

RESEARCH

Open Access



Open chromatin analysis in *Trypanosoma cruzi* life forms highlights critical differences in genomic compartments and developmental regulation at tDNA loci

Alex Ranieri Jerônimo Lima^{1,2†}, Herbert Guimarães de Sousa Silva^{1,2,3†}, Saloe Poubel^{1,2}, Juliana Nunes Rosón^{1,2,3}, Loyze Paola Oliveira de Lima^{1,2}, Héllida Marina Costa-Silva^{1,2}, Camila Silva Gonçalves^{4,5}, Pedro A. F. Galante⁶, Fabiola Holetz⁷, Maria Cristina Machado Motta^{4,5}, Ariel M. Silber⁸, M. Carolina Elias^{1,2} and Julia Pinheiro Chagas da Cunha^{1,2*} 

Abstract

Background: Genomic organization and gene expression regulation in trypanosomes are remarkable because protein-coding genes are organized into codirectional gene clusters with unrelated functions. Moreover, there is no dedicated promoter for each gene, resulting in polycistronic gene transcription, with posttranscriptional control playing a major role. Nonetheless, these parasites harbor epigenetic modifications at critical regulatory genome features that dynamically change among parasite stages, which are not fully understood.

Results: Here, we investigated the impact of chromatin changes in a scenario commanded by posttranscriptional control exploring the parasite *Trypanosoma cruzi* and its differentiation program using FAIRE-seq approach supported by transmission electron microscopy. We identified differences in *T. cruzi* genome compartments, putative transcriptional start regions, and virulence factors. In addition, we also detected a developmental chromatin regulation at tRNA loci (tDNA), which could be linked to the intense chromatin remodeling and/or the translation regulatory mechanism required for parasite differentiation. We further integrated the open chromatin profile with public transcriptomic and MNase-seq datasets. Strikingly, a positive correlation was observed between active chromatin and steady-state transcription levels.

Conclusion: Taken together, our results indicate that chromatin changes reflect the unusual gene expression regulation of trypanosomes and the differences among parasite developmental stages, even in the context of a lack of canonical transcriptional control of protein-coding genes.

Keywords: *Trypanosoma cruzi*, FAIRE-seq, tRNA, Active chromatin, Gene expression control, Polycistronic transcription, Virulence factors, Epigenomics, Chromatin remodeling, Cell differentiation

Background

Chromatin is the stage where gene expression and epigenetic regulation occur. The epigenome of a cell consists of a series of heritable chromatin modifications that orchestrate and mirror essential aspects of gene regulation. For example, in general, the promoter regions of

[†]Alex Ranieri Jerônimo Lima and Herbert Guimarães de Sousa Silva contributed equally

*Correspondence: julia.cunha@butantan.gov.br

¹ Laboratório de Ciclo Celular, Instituto Butantan, São Paulo, SP, Brazil
Full list of author information is available at the end of the article



© The Author(s) 2022, corrected publication 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

highly expressed genes have an open chromatin state and are enriched in active histone marks such as H3K4me. In contrast, genomic regions with lower transcription rates are present in a more compact chromatin state and enriched in histone marks such H3K9me3 and H3K27me3 [1, 2]. Epigenomic changes can be explored using next-generation sequencing coupled with traditional approaches to detect DNA regulatory sequences. The FAIRE (formaldehyde-assisted isolation of regulatory elements)-seq technique is one of these approaches that uses formaldehyde to cross-link proteins, mainly histones, to DNA [3]. This results in an enrichment of genomic regions associated with regulatory activity, previously detected by DNase I and active chromatin marks, such as H3K4me and RNA Pol II [4]. Promoter regions of highly expressed genes are more enriched in FAIRE than poorly transcribed regions, indicating that FAIRE-enriched sequences are associated with active genomic regions [5].

Trypanosomatids are early-divergent eukaryotes that present some unique molecular characteristics, such as a hypermodified thymine base, known as base J, intronless genes, polycistronic transcription of functionally unrelated genes, mRNA maturation by trans-splicing, and posttranscriptional regulation of gene expression [6]. RNA Pol II transcription starts mainly at the intergenic regions between two polycistronic transcription units (PTUs) of divergent transcription (dSSR—divergent strand switch region) and ends between two convergent PTUs (cSSR—convergent strand switch region) or at regions transcribed by other polymerases [6]. Although no gene-specific promoters have been described, transcription starts at dispersed promoter regions rich in GT [7] and decorated by epigenetic marks and histone variants (H2A. Z and H2B. V) [8–10]. Transcription termination regions (TTRs) also contain epigenetic marks, such as histone variant H3. V and H4. V, which favor ending RNA Pol II transcription [9, 11]. In addition, dSSRs and cSSRs were shown to be enriched in base J [12].

Trypanosoma cruzi, the etiologic agent of Chagas disease, alternates between a vertebrate host and a triatomine vector [13]. In addition to the differences in cell shape, organization, and metabolism, *T. cruzi* life forms are also distinguished by the capacity of cell replication (epimastigotes and amastigotes) and cell infection (tissue culture trypomastigotes—herein called only trypomastigotes; and metacyclic trypomastigotes—herein called only metacyclic) [14–16]. Profound changes in global transcription are found during the differentiation of epimastigotes into metacyclics and between epimastigotes and trypomastigotes [17]. Evaluation of transcription rate dynamics during metacyclogenesis indicates that the decrease in transcription rates only occurs in

fully differentiated metacyclics [18]. Differentiation to metacyclics is followed by a global downregulation of the transcriptome except by a translational upregulation of virulence factors, such as TS and GP63, in metacyclics [19]. The phosphorylation status of the translation factors eIF2 α and eIF5a seems to be a key mechanism controlling translation and, therefore, *T. cruzi* metacyclogenesis [20, 21].

Nucleus and chromatin also change throughout the parasite life cycle. In epimastigotes and amastigotes, the nucleus is spherical, containing a large nucleolus and high amounts of euchromatin, which is considered transcriptionally active. In contrast, the trypomastigote forms (TCTs and metacyclics) present an elongated nucleus, absence of nucleolus, and abundant heterochromatin, which is dispersed in the nucleoplasm [15, 17, 18]. Changes in nuclear morphology start at the beginning of metacyclogenesis, while heterochromatin spreads progressively along with the nucleus in intermediate forms [15]. The mechanisms that drive and control these ultrastructural alterations during metacyclogenesis are not fully understood.

