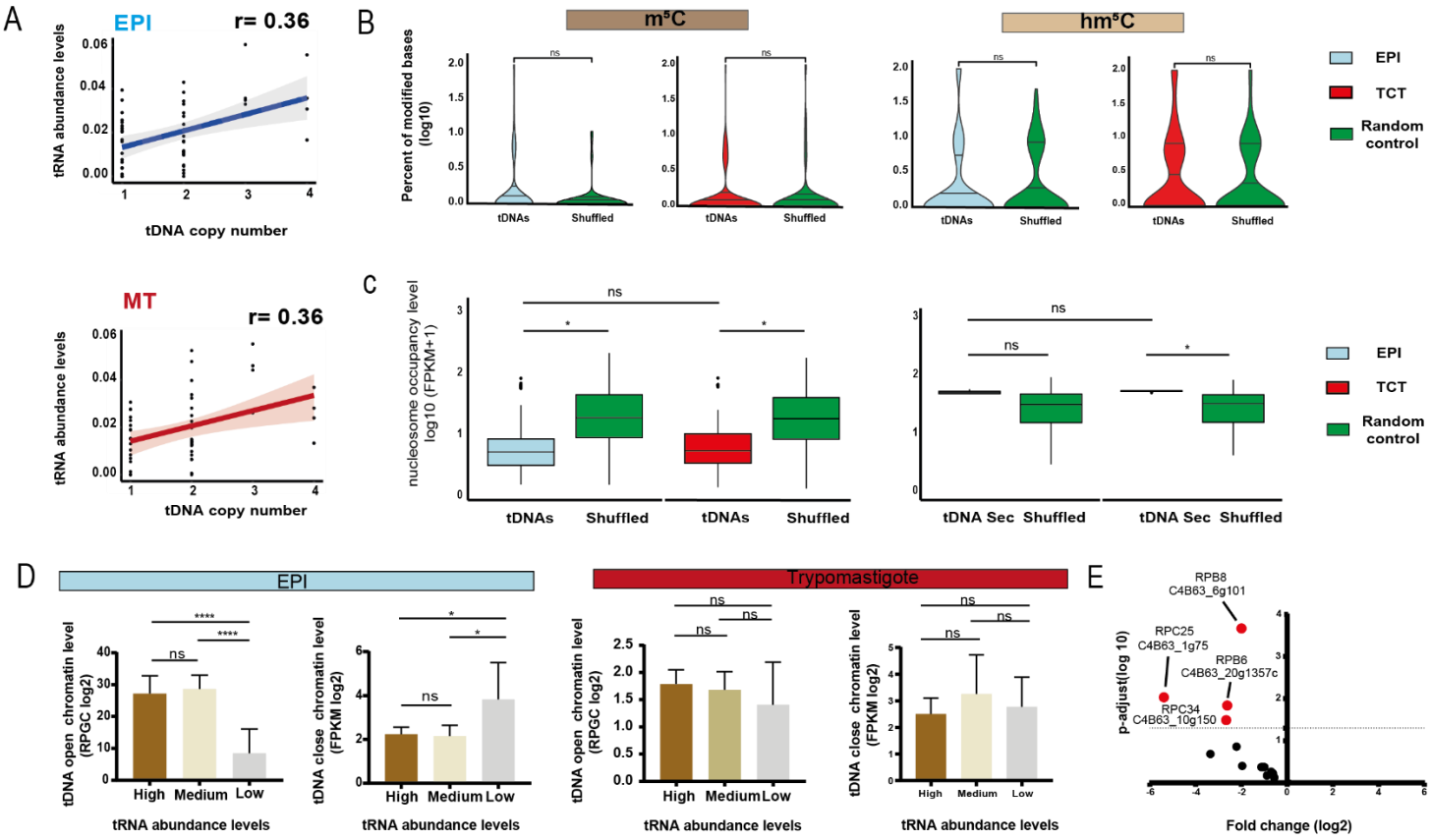


Figure 4. Correlation between tDNA chromatin and tRNA abundance in *T. cruzi* life forms. **A)** Spearman's correlation between tRNA abundance and the number of corresponding tDNA copies in the EPI and MT forms. $tRNA^{Sec}$ is not included in the scatter plot. Red or blue shading, 95% confidence interval. **B)** Percentage of m^5C and hm^5C modifications on tDNA nucleotides present in the EPI and TCT forms. **C)** Nucleosome enrichment in tDNAs and random controls (shuffled) in the EPI and TCT

forms using fragments per kilobase per million mapped reads (FPKM) +1 (log10) values from MNase-seq data. Shuffled was used as random genomic controls. **D)** Bar chart showing the associations between high-, medium- or low-abundance tRNAs and their corresponding tDNA chromatin profiles. The open chromatin levels were normalized to the reads per genomic content (RPGC) from FAIRE-seq data obtained with the EPI and MT forms. The closed chromatin were normalized to the FPKM values from MNAse-seq data obtained with the EPI and TCT forms. Statistical significance tests were performed with the Wilcoxon–Mann–Whitney test (for p values: **** = 0.0001; * = 0.05; ns = not significant). **E)** Volcano plot depicting the fold changes and p values of RNAP III subunits in *T. cruzi* life forms (EPI vs. MT) using Ribo-seq data and DESeq2 algorithm. A p value of ≤ 0.05 and fold change ≥ 1.5 were considered statistically significant.

tRNA abundance coadapts with codon frequencies of highly translated transcripts

We investigated whether tRNA abundance exhibits coadaptation with the corresponding codon frequency. We found a moderate correlation (R=0.63) between codon frequency for all CDSs and tRNA abundance in EPIs (Figure 4A). Notably, core CDSs (R=0.63) exhibited a stronger correlation than did disruptive CDSs (R=0.42). As MTs express more disruptive proteins than EPIs (33–35), we asked whether codon frequency and tRNA abundance are more strongly correlated for disruptive genes in MTs. However, similar to that for EPIs, the correlation between codon frequency for disruptive genes and tRNA abundance was lower than that for core genes (R= 0.33 and R=0.56) (Figure 4A), which is consistent with the data shown in Fig. 3C, where tRNA abundance profiles only slightly changed between EPIs and MTs.

Next, we asked whether the codon frequency of highly expressed transcripts in EPIs or MTs matches their tRNA pools. The genes upregulated in the EPIs compared to the MTs (Figure 4C and Supplementary Table 1B-E) had a similar correlation with all of the other CDSs (R=0.65 vs. R=0.63) (Figure 4A and C). The genes upregulated in MTs compared to EPIs (Figure 4C) also exhibited a similar correlation to all CDSs (R=0.54 vs. R=0.56) (Figure 4A and Table 1B-E), indicating that differentially expressed genes do not have a more optimal codon bias for efficient expression in either form.