The genome of this parasite was proposed to be compartmentalized in two regions that differ in gene and GC content: the core compartment (composed of conserved and hypothetical conserved genes) and the disruptive compartment (containing the multigenic families trans-sialidase—TS, mucin-associated surface proteins—MASP, and mucins). GP63, dispersed gene family 1 (DGF-1), and retrotransposon hot spot (RHS) multigene families can occur in both compartments [22]. Multigene family members such as TS, mucins, MASP, and GP63 code for virulence factors present on the parasite surface and are upregulated in infective forms [23].

In recent years, our group has demonstrated that changes in epigenetic marks, such as histone posttranslational modifications (PTMs), histone variant deposition, and nucleosome occupancy, change among *T. cruzi* life forms and follow parasite differentiation [24–26]. As *T. cruzi* differentiation is followed by a global decrease in transcription, translation, and alterations in nuclear morphology in a background of transcriptional regulation absence of protein-coding genes, we comprehensively investigated the impact of chromatin changes in a scenario commanded by posttranscriptional control. By FAIRE-seq, we identified important differences in *T. cruzi* genome compartments and a developmental change at tRNA loci (tDNA). Of note, the latter may be associated with the regulatory mechanism associated with translation control required for parasite differentiation and/or the intense chromatin remodeling detected as differentiation advances. The integrative analysis of FAIRE-seq with public transcriptomic and nucleosome mapping datasets

allowed us to shed light on the mechanisms related to the differentiation of infective forms and help disentangle the impact of chromatin remodeling in a scenario commanded mainly by posttranscriptional control.

Results

Core and disruptive genomic compartments of *T. cruzi* have a different open chromatin profile

To investigate whether the lack of transcriptional regulation of protein-coding genes would be reflected in the open chromatin profile of *T. cruzi*, we performed a FAIRE-seq genome-wide approach. We rescued these regions from the genome of epimastigotes under exponential proliferation. Biological duplicates were processed together with input control samples (reversed cross-link before DNA extraction) to account for any bias in genome sequencing and assembly. FAIRE-seq reads presented high duplication levels (66.2–72.1% in epimastigotes—Additional file 1: Figure S1A) and high Phred score (quality) values (>30—Additional file 1: Figure S1B). After the mapping process (Additional file 1: Figure S1C), only mapped reads with mapping quality scores (MAPQs) above 10 were retained, resulting in approximately 116 million reads mapped (Additional file 1: Figure S1D). Spearman correlation analysis (Additional file 1: Figure S1E) and PCA (Additional file 1: Figure S1F) using genome-wide read coverage of 50-bp windows for each replicate indicated a high correlation among them, showing good reproducibility and agreement among biological replicates. To avoid bias at repetitive regions, we removed multimapped reads, resulting in reduced (or zero) read coverage at repetitive regions of the genome (Additional file 1: Figure S2). Most of the repetitive regions occurred in intergenic sections (92.75%) of the genome (Additional file 2: Table S1). Of note, strand switch regions (SSRs) were not considered intergenic regions; nevertheless, no repeat annotations were found there. Only 309 CDSs (5.65% of total repetitive regions, in bp) were located in annotated repetitive regions.

To obtain an overview of the open chromatin distribution in large genomic regions, we compared the data of epimastigote FAIRE-seq (E) and their corresponding control samples (C). Figure 1 shows the distribution of reads obtained from both samples on two contigs from Dm28c (PRFA1000011—594 kb and PRFA1000027—454 kb). Datasets were normalized to reads per genomic content (RPGC) to account for different sequencing depths. Visual inspection of RPGC levels indicates that the FAIRE sample has higher overall levels along the contigs with some clear enrichments when compared to the control sample. Ratio levels in log₂ (E/C) indicated that most regions were enriched in FAIRE samples (Fig. 1A). Repetitive regions are depleted in both FAIRE and control

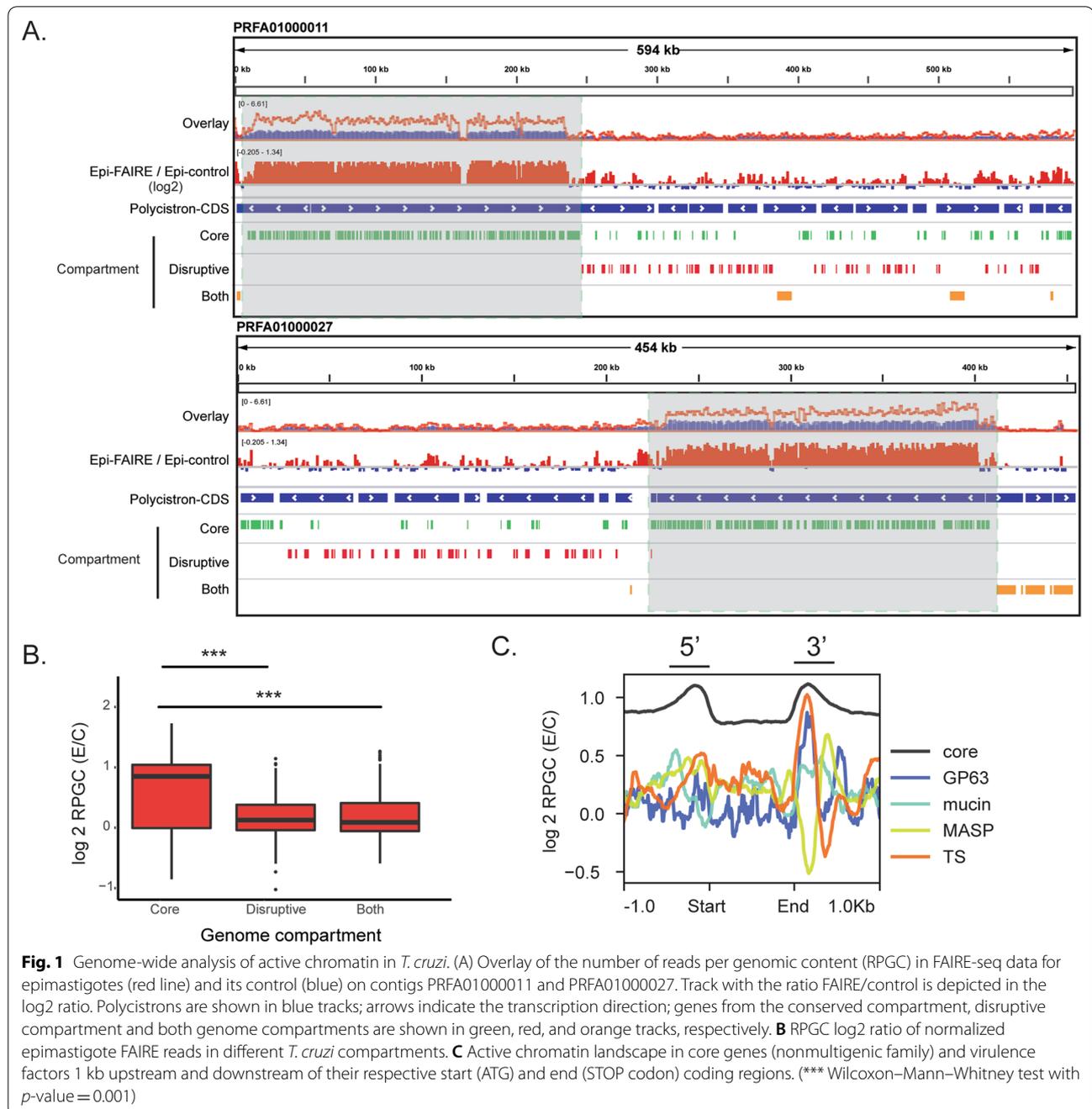
samples, likely due to the applied mapping and filtering steps (Additional file 1: Figure S2).

Regions covered by the core compartment (composed of conserved and hypothetical conserved genes, green tracks) had higher RPGC-normalized counts than the disruptive compartment regions (red tracks) (Fig. 1A and B, Additional file 2: Tables S2, S3). Differences in RPGC counts between these two compartments were also observed in control samples; however, this difference was even more evident in FAIRE samples, which was reflected by higher levels of log₂ ratio (E/C) in the core than in the disruptive compartment (Fig. 1B). Remarkably, removing multimapped reads (see [methods](#)—Additional file 1: Figure S2B) slightly affected this difference since the same criteria were used for experimental and control samples. Instead, this can be explained by the less open chromatin at the multigenic family in contrast to a more open chromatin profile in the nonmultifamily CDS.

Because the distribution of open chromatin was revealed to be distinct between genome compartments, we asked whether their landscapes, at a gene level, were also different. Strikingly, the landscape of open chromatin greatly differs among virulence factors (present in the disruptive compartment), among them, and among the remaining protein-coding genes. Figure 1C shows that at most core-gene members, the regions coding for their 5' and 3' UTRs are enriched in open chromatin, corroborating the impoverishment of nucleosomes seen previously by MNase-seq (Additional file 1: Figure S3B). In contrast, no clear enrichment of open chromatin was found around the genomic regions coding for the 5' UTR of virulence factors, while the regions coding for the 3' UTR greatly differed among them: TS and GP63 exhibited an enrichment similar to other CDSs. At the same time, MASPs had a clear depletion followed by an enrichment a few base pairs downstream (Fig. 1C). Nucleosome occupancy partially explains the active chromatin distribution around the virulence factor genes, as some regions (such as the upstream regions of mucins, GP63, DGF-1 and RHS) exhibit enrichment or decrease both in FAIRE and MNase data (Additional file 1: Figure S3A).

Open chromatin is enriched at divergent SSRs (dSSRs) and uniformly distributed along PTUs

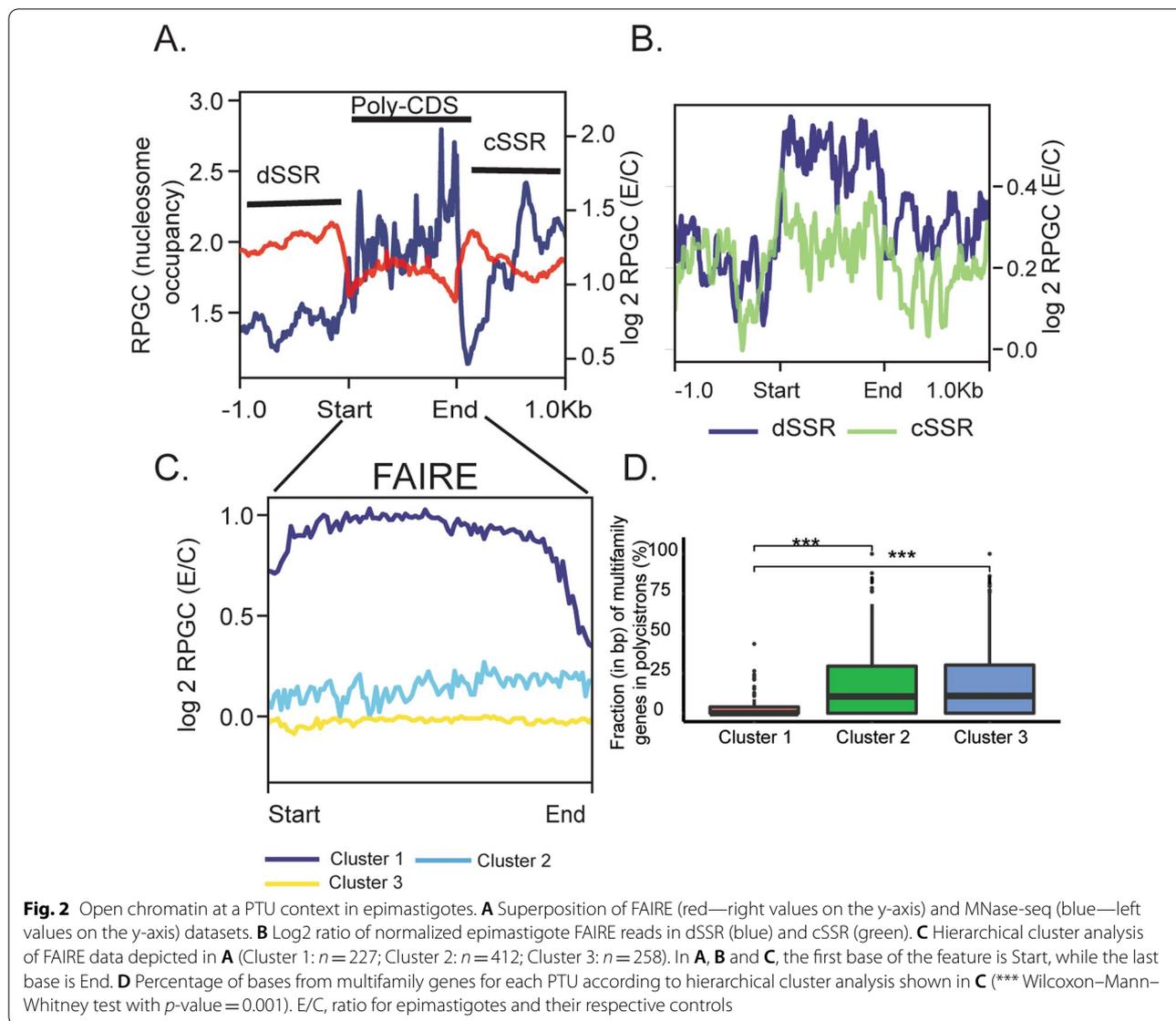
In trypanosomatids, transcription of protein-coding genes initiates mainly at dSSRs and terminates at cSSRs; however, transcription might also start at some non-SSRs, mainly close to tRNA genes [6, 27, 28]. This latter has not yet been described in *T. cruzi*, we inspected dSSRs as a proxy of transcription start sites, which resulted in an enrichment of open chromatin compared to PTUs, as evidenced in Fig. 2A and 2B. Combining a previous nucleosome mapping and occupancy profile



[25] with the FAIRE-seq data along PTUs revealed complementary opposite landscape profiles: open chromatin regions mainly reflect nucleosome-depleted regions (Fig. 2A).