Subsequently, we evaluated the association between codon frequency and tRNA abundance for the more abundant and less abundant translated transcripts from the reference set (Supplementary Table 1D-E), which were created to avoid multiple-copy genes and transcripts that exhibited large differences (fold change log2 ≥ 2) in expression between life forms. We found that the correlation was greater for the highly abundant transcripts (top 100) than for the less abundant transcripts (bottom 100) in both EPIs and MTs, suggesting that a regulatory mechanism based on tRNA availability operates

similarly across *T. cruzi* life forms (Figure 4E). However, the correlations between tRNA abundance and the corresponding codon frequencies in MTs were lower than those observed in EPIs.

Finally, we compared the codon usage bias (considering the codon adaptive weights—both the optimal and nonoptimal codons) among all groups of genes mentioned above. We found that codon bias was most pronounced when comparing the core and disruptive compartments (R=0.75), followed by the top 100 and bottom 100 genes (R=0.82) and finally the upregulated genes (R=0.94) (Figure 1H, Figure 4B and 4D).

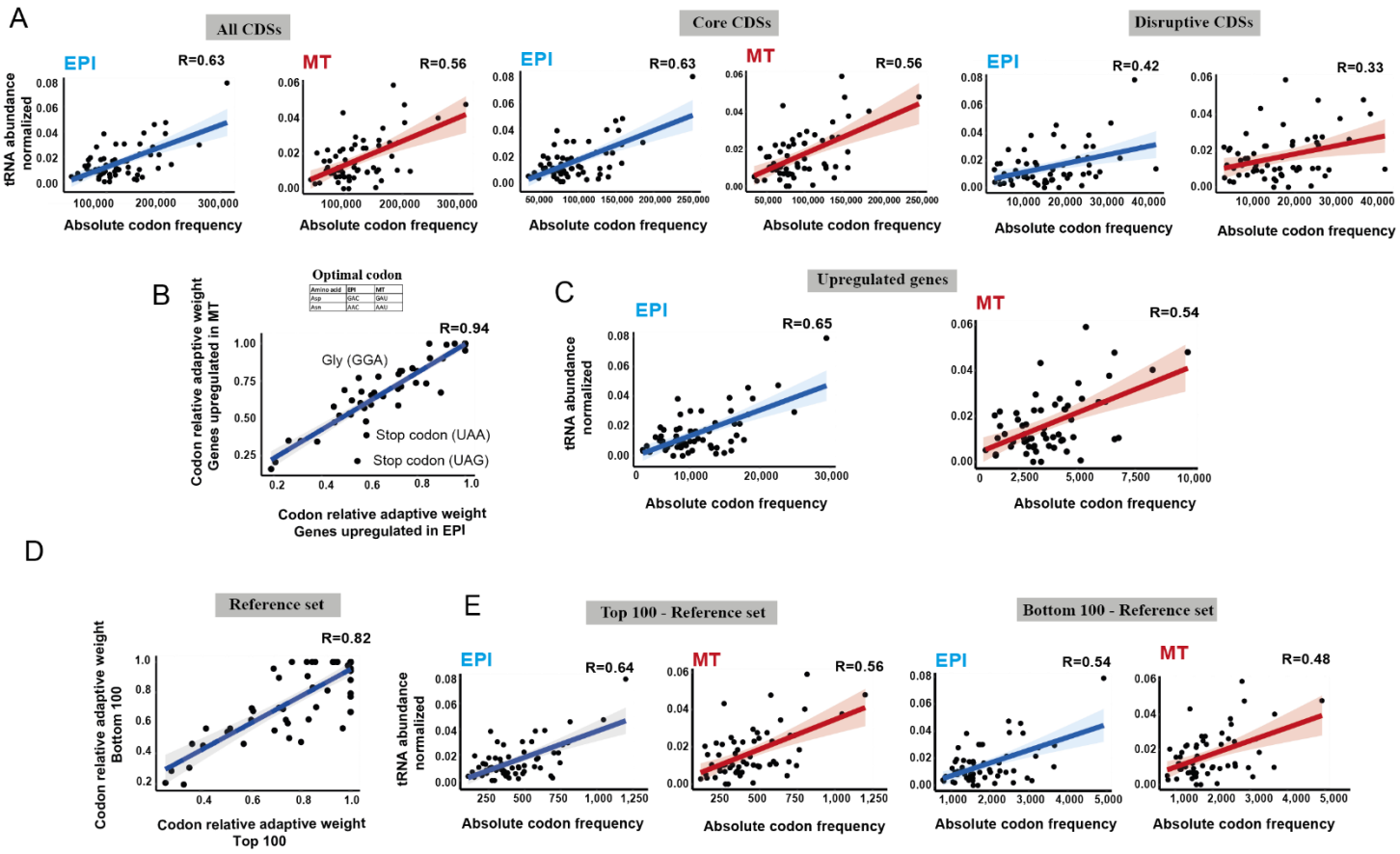


Figure 5. Relationship between tRNA abundance and codon usage frequency in *T. cruzi*. **A)** Pearson's correlation between tRNA abundance in the EPI and MT forms with codon usage frequencies in all CDSs, core genes and disruptive genes. **B)** Pearson's correlation for the relative adaptive weights of codons in the upregulated CDSs for the EPI (1,752 genes) and MT forms (361 genes), obtained with translatome data (Ribo-seq). The distinct optimal codons in the upregulated genes in the EPI and MT forms are illustrated in the table below the scatter plot. The highlighted codons are at least 20% different between the indicated conditions. **C)** Pearson's correlation between tRNA abundance in the EPI and MT forms with their absolute codon frequencies from upregulated genes. **D)** Pearson's correlation for

the relative adaptive weights of codons in the top 100 and bottom 100 genes from a reference dataset containing more and less abundant transcripts translated from *T. cruzi* life forms. **E)** Pearson's correlation between tRNA abundance in the EPI and MT forms with codon frequencies of higher (top 100) and lower (bottom 100) translated transcripts. The expression level was based on transcripts per million (TPM) values from Ribo-seq data. Shading- 95% confidence interval.