PTU regions were hierarchically clustered into three groups based on their log₂ (E/C) RPGC level (Fig. 2C). Clusters 2 and 3 are very similar, exhibiting a near-flat

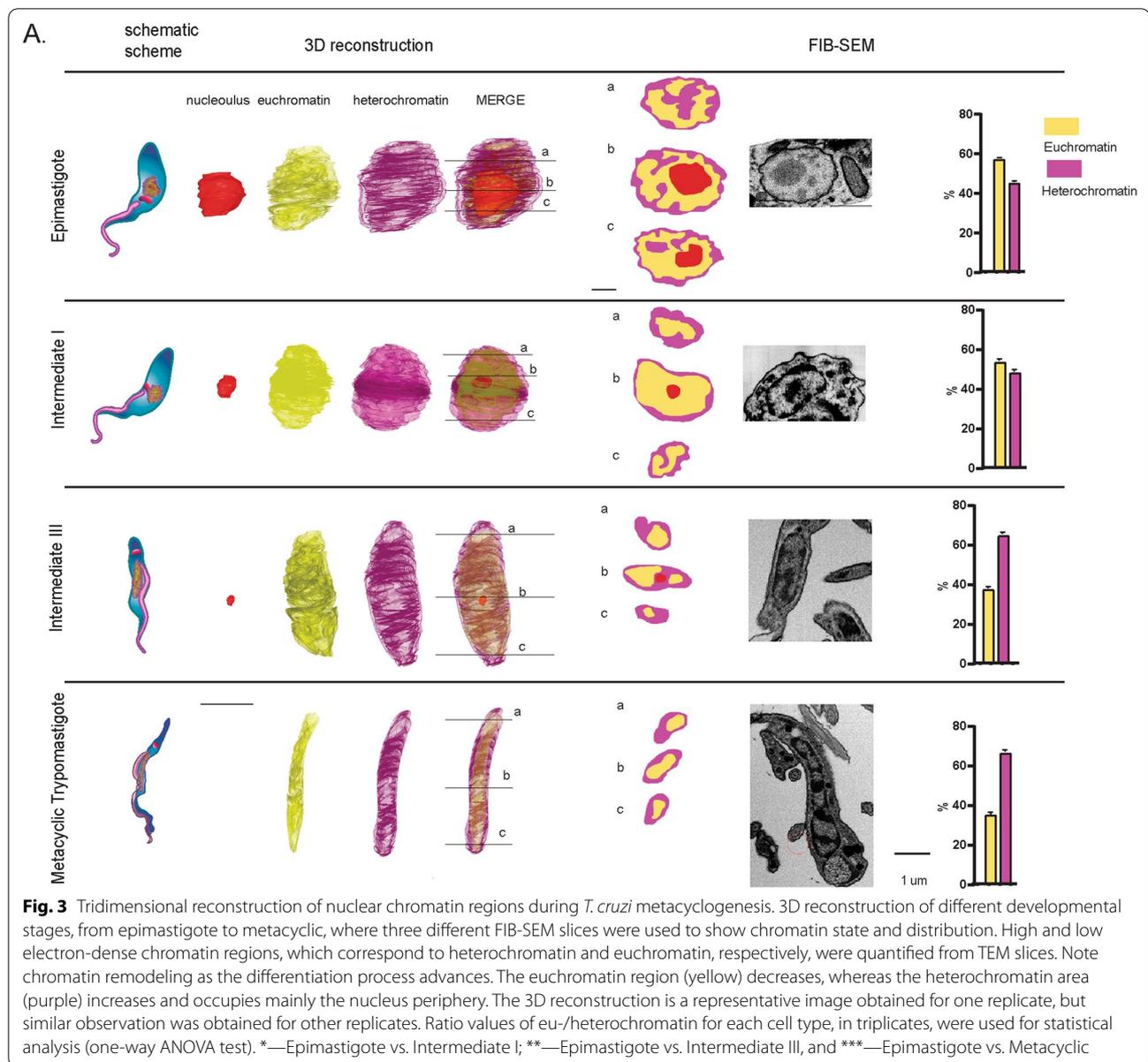
pattern with low overall RPGC levels. In contrast, Cluster 1 showed higher overall levels of open chromatin, with a decrease at the edges, mainly at the PTU ending region. Clusters 2 and 3 contained significantly more genes (per bp) from the multigenic family than Cluster 1, which was enriched mainly from genes of the core compartment (Fig. 2D).



Levels and relative nuclear position of eu- and heterochromatin changes during metacyclogenesis

Many morphological changes are observed during the differentiation of epimastigotes to metacyclics, including nuclear elongation and kinetoplast repositioning to the parasite posterior end. Previously, it has been reported that the heterochromatin near the nuclear envelope in the epimastigote forms spreads progressively along with the nucleus in the intermediate forms, reaching a higher level of compaction in metacyclics [15]. However, only in this work a systematic evaluation and 3D reconstruction of the two traditional chromatin classes, eu- and heterochromatin, was performed to check chromatin remodeling during parasite differentiation. The obtained results indicate that in *T. cruzi*

epimastigotes, euchromatin resides in the central area, whereas heterochromatin is mainly distributed close to the nuclear envelope and surrounding the nucleolus (Fig. 3A—epimastigote). The euchromatin volume is higher in epimastigote and intermediate I forms and decreases during differentiation to metacyclic. Heterochromatin, in the epimastigote form, spreads throughout the nucleus, and as metacyclogenesis advances, its percentage increases and its location becomes increasingly peripheral. These results indicate that progressive chromatin remodeling occurs during parasite differentiation. It is worth observing that during metacyclogenesis, the nucleolus reduces its size, as it undergoes disassembly and dispersion throughout the nuclear matrix, which may be related to the decrease in



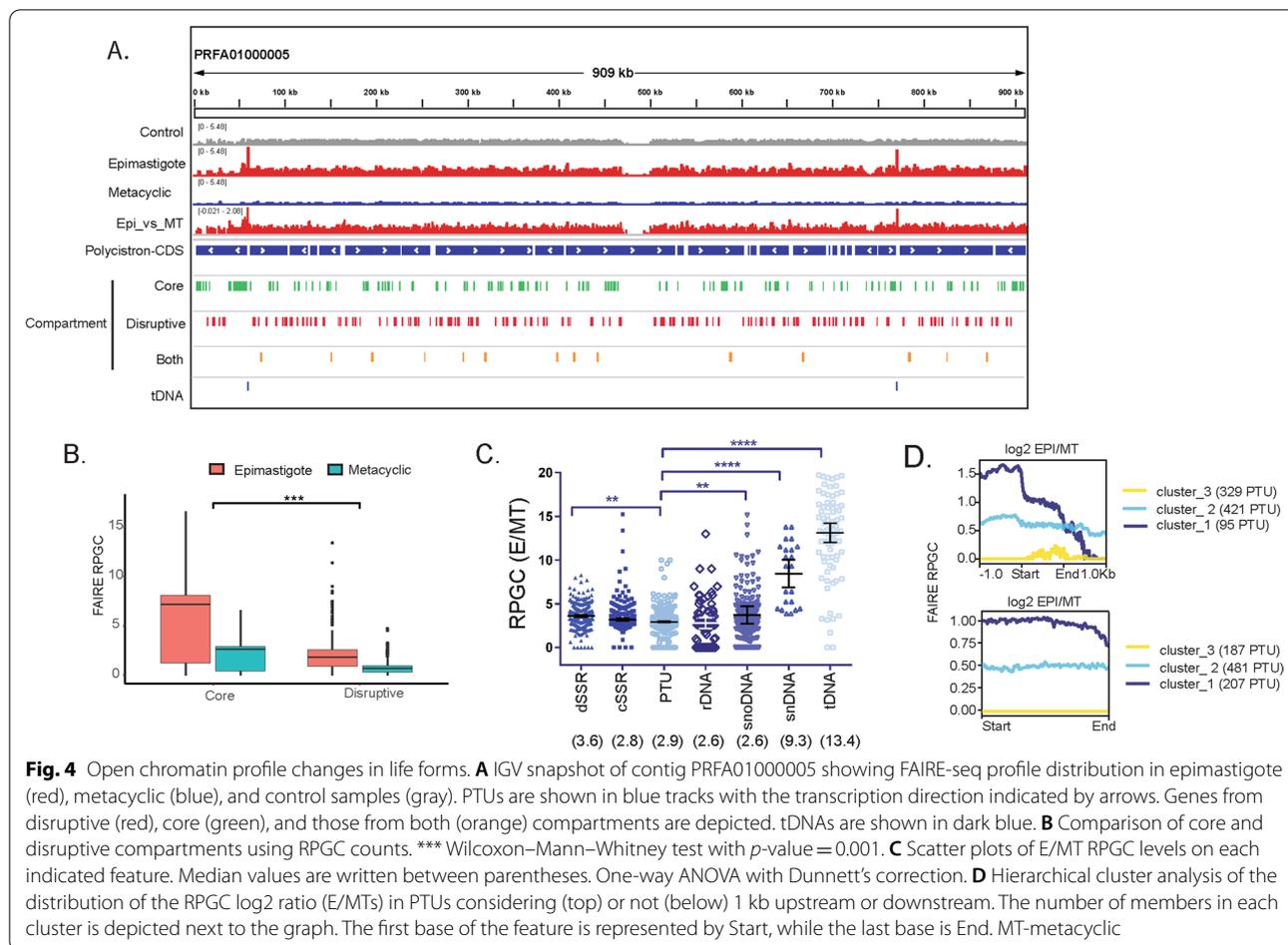
ribosomal biogenesis (Fig. 3, Additional file 3: Videos S1–S4).

Genome-wide analysis of open chromatin regions in epimastigotes and metacyclics

Tridimensional reconstruction of chromatin areas along epimastigote-to-metacyclic differentiation indicated a significant reduction in euchromatin regions (Fig. 3), which are considered open chromatin areas. To gain more insights into these changes, we performed a comparison of FAIRE-seq data from epimastigote and metacyclic forms. Visual inspection of normalized RPGC levels along the *T. cruzi* genome confirms a greater

abundance in open chromatin and few clear enrichments in epimastigote forms than metacyclics (Fig. 4A).

As mentioned above, the disruptive compartment is enriched in virulence factors that are mainly expressed in infective forms [23]. To address whether the disruptive compartment would be enriched in open chromatin in metacyclics relative to epimastigotes, RPGC-normalized FAIRE-seq data were compared between life forms, obtaining log₂ ratios. Median RPGC values from the disruptive and core compartments were lower in metacyclics (Fig. 4B, Additional file 2: Tables S2, S3). Differences between compartments within life forms are very similar: the core compartment has 3.9 times more open



chromatin than the disruptive compartment in both epimastigotes and metacyclics. In the MNase data, core and disruptive compartments also presented different RPGC levels, with the former being significantly higher (Additional file 1: Figure S3B). In general, virulence factors (TS, GP63, MASP, and mucins) and core compartments have 2.6 and 2.8 times (fold change, median values) less open chromatin in metacyclics than in epimastigotes (Additional file 1: Figure S3A), which agrees with global changes in other genomic features (see Fig. 4C, Additional file 2: Table S4). The open chromatin landscape of virulence factor genes is similar in both life forms, with a slight difference at the upstream coding region of mucins (Additional file 1: Figure S3D). Taken together, these results indicate that virulence factors indeed have a different pattern of open chromatin when compared to other CDSs (as we discussed below) that seemed to be maintained between life forms, which reinforces major posttranscriptional control of these genes.

Previously, we found that dSSRs from trypomastigotes were enriched in nucleosomes compared to

epimastigotes [25]. Then, we speculated whether dSSRs from epimastigotes would be more abundant in open chromatin. Indeed, dSSRs are approximately 3.6 more enriched in open chromatin at epimastigotes (Fig. 4C) (Additional file 1: Figure S4A), which corroborates a lower nucleosome occupancy in replicative forms compared to nonreplicative forms.

Upon hierarchical cluster analysis, the enrichment of open chromatin at dSSRs can be further visualized in a PTU context (Fig. 4D—top, Additional file 2: Table S5). Most notably, Cluster 1 showed enrichment at dSSRs followed by a decreasing signal level toward the cSSRs. A detailed inspection of Cluster 1 elements revealed that approximately 10% of their PTUs are located downstream of tDNA loci. Cluster 2 represents PTUs whose dSSRs have similar levels of openness compared to their adjacent PTUs. Furthermore, interestingly, Cluster 3 encompasses PTUs whose levels of open chromatin are equal between E and metacyclics (\log_2 E/metacyclic = 0). Taken together, the obtained data indicate that different PTUs have distinct levels of openness at their

transcription initiation regions. Clusters 2 and 3 were enriched in genes from the dispersed compartment (data not shown).