Base pairing modes between codons and anticodons are linked to translated transcript levels

To further understand the observation that highly expressed transcripts (top 100) are more strongly correlated with tRNA abundance than are less expressed transcripts (bottom 100) (Figure 5E), we performed a detailed analysis of the codon composition of these genes. We identified forty-six tRNA anticodons and sixty-two tRNA codons in the *T. cruzi* genome, implying that some tRNA transcripts must recognize more than one codon via tRNA modification (inosine) or wobble base pairing (G-U) (Figure 2D). We found nine distinct optimal codons between the top 100 or bottom 100 transcripts (Figure 6A). With the exception of Ala (GCC/GCG), all of these residues are recognized by the same set of tRNA species. Strikingly, six (out of nine) optimal codons from the top 100 abundant transcripts interact via Watson–Crick base pairing only, while six (out of nine) optimal codons from the bottom 100 form only wobble base pairing (Figure 6B and C). This observation suggested that the anticodon:codon base pairing mode is associated with the expression level of the transcripts in *T. cruzi.* Preferential usage of codons forming Watson-Crick pairing in highly expressed transcripts were also observed in other organism, such as *Plasmodium falciparum* (65). Such skewed synonymous codon usage can be a common strategy to optimize translation among parasites.

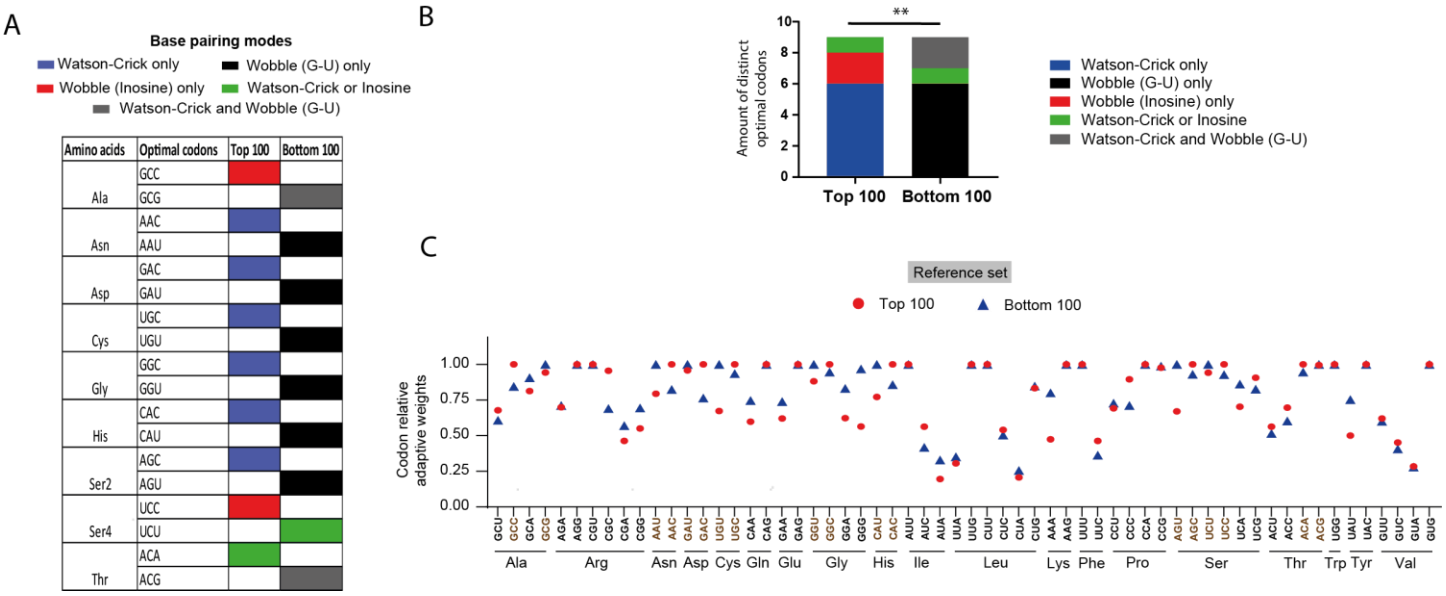


Figure 6. Optimal codons and anticodon:codon pairing modes in the top 100 and bottom 100 genes in *T. cruzi*. **A)** Distinct optimal codons between the highest 100 (top 100) and lowest 100 (bottom 100) translated transcripts in the EPI and MT forms. **B)** Distribution of the anticodon:codon base pairing modes (**i.**Watson–Crick only, **ii.** wobble (G-U) only, **iii.** wobble (inosine) only, **vi.** Watson–Crick and wobble (inosine) and Watson–Crick and wobble (G:U)) were present in distinct optimal codons between the top 100 and bottom 100 genes. Chi-square - p value ≤ 0.01 (**) was considered. **C)** Codon relative adaptive weights compared to optimal codons from the top 100 and bottom 100 genes.

Discussion

Codon usage bias has been proposed to be a significant factor directing transcript and protein expression levels in trypanosomes (6, 37). Codon frequency is linked to the abundance of tRNA pools in some species, but measuring tRNA expression levels has been only a recent advancement (45). Here, we sequenced tRNAs in two different developmental forms of *T. cruzi*, which enabled direct comparisons between tRNA abundance and codon usage biases. We observed a stronger correlation between codon usage and tRNA abundance than between the copy number of tRNA genes (Figure 4A), highlighting the importance of obtaining actual tRNA expression profiles to analyze the relationship between codon bias and tRNA supply. Surprisingly, tRNA abundance profiles were very similar in EPIs and MTs (Figure 3C), although these two forms have distinct gene expression profiles (33, 35). tRNA abundances correlated with the core compartment of the genome and highly expressed mRNA transcripts in both life forms (Figure 5A, C and E).

The transcription of tRNAs relies on RNA Pol III and interactions between transcription factors and promoters within tRNA loci. This process is facilitated by the depletion of nucleosomes at these loci (24, 66). We showed that tDNAs are in regions that are lacking in nucleosomes in both the EPI and TCT forms, except for $tDNA^{Sec}$ (Figure 4C). Consequently, these results indicate that the chromatin status can influence tRNA transcription (Figure 4D). Notably, this correlation was exclusive to EPI forms (Figure 4D). In trypomastigote forms (both TCTs and MTs), tRNA expression appears to be independent of chromatin status, suggesting that other regulatory mechanisms maintain the tRNA abundance profiles in MTs similar to those in the EPI form. Considering the presence of NDRs in the tDNA of MTs (as observed in TCT forms), it is plausible that chromatin closure in MTs (as observed in FAIRE-seq assays) (47) may be attributable to the deposition of proteins associated with chromatin compaction but due to nucleosome deposition. Future investigations are warranted to validate this hypothesis.

Another possible explanation for the similar tRNA abundance profiles in the MT and EPI forms is that de novo tRNA synthesis is not active in the MT form, resulting in maintenance of the same tRNA profiles as in the EPI form. The low amount of mature tRNAs (per parasite) in MT forms (47), alongside a downregulation of the RNAP III machinery in these forms (Figure 4E), suggests that tRNA synthesis is not active in MTs, while nascent expression via RNA Pol III has not yet been assessed in *T. cruzi*. Pol III RNA expression can also be arrested by Maf1, which was described in *T. brucei* to inhibit RNAP III transcription in procyclic forms (67). Since Maf1 is known to bind to RNAP III under stress conditions in some organisms (66), we hypothesized that a similar mechanism may also play a role during *T. cruzi* differentiation.

To assess the effect of codon bias on *T. cruzi* gene expression, we evaluated genes from the core and disruptive regions and genes with higher and lower abundances. Genes in the disruptive regions showed increased expression in *T. cruzi*'s infective forms (33, 35) but comprised only three gene classes, while the core compartment encompassed all of the remaining CDS (36). Therefore, we expected these two compartments to differ greatly in codon usage, which was confirmed in the present study (Figure 1H). Additionally, we also found nine pairs of codons that were optimized differently among the most and least abundant transcripts (Figure 6A). Together, these results indicate the presence of codon bias associated with both gene expression and genomic compartments in *T. cruzi* life forms. Notably, we found that the most and least abundant transcripts use codons that are more or less adapted to their tRNA pool (Figure 5E), suggesting that the tRNA abundance profile is adapted to codon usage in highly expressed genes.

We also found that the stop codon UGA is prevalent in disruptive genes, while the core and GpDR compartments contain a homogeneous distribution of the three stop codons present in the *T. cruzi* genome (Figure 1E). The enrichment of the UGA stop codon in disruptive genes could provide a significant advantage for rapid upregulation of these genes observed in MTs compared to EPIs, as UGA exhibits greater stop translation efficiency than UAA and UAG in eukaryotes (68–71). In addition, the UGA codon can be recognized by $tRNA^{Sec}$ if a SECIS element is present downstream of the mRNA (72, 73). Interestingly, we found a fourfold increase in the level of tRNA-Sec in MTs (nonproliferative cells) (Figure 4C). Among the disruptive genes enriched in UGA, we found that only 1.4% of genes (mainly TSs) had putative SECIS elements (data not shown), indicating that in the majority of cases, UGA at disruptive genes is indeed associated with a stop in translation and not with the insertion of a selenocysteine amino acid.

Our results suggest that tRNA abundance may be less coadapted to gene expression in MTs than in EPIs. However, this coadaptation can also be further modulated by the presence of PTMs in tRNAs. In this regard, we found that tRNAs from MTs contain fewer putative PTMs, mainly those located within the tRNA body (Figure 3E), which can affect stability (20–22). Thus, it is plausible that in MTs, the coadaptation of tRNAs and codon frequency may be even lower. Evaluating the full repertoire of PTMs in tRNAs of different life forms will shed light on the role of these PTMs in tRNA biology and gene expression in *T. cruzi*.

Trypanosomatids have fewer copies of tRNA genes (fewer isodecoders) (Figure 2A) than vertebrates (74, 75), suggesting that these genes may be under selective pressure to avoid deleterious mutations. Here, we detected forty-six anticodons for sixty-two codons (Figure 2D and Supplementary Table 2B), which contrasts with findings in humans, where there are fifty-seven anticodons for sixty-

one codons (76, 77). This indicates that the parasite needs to employ more strategies to recognize the entire codon repertoire of the genome, such as employing more modifications in anticodons and codon–anticodon wobble pairing. In this sense, we found that *T. cruzi* may take advantage of another layer of gene expression regulation based on anticodon:codon pairing modes. We found that the most abundant genes from each life form predominantly use codons that utilize Watson–Crick base pairing (Figure 6A–C). In contrast, the least abundant transcripts had greater frequencies of wobble-pairing codon–anticodon interactions. tRNA pairing affects translation efficiency, with the latter type being associated with lower efficiency (65, 78). Future studies should experimentally assess whether the expression of a protein is indeed affected by swapping codons recognized by different pairing modes. In brief, we propose that *T. cruzi* protein expression in EPIs and MTs may be affected by a combination of codon usage bias, tRNA abundance and anticodon:codon pairing mode. Subsequent studies should explore the potential impact of perturbed tRNA biology on parasite viability, thereby offering promising avenues for novel disease-control strategies, particularly in the realm of tRNA-based therapeutics.