Considering only the PTU region, epimastigotes showed a significantly increased signal (2.9 times) compared to metacyclics (Fig. 4C). The hierarchical cluster analysis reflected PTUs with different levels of open chromatin among life forms and, importantly, near-flat enrichment along the PTU region (Fig. 4D bottom, Additional file 2: Table S6). In accordance with hierarchical clustering of PTUs with SSRs (Fig. 4D top), Clusters 2 and 3 were also enriched in the multigenic family (data not shown).

FAIRE-seq data correlate with steady-state gene expression levels in both life forms

Given the different open chromatin profiles observed among PTUs (Fig. 4D), we investigated whether a similar pattern occurs in distinct CDSs from different life stages. Thus, CDSs were hierarchically clustered into three groups based on their log₂ RPGC level (Additional file 1: Figure S4D). An increase in open chromatin was found mainly at genomic regions encompassing 5' and 3' UTRs at Clusters 1 and 2. Similar to the results shown above, Cluster 1, which has the highest log₂ ratio, was enriched in genes from the core compartment, whereas Clusters 2 and 3 were enriched in genes from the disruptive compartment and repetitive genes (data not shown).

Then, we explored whether regions enriched in open chromatin would have transcripts expressed at higher levels. CDSs were first classified as high, medium or low expressed in each life form based on their TPM counts obtained from a transcriptomic study published elsewhere [19] (Additional file 1: Figure S5). RPGC-normalized counts were retrieved for each CDS according to their respective expression class and each life form (Fig. 5). Surprisingly, a positive correlation of open chromatin with steady-state transcription levels was found. For epimastigotes, significant differences in FAIRE enrichment were observed when comparing all expression classes. In contrast, for metacyclics, significance was observed mainly when compared to the low expressed genes (high versus low and medium versus low).

Open chromatin is developmentally regulated at tDNA loci

FAIRE-seq analysis highlighted a global decrease in the levels of open chromatin in metacyclics compared to epimastigotes. However, we wondered whether the open chromatin profile would give us more clues about the differences in gene expression and phenotype found between life forms, especially those related to their differentiation program. Comparing all genomic features, we detected striking differences in open chromatin

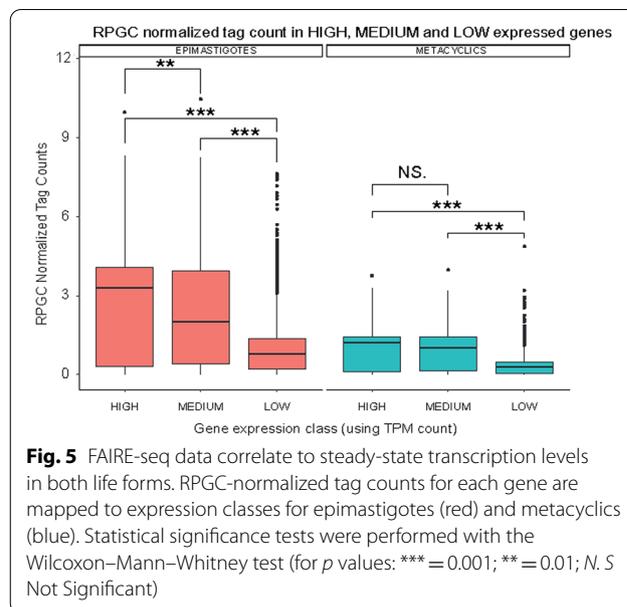


Fig. 5 FAIRE-seq data correlate to steady-state transcription levels in both life forms. RPGC-normalized tag counts for each gene are mapped to expression classes for epimastigotes (red) and metacyclics (blue). Statistical significance tests were performed with the Wilcoxon–Mann–Whitney test (for *p* values: *** = 0.001; ** = 0.01; N.S. Not Significant)

enrichment at the regions coding for the small nuclear RNAs (snRNAs) and tRNA genes. Of note, this enrichment was significantly higher (fold changes of 13.4 and 9.3 for tDNAs and snDNA, respectively) in epimastigotes than in metacyclics (Figs. 4C and 6A, B).

Hierarchical cluster analysis of the E/metacyclic ratio revealed that the majority of tDNAs had at least six times more open chromatin in epimastigotes than in metacyclics (Fig. 6C, Additional file 1: Figure S6, Additional file 2: Table S7). Clusters were not related to the tRNA isoacceptor type or class (Additional file 1: Figure S6B); however, their distribution reflects their location regarding the transcription direction of the adjacent PTU (Additional file 1: Figure S6C). For example, Clusters 1 and 2 were enriched in tDNAs that were mainly located between codirectional (60% and 30%, respectively) and divergent (10% and 40%, respectively) PTUs (Fig. 6D). This distribution suggests that chromatin alterations at tDNAs may affect transcription initiation at adjacent PTUs.

Finally, we addressed whether the increase in open chromatin within tDNAs would be associated with tRNA expression levels. Previously, Garcia-Silva [29] showed by RNA-FISH assays that compared to metacyclics, epimastigotes have higher amounts of tRNAs. To confirm this finding, epimastigotes and metacyclics were probed simultaneously using one of the four probes against 5' and 3' Asp and Glu tRNAs by RNA-FISH assays. Additional file 1: Figure S6D indicates that tRNAs are much lower in metacyclics (about 50% less abundant). Evaluation of tRNA expression by PCR is not a straightforward approach once this molecule is highly modified. Then,

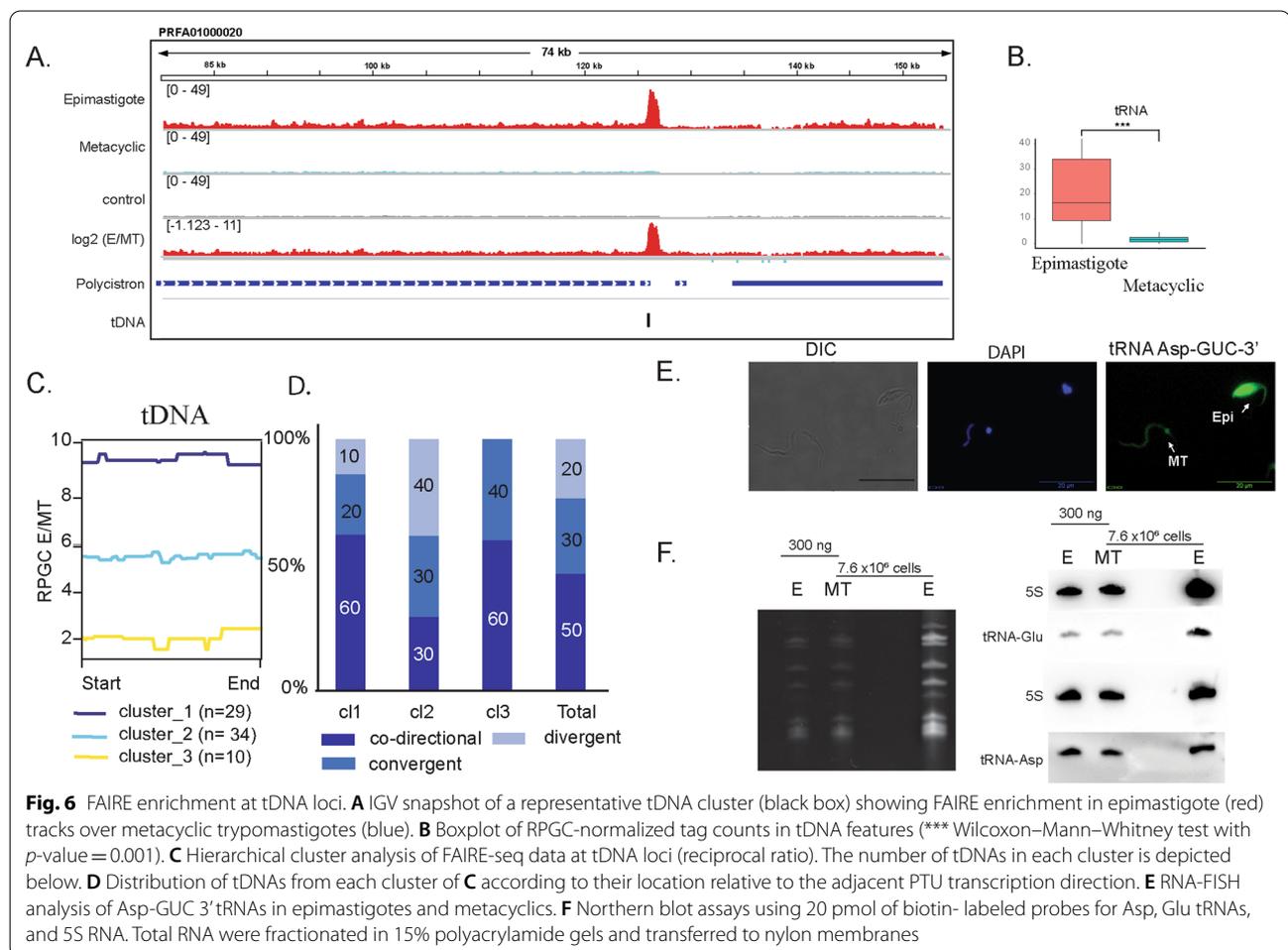
we also evaluated tRNA expression in life forms using Northern blot. Asp and Glu tRNAs are more abundant (per cell) in epimastigotes than at metacyclics (Fig. 6E–F, Additional file 1: Figure S6F) corroborating with RNA-FISH analysis; but tRNAs are equally abundant when similar amounts of total RNA (in ng) from life forms where compared (Fig. 6F). It is important to note that in both analyses we are evaluating steady-state tRNA transcripts, instead of nascent transcripts. The nascent transcript analysis is hampered due to the very low abundance of transcription in metacyclics [17].

Discussion

The differentiation of *T. cruzi* into nonreplicative life forms is followed by a global decrease in RNA transcription, which is mainly observed in fully differentiated forms, such as trypomastigotes and metacyclics [17, 18]. Chromatin is the stage where gene expression regulation occurs, and trypanosomes offer an exciting scenario to evaluate its changes once their protein-coding genes

are devoid of transcription regulation in a gene-specific manner [6]. Here, we showed that the differentiation of epimastigotes to metacyclics is followed by a progressive increase in the percentage of heterochromatin area, which becomes restricted to the nuclear periphery (Fig. 3). Using a conformation chromosome capture approach, euchromatic and heterochromatic regions were further associated with structural genome compartments A and B, which are related to transcriptionally active and inactive regions, respectively [30, 31]. Our work indicates that the differentiation program that drives morphological (such as repositioning of kinetoplast and flagellum, nuclear elongation) and metabolic changes in trypanosomes, is also followed by intense remodeling of chromatin domains.

To further explore the genomic context of open chromatin, we used the FAIRE approach, a methodology previously known to enrich for active genomic regions [3, 5, 32]. Here we not only confirm, but also highlight many new findings compared to our previous MNase-seq



analysis [25]. Nevertheless, the agreement between both approaches not only reinforces the results, but also suggests a molecular explanation (nucleosome enrich or depletion) for a given “closed” or “open” chromatin region. FAIRE-seq analysis indicated that dSSRs are enriched in open chromatin relative to PTUs, corroborating the presence of fewer nucleosomes at dSSRs than PTUs [25]. Interestingly, the *T. cruzi* genome has PTUs whose dSSRs have different levels of open chromatin. Whether it is associated with differential deposition of the RNA Pol II machinery, histone variants, and histone PTMs and, more importantly, whether it has any consequences on the PTU transcription levels needs to be further investigated. Global nascent transcriptomic analysis comparing transcription along and within PTUs is absent in trypanosomes but can clarify this issue. In addition, hierarchical cluster analysis of PTUs without SSRs indicates different levels of open chromatin between PTUs, but with a clear flat pattern along the entire PTU. The latter indicates a uniform open chromatin status along PTUs, possibly reflecting the polycistronic transcription nature of trypanosomes.

We envisaged that the presence of two genomic compartments in *T. cruzi*, proposed by Berná et al. [22], may induce the formation of a complex 3D chromatin structure. Here, by evaluating open chromatin status in a genome-wide fashion, the core and disruptive genomic compartments of *T. cruzi* exhibited great differences in the enrichment of open chromatin, with the former highly enriched in open chromatin compared to the latter. The disruptive compartment is composed of virulence factors that have a very repetitive nature. Therefore, we cannot rule out that mapping and genome assembly bias influence these data, although removing multimapping reads did not affect the results. Regardless, because virulence factors are mainly expressed in infective forms, it was expected that these chromatin regions would be more open in metacyclics if a gene-specific transcriptional regulation scenario existed. However, it was determined that both the disruptive and core compartments were lower in metacyclics. These findings confirm that posttranscriptional regulation of virulence factors indeed plays a key role in their expression, as previously indicated [33, 34].