Supplementary information

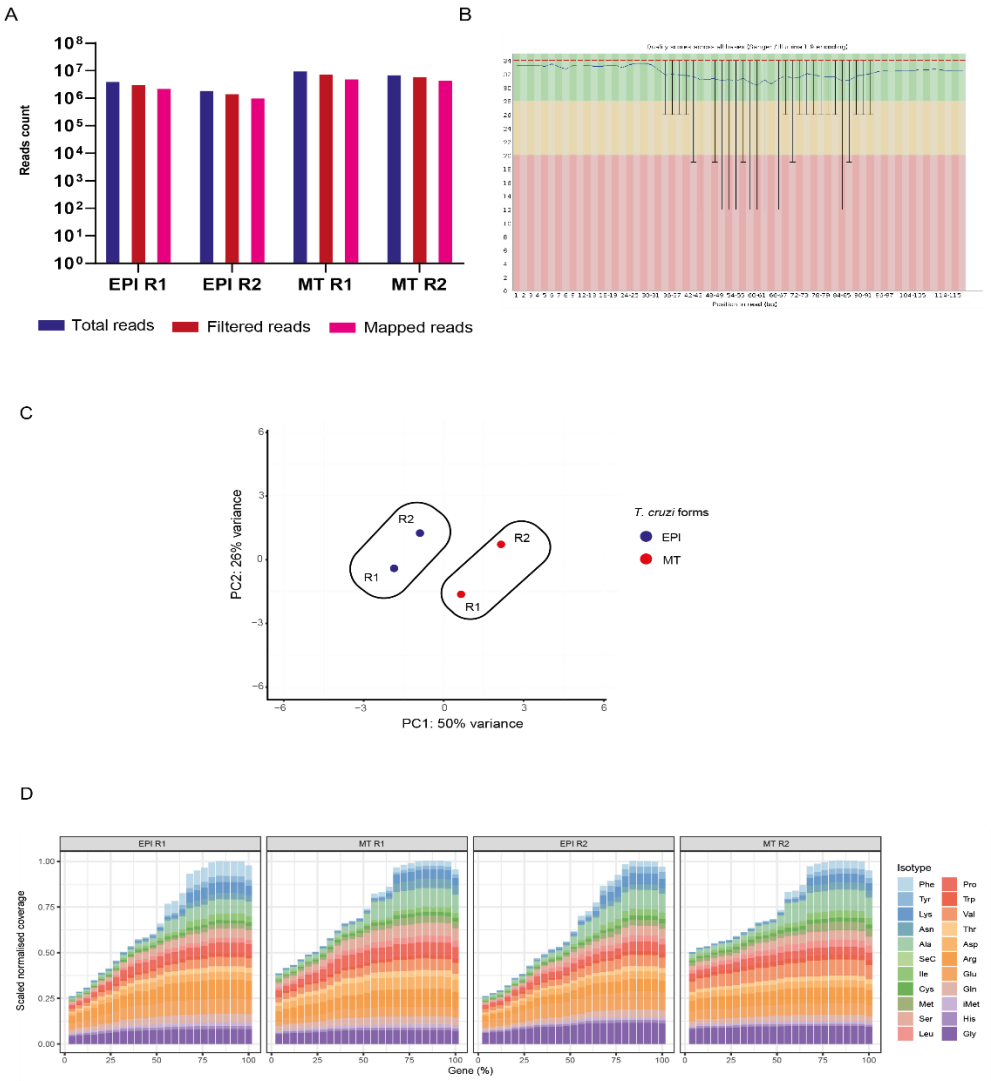


Figure S1. Quantity of reads and sequencing data quality for tRNA-seq performed in the EPI and MT forms. **A)** Number of reads sequenced, filtered and mapped on tDNAs from *T. cruzi* Dm28c. **B)** Quality of base pair sequences, assessed by Phred scores, of a representative sample (MT R1). **C)** PCA plot of tRNA-seq-normalized samples with the varianceStabilizingTransformation tool from the DESeq2 package. R1= replicate 1; R2= replicate 2. **D)** Metagene analysis of the normalized sequence coverage of tRNA isotypes.

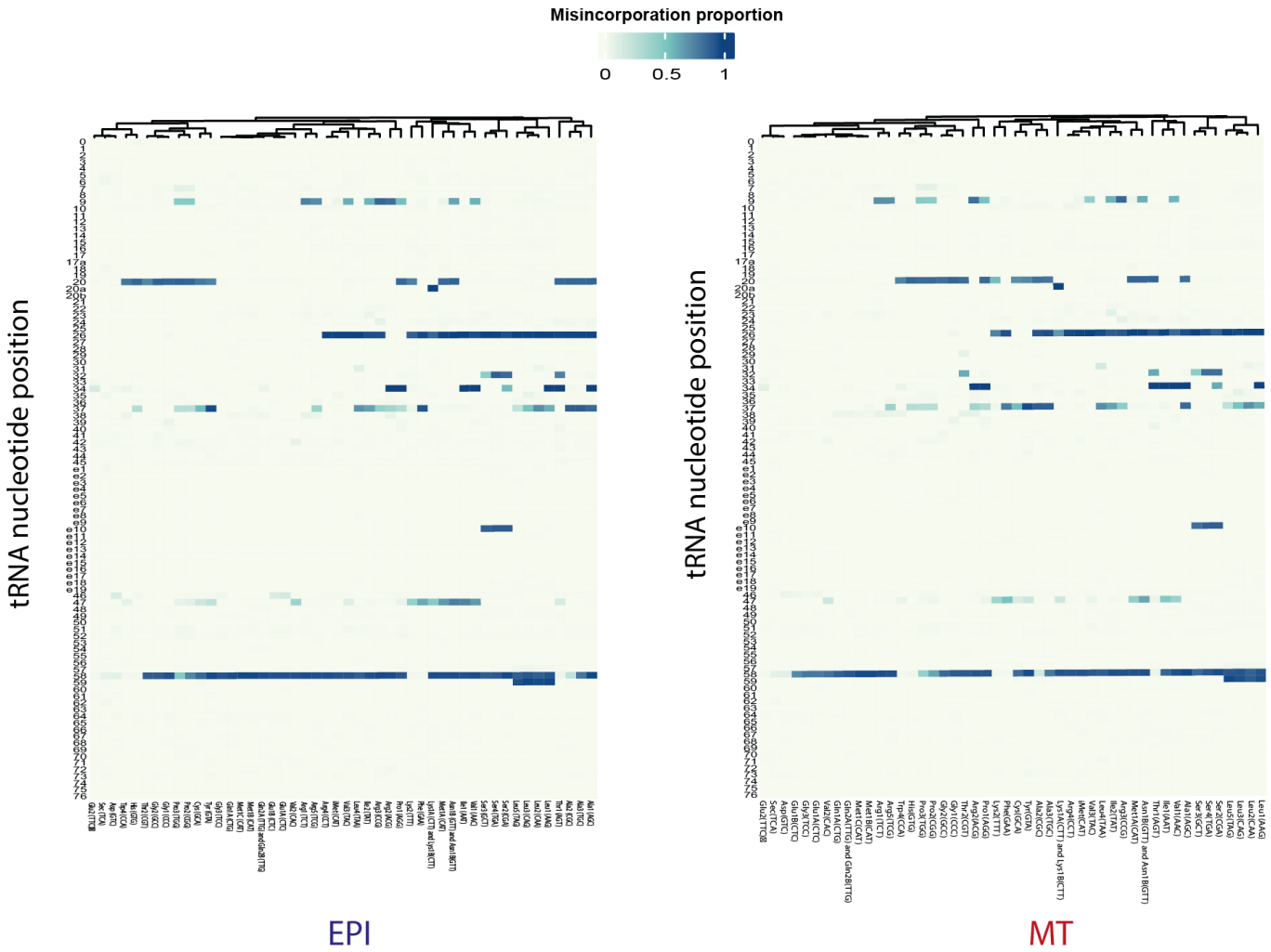


Figure S2. Reverse transcriptase signatures (RT-signatures) in tRNA species from the EPI and MT forms. Heatmap containing the misincorporation proportion values for each tRNA nucleotide position from *T. cruzi*.

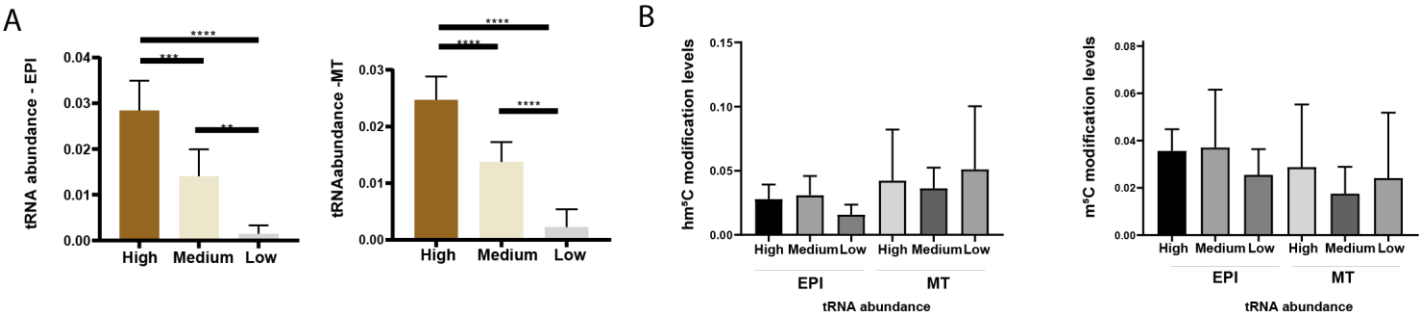


Figure S3. Abundance of twenty tRNA transcripts of a single copy in *T. cruzi.* **A)** Bar charts depicting the seven tRNAs with the highest abundance, the seven with medium expression, and the remaining six with the lowest abundance in either the EPI or MT form. **B)** Relationships between tRNA abundance levels (high, medium and low) and hm^5C or m^5C modification levels in twenty single-copy tDNAs in the EPI and MT forms. Statistical significance tests were performed with the Wilcoxon–Mann–Whitney test. (ns = not significant).

ADDITIONAL FILES

Table 1. Codon usage in the *T. cruzi* genome. **A)** Codon frequencies, RSCU and codon relative adaptive weight in the core, disruptive and GpDR compartments. **B)** TPM values for all genes in the EPI and MT forms obtained from the translatome data (Ribo-seq). **C)** DESeq2 results from Ribo-seq data performed in the EPI and MT forms. **D)** Gene IDs of the reference set (top 100 and bottom 100) and genes upregulated in the EPI and MT forms. E) Analysis of the distribution of optimal codons, codon frequencies, and codon adaptive relative weight for either the top 100 or bottom 100 genes in a reference set, as well as for the dataset of genes upregulated in the EPI and MT forms.

Table 2. Characterization of tDNA and tRNA pools in *T. cruzi* Dm28c. **A)** Genomic coordinates of tDNAs, gene IDs, sense strands and copy numbers. **B)** Identification of tRNA anticodons and their base pairing types with corresponding codons in the *T. cruzi* genome**. C)** tRNA abundance of replicates (R1 and R2) and their averages in both the EPI and MT forms. **D)** DESeq results of tRNA abundance in EPIs compared to the MTs. **E)** Correlations between tRNA abundance and base pairing types with their corresponding codons. **F)** BLAST and HMMER results containing the best hit of the RNAP III subunit candidates present in the *T. cruzi* genome.

Table 3. Abundance of single-copy tRNAs and their levels of open (FAIRE-seq) and closed (MNase-seq) chromatin in the EPI and MT forms.

Author Contributions

Silva, H.G. and Cunha, J.P.C. contributed to the writing, analysis, and rationale of this article. Kimura, S. and Waldor, K. M. assisted in executing the tRNA sequencing. Lima, P.L.C. obtained the RPKM values from the MNase-seq data, and Pires, D obtained the tDNA modification data. All authors contributed to the revision of this work.

Acknowledgments

We thank Ivan Novaski Avino and Karin Navarro for technical assistance and Dr. Alex Ranieri Lima for important input on the data and bioinformatic analysis. We thank Dr. Janaina de Freitas Nascimento and Dr. Mark Carrington for their valuable discussions on codon usage data. We thank Leticia de Sousa Lopes for reviewing this manuscript.

Data availability

The t-RNA-seq, MNAse-seq, and FAIRE-seq data can be accessed at the Sequence Read Archive (SRA) (https://www.ncbi.nlm.nih.gov/sra) under the following accession numbers: PRJNA1124437, PRJNA665060 and PRJNA763084. Oxford Nanopore sequencing datasets are available at(28).

Funding

This work was supported by the São Paulo Research Foundation (FAPESP) [#13/07467-1, #18/15553-9, #21/11419-9].

Conflict of interest

The authors declare no competing interests.

References

1. Crick,F.H.C. (1968) The origin of the genetic code. *J Mol Biol*, **38**, 367–379.

2. Bahiri-Elitzur,S. and Tuller,T. (2021) Codon-based indices for modeling gene expression and transcript evolution. *Comput Struct Biotechnol J*, **19**, 2646–2663.

3. Rudolph,K.L.M., Schmitt,B.M., Villar,D., White,R.J., Marioni,J.C., Kutter,C. and Odom,D.T. (2016) Codon-Driven Translational Efficiency Is Stable across Diverse Mammalian Cell States. *PLoS Genet*, **12**, e1006024.

4. Hershberg,R. and Petrov,D.A. (2009) General Rules for Optimal Codon Choice. *PLoS Genet*, **5**, e1000556.

5. Sharp,P.M. and Li,W.-H. (1987) The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res*, **15**, 1281–1295.

6. Jeacock,L., Faria,J. and Horn,D. (2018) Codon usage bias controls mRNA and protein abundance in trypanosomatids. *Elife*, **7**.

7. Gingold,H., Dahan,O. and Pilpel,Y. (2012) Dynamic changes in translational efficiency are deduced from codon usage of the transcriptome. *Nucleic Acids Res*, **40**, 10053–10063.

8. Liu,Y. (2020) A code within the genetic code: codon usage regulates co-translational protein folding. *Cell Communication and Signaling*, **18**, 145.

9. Hanson,G. and Coller,J. (2018) Codon optimality, bias and usage in translation and mRNA decay. *Nat Rev Mol Cell Biol*, **19**, 20–30.

10. Rak,R., Polonsky,M., Eizenberg-Magar,I., Mo,Y., Sakaguchi,Y., Mizrahi,O., Nachshon,A., Reich-Zeliger,S., Stern-Ginossar,N., Dahan,O., *et al.* (2021) Dynamic changes in tRNA modifications and abundance during T cell activation. *Proceedings of the National Academy of Sciences*, **118**.

11. Sabi,R. and Tuller,T. (2014) Modelling the Efficiency of Codon–tRNA Interactions Based on Codon Usage Bias. *DNA Research*, **21**, 511–526.

12. Warner,J.R. and Rich,A. (1964) THE NUMBER OF SOLUBLE RNA MOLECULES ON RETICULOCYTE POLYRIBOSOMES. *Proceedings of the National Academy of Sciences*, **51**, 1134–1141.

13. J D WATSON (1964) THE SYNTHESIS OF PROTEINS UPON RIBOSOMES. *Bull Soc Chim Biol*, **46**, 1399–1425.

14. Cleaves,H.J. (2011) Watson–Crick Pairing. In *Encyclopedia of Astrobiology*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1775–1776.

15. WATSON,J.D. and CRICK,F.H.C. (1953) Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, **171**, 737–738.

16. Varani,G. and McClain,W.H. (2000) The G·U wobble base pair. *EMBO Rep*, **1**, 18–23.

17. Crick,F.H.C. (1966) Codon—anticodon pairing: The wobble hypothesis. *J Mol Biol*, **19**, 548–555.

18. Agris,P.F., Eruysal,E.R., Narendran,A., Väre,V.Y.P., Vangaveti,S. and Ranganathan,S. V. (2018) Celebrating wobble decoding: Half a century and still much is new. *RNA Biol*, **15**, 537–553.

19. Kato,M., Nishikawa,K., Uritani,M., Miyazaki,M. and Takemura,S. (1990) The Difference in the Type of Codon-Anticodon Base Pairing at the Ribosomal P-Site Is One of the Determinants of the Translational Rate1. *The Journal of Biochemistry*, **107**, 242–247.

20. Ohira,T. and Suzuki,T. (2024) Transfer RNA modifications and cellular thermotolerance. *Mol Cell*, **84**, 94–106.

21. Yared,M.-J., Marcelot,A. and Barraud,P. (2024) Beyond the Anticodon: tRNA Core Modifications and Their Impact on Structure, Translation and Stress Adaptation. *Genes (Basel)*, **15**, 374.

22. Cui,W., Zhao,D., Jiang,J., Tang,F., Zhang,C. and Duan,C. (2023) tRNA Modifications and Modifying Enzymes in Disease, the Potential Therapeutic Targets. *Int J Biol Sci*, **19**, 1146–1162.

23. Mahapatra,S., Dewari,P.S., Bhardwaj,A. and Bhargava,P. (2011) Yeast H2A.Z, FACT complex and RSC regulate transcription of tRNA gene through differential dynamics of flanking nucleosomes. *Nucleic Acids Res*, **39**, 4023–4034.

24. Shukla,A. and Bhargava,P. (2018) Regulation of tRNA gene transcription by the chromatin structure and nucleosome dynamics. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, **1861**, 295–309.

25. Torrent,M., Chalancon,G., de Groot,N.S., Wuster,A. and Madan Babu,M. (2018) Cells alter their tRNA abundance to selectively regulate protein synthesis during stress conditions. *Sci Signal*, **11**.

26. Pinkard,O., McFarland,S., Sweet,T. and Coller,J. (2020) Quantitative tRNA-sequencing uncovers metazoan tissue-specific tRNA regulation. *Nat Commun*, **11**, 4104.

27. Gao,L., Behrens,A., Rodschinka,G., Forcelloni,S., Wani,S., Strasser,K. and Nedialkova,D.D. (2024) Selective gene expression maintains human tRNA anticodon pools during differentiation. *Nat Cell Biol*, **26**, 100–112.

28. Du,M.-Z., Wei,W., Qin,L., Liu,S., Zhang,A.-Y., Zhang,Y., Zhou,H. and Guo,F.-B. (2017) Co-adaption of tRNA gene copy number and amino acid usage influences translation rates in three life domains. *DNA Research*, **24**, 623–633.

29. Iben,J.R. and Maraia,R.J. (2012) tRNAomics: tRNA gene copy number variation and codon use provide bioinformatic evidence of a new anticodon:codon wobble pair in a eukaryote. *RNA*, **18**, 1358–1372.

30. Liu,C., Yuan,J., Zhang,X., Jin,S., Li,F. and Xiang,J. (2021) tRNA copy number and codon usage in the sea cucumber genome provide insights into adaptive translation for saponin biosynthesis. *Open Biol*, **11**.

31. Clayton,C. (2019) Regulation of gene expression in trypanosomatids: living with polycistronic transcription. *Open Biol*, **9**.

32. de Souza,W., de Carvalho,T.M.U. and Barrias,E.S. (2010) Review on *Trypanosoma cruzi* : Host Cell Interaction. *Int J Cell Biol*, **2010**, 1–18.

33. Smircich,P., Eastman,G., Bispo,S., Duhagon,M.A., Guerra-Slompo,E.P., Garat,B., Goldenberg,S., Munroe,D.J., Dallagiovanna,B., Holetz,F., *et al.* (2015) Ribosome profiling reveals translation control as a key mechanism generating differential gene expression in Trypanosoma cruzi. *BMC Genomics*, **16**, 443.

34. Freire-de-Lima,L., Fonseca,L., Oeltmann,T., Mendonça-Previato,L. and Previato,J. (2015) The *trans* -sialidase, the major *Trypanosoma cruzi* virulence factor: Three decades of studies. *Glycobiology*, **25**, 1142–1149.

35. de Godoy,L.M.F., Marchini,F.K., Pavoni,D.P., Rampazzo,R. de C.P., Probst,C.M., Goldenberg,S. and Krieger,M.A. (2012) Quantitative proteomics of *Trypanosoma cruzi* during metacyclogenesis. *Proteomics*, **12**, 2694–2703.

36. Berná,L., Rodriguez,M., Chiribao,M.L., Parodi-Talice,A., Pita,S., Rijo,G., Alvarez-Valin,F. and Robello,C. (2018) Expanding an expanded genome: long-read sequencing of Trypanosoma cruzi. *Microb Genom*, **4**.

37. de Freitas Nascimento,J., Kelly,S., Sunter,J. and Carrington,M. (2018) Codon choice directs constitutive mRNA levels in trypanosomes. *Elife*, **7**.

38. Padilla-Mejía,N.E., Florencio-Martínez,L.E., Figueroa-Angulo,E.E., Manning-Cela,R.G., Hernández-Rivas,R., Myler,P.J. and Martínez-Calvillo,S. (2009) Gene organization and sequence analyses of transfer RNA genes in Trypanosomatid parasites. *BMC Genomics*, **10**, 232.

39. Contreras,V.T., Salles,J.M., Thomas,N., Morel,C.M. and Goldenberg,S. (1985) In vitro differentiation of Trypanosoma cruzi under chemically defined conditions. *Mol Biochem Parasitol*, **16**, 315–327.

40. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J Mol Biol*, **215**, 403–410.

41. Chan,P.P., Lin,B.Y., Mak,A.J. and Lowe,T.M. (2021) tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res*, **49**, 9077–9096.

42. Kimura,S., Dedon,P.C. and Waldor,M.K. (2020) Comparative tRNA sequencing and RNA mass spectrometry for surveying tRNA modifications. *Nat Chem Biol*, **16**, 964–972.

43. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*, **17**, 10.

44. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

45. Behrens,A., Rodschinka,G. and Nedialkova,D.D. (2021) High-resolution quantitative profiling of tRNA abundance and modification status in eukaryotes by mim-tRNAseq. *Mol Cell*, **81**, 1802-1815.e7.

46. Mistry,J., Finn,R.D., Eddy,S.R., Bateman,A. and Punta,M. (2013) Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res*, **41**, e121–e121.

47. Lima,A.R.J., Silva,H.G. de S., Poubel,S., Rosón,J.N., de Lima,L.P.O., Costa-Silva,H.M., Gonçalves,C.S., Galante,P.A.F., Holetz,F., Motta,M.C.M., *et al.* (2022) Open chromatin analysis

in Trypanosoma cruzi life forms highlights critical differences in genomic compartments and developmental regulation at tDNA loci. *Epigenetics Chromatin*, **15**, 22.

48. Lima,A.R.J., de Araujo,C.B., Bispo,S., Patané,J., Silber,A.M., Elias,M.C. and da Cunha,J.P.C. (2021) Nucleosome landscape reflects phenotypic differences in Trypanosoma cruzi life forms. *PLoS Pathog*, **17**, e1009272.

49. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*, **9**, 357–359.

50. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

51. Pongor,L.S., Gross,J.M., Vera Alvarez,R., Murai,J., Jang,S.-M., Zhang,H., Redon,C., Fu,H., Huang,S.-Y., Thakur,B., *et al.* (2020) BAMscale: quantification of next-generation sequencing peaks and generation of scaled coverage tracks. *Epigenetics Chromatin*, **13**, 21.

52. Lima,P.L.C., Lopes,L. de S., Rosón,J.N., Borges,A., Bellini,N.K., Tahira,A., da Silva,M.S., Pires,D., Elias,M.C. and Cunha,J.P.C. (2024) Comprehensive Analysis of Nascent Transcriptome Reveals Diverse Transcriptional Profiles Across the Trypanosoma cruzi Genome Underlining the Regulatory Role of Genome Organization, Chromatin Status, and Cis-Acting Elements. *bioRxiv*.

53. Díaz-Viraqué,F., Chiribao,M.L., Libisch,M.G. and Robello,C. (2023) Genome-wide chromatin interaction map for Trypanosoma cruzi. *Nat Microbiol*, **8**, 2103–2114.

54. Li,H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**, 3094–3100.

55. Bray,N.L., Pimentel,H., Melsted,P. and Pachter,L. (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*, **34**, 525–527.

56. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, **15**, 550.

57. Li,Q., Luo,Y., Sha,A., Xiao,W., Xiong,Z., Chen,X., He,J., Peng,L. and Zou,L. (2023) Analysis of synonymous codon usage patterns in mitochondrial genomes of nine Amanita species. *Front Microbiol*, **14**.

58. Viraqué,F.D., Ehrlich,R. and Robello,C. Genomic organization of Trypanosoma cruzi tRNA genes. https://doi.org/10.1101/2023.12.15.571829.

59. Schwartz,S. and Motorin,Y. (2017) Next-generation sequencing technologies for detection of modified nucleotides in RNAs. *RNA Biol*, **14**, 1124–1137.

60. Shi,D.-Q., Ali,I., Tang,J. and Yang,W.-C. (2017) New Insights into 5hmC DNA Modification: Generation, Distribution and Function. *Front Genet*, **8**.

61. Hummel,G., Berr,A., Graindorge,S., Cognat,V., Ubrig,E., Pflieger,D., Molinier,J. and Drouard,L. (2020) Epigenetic silencing of clustered tRNA genes in Arabidopsis. *Nucleic Acids Res*, **48**, 10297–10312.

62. Chereji,R. V., Bryson,T.D. and Henikoff,S. (2019) Quantitative MNase-seq accurately maps nucleosome occupancy levels. *Genome Biol*, **20**, 198.

63. Simon,J.M., Giresi,P.G., Davis,I.J. and Lieb,J.D. (2012) Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. *Nat Protoc*, **7**, 256–267.

64. Giresi,P.G., Kim,J., McDaniell,R.M., Iyer,V.R. and Lieb,J.D. (2007) FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res*, **17**, 877–885.

65. Chan,S., Ch'ng,J.-H., Wahlgren,M. and Thutkawkorapin,J. (2017) Frequent GU wobble pairings reduce translation efficiency in Plasmodium falciparum. *Sci Rep*, **7**, 723.

66. Graczyk,D., Cieśla,M. and Boguta,M. (2018) Regulation of tRNA synthesis by the general transcription factors of RNA polymerase III - TFIIIB and TFIIIC, and by the MAF1 protein. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, **1861**, 320–329.

67. Romero-Meza,G., Vélez-Ramírez,D.E., Florencio-Martínez,L.E., Román-Carraro,F.C., Manning-Cela,R., Hernández-Rivas,R. and Martínez-Calvillo,S. (2017) Maf1 is a negative regulator of transcription in Trypanosoma brucei. *Mol Microbiol*, **103**, 452–468.

68. Trexler,M., Bányai,L., Kerekes,K. and Patthy,L. (2023) Evolution of termination codons of proteins and the TAG-TGA paradox. *Sci Rep*, **13**, 14294.

69. Belin,D. and Puigbò,P. (2022) Why Is the UAG (Amber) Stop Codon Almost Absent in Highly Expressed Bacterial Genes? *Life*, **12**, 431.

70. Anzalone,A. V., Zairis,S., Lin,A.J., Rabadan,R. and Cornish,V.W. (2019) Interrogation of Eukaryotic Stop Codon Readthrough Signals by *in Vitro* RNA Selection. *Biochemistry*, **58**, 1167–1178.

71. Poole,E.S., Brown,C.M. and Tate,W.P. (1995) The identity of the base following the stop codon determines the efficiency of in vivo translational termination in Escherichia coli. *EMBO J*, **14**, 151–158.

72. Berry,M.J., Banu,L., Harney,J.W. and Larsen,P.R. (1993) Functional characterization of the eukaryotic SECIS elements which direct selenocysteine insertion at UGA codons. *EMBO J*, **12**, 3315–3322.

73. Mariotti,M., Lobanov,A. V., Guigo,R. and Gladyshev,V.N. (2013) SECISearch3 and Seblastian: new tools for prediction of SECIS elements and selenoproteins. *Nucleic Acids Res*, **41**, e149–e149.

74. Goodenbour,J.M. and Pan,T. (2006) Diversity of tRNA genes in eukaryotes. *Nucleic Acids Res*, **34**, 6137–6146.

75. Santos,F.B. and Del-Bem,L.-E. (2022) The Evolution of tRNA Copy Number and Repertoire in Cellular Life. *Genes (Basel)*, **14**, 27.

76. Iben,J.R. and Maraia,R.J. (2014) tRNA gene copy number variation in humans. *Gene*, **536**, 376–384.

77. Kanduc,D. (2021) Human Codon Usage: The Genetic Basis of Pathogen Latency. *Glob Med Genet*, **08**, 109–115.

78. Stadler,M. and Fire,A. (2011) Wobble base-pairing slows in vivo translation elongation in metazoans. *RNA*, **17**, 2063–2073.