Moreover, the multigenic families MASP, GP63, mucin, and TS exhibited distinct open chromatin landscapes compared to the other nonmultifamily members. The reason for this may be related to the greater GC content of the disruptive compartment than the core compartment [22], because the G plus C content is intrinsically correlated with nucleosome occupancy [35]. Although an expected opposite correlation among FAIRE and MNase-seq data was observed in the majority of evaluated

genomic regions, clearly it was possible to note that some regions do not contain nucleosomes, but are depleted in active chromatin and vice versa, indicating the presence/absence of other chromatin players. It is important to stress that different life forms were used in each study, which could account for the observed divergence. In MNase-seq approach, we used trypanomastigotes and here, metacyclics. Nevertheless, the fact that some aspects of MNase and FAIRE data from TCTs and metacyclics corroborate indicates that the epigenome landscape of these life forms is well conserved even though they are different cell types. Another important variable is the fact that kinetoplasts contain a thymine modification called base J that is mainly located in telomeric regions. In *T. cruzi*, base J was also found flanking PTUs and within subtelomeric regions, which are rich in virulent factors (located at the disrupted compartment). Interestingly, the presence of base J at these genes was shown to be developmentally regulated [36, 37]. Thus, it remains to be investigated whether the presence of base J at the disruptive compartment may interfere with chromatin openness and/or if other protein complexes may be in these regions, thereby affecting chromatin architecture.

Strikingly, the open chromatin state based on FAIRE-seq read enrichment is associated with steady-state transcription levels in *T. cruzi*. These findings are remarkable once trypanosome gene expression relies mainly on posttranscriptional events. However, this is not unprecedented. Previously, it was determined that the nucleosome landscape reflects steady-state gene expression mainly regarding the depth of the nucleosome-depleted region around 5'UTR [7, 25, 38]. A proposed explanation is the association of cotranscriptional splicing resulting in putative interference with the final levels of transcripts. It can be also hypothesized that the machinery associated with RNA processing (transplicing and polyadenylation) and some RNA binding proteins would closely associate with chromatin, thereby influencing local chromatin structure. In fact, *T. cruzi* chromatin proteomics reveals the presence of at least 13 mRNA binding proteins [24], but further experiments should be performed to determine whether these proteins indeed interact with chromatin and if they may play a role in chromatin architecture.

Of note, the open chromatin landscape seems to be conserved between life forms, except for a global decrease at their levels on metacyclics. This agrees with a global decline in transcription levels along with parasite differentiation [18]. One of the main differences in FAIRE enrichment between life forms was found at loci that code for snRNAs and tRNAs. snRNAs are uridine-rich sequences that compose the spliceosome and are transcribed by RNA polymerase III [39, 40]. tDNAs are

approximately 78 bp long and contain intrinsic promoters, known as Box A and Box B, which are the binding sites for the transcription factor complex named TFIIC and are responsible for recruiting the TFIIB complex to enable the assembly of RNA polymerase III at the TSS region of these genes [41, 42]. Due to the higher expression of tDNAs on epimastigotes per cell, we believe that the RNA Pol III complex would be enriched in tDNA loci, likely hampering the deposition of nucleosomes and generating an open chromatin status. In fact, epimastigotes indeed have a lower enrichment of nucleosomes at tDNAs (compared to trypomastigotes). Nevertheless, in trypomastigote life forms, tDNA loci are depleted of nucleosomes compared to their surrounding regions [25].

Of note, the absence of transcription machinery at tDNA loci can be the cause or consequence of the closed chromatin status. However, the implications of the developmental regulation of this status can be hypothesized. The mentioned changes may influence the expression of the nearby PTUs, which, in turn, would be preferentially affected along with differentiation. In this sense, it is interesting that tDNA hierarchical cluster analysis pointed out a distinct distribution of tDNA openness based on their location related to adjacent PTUs (convergent, divergent, or codirectional). We hypothesize that the loss of open chromatin status at tDNAs located within a codirectional PTU may affect transcription mainly at the downstream PTU. This is supported by the findings that indicate that tDNA loci and their associated transcription factors, mainly TFIIC, may function as insulators. The latter has a critical role in chromatin domain organization, working on separating active and silenced domains [43].

Developmentally regulated chromatin changes at tDNA loci are especially intriguing considering that translation control has a key role in trypanosomatid differentiation. During the differentiation of epimastigotes to metacyclics, a global downregulation of a large proportion of the transcriptome occurs [19, 21, 44]. Although the underlying mechanism is not fully understood, phosphorylation of eIF2 α increases upon nutritional stress attenuates protein synthesis during *T. cruzi* differentiation [21]. Here, we confirmed that, per cell, tRNAs are less expressed in metacyclics than in epimastigotes, correlating with the diminished amount of open chromatin in the former than in the latter. Thus, chromatin changes at tDNA loci may be another layer of translation regulation during development. In fact, many studies indicate that tRNA levels are indeed a limiting step in translation [45], blocking it when presented below a given threshold. However, tRNAs have similar abundance in epimastigotes and metacyclics when equal amounts of total RNAs (in nanograms) were compared. This confirms a global

downregulation of transcription in metacyclics [18], indicating a similar behavior of tRNAs abundance over other RNA types (5S and 18S—data not shown). However, we could not affirm a direct association of open chromatin profile and tRNA transcription because no nascent transcription expression analysis was performed, but we hypothesize that the critical differences in open chromatin at tRNA loci could be associated with an intense chromatin remodeling as metacyclogenesis advances (Fig. 3). As mentioned above, tDNAs have a critical role in chromatin organization working as insulators and enriched at topological-associated domains boundaries [43]. In yeast stationary phase, members of tRNA expression machinery (mainly Pol III and TFIIB) are released from chromatin, allowing nucleosome deposition [46]. We envisage a similar scenario as metacyclogenesis advances: tRNA machinery associated with tDNAs in epimastigotes would be released, allowing binding of chromatin architectural proteins and nucleosomes (justifying the depletion of open chromatin) in metacyclics leading to a possible 3D genome structure shaping.

Conclusion

Overall, the open chromatin genome-wide approach recapitulates unique aspects of trypanosomatid gene regulation such as polycistronic transcription and lack of RNA pol II transcription regulation. However, it also indicated that different dSSRs may have different levels of open chromatin, suggesting PTU-specific regulation. In addition, the status of open chromatin correlates with the presence of conserved and disruptive genome compartments, which further correlates with the expression of virulence factors. Finally, the obtained data indicated developmental chromatin regulation at tDNA loci, which may orchestrate an essential step in posttranscriptional regulation control necessary for parasite differentiation and survival.

Methods

Parasites

T. cruzi (Dm28c strain) epimastigotes were cultured at 28 °C in LIT medium supplemented with 10% fetal bovine serum (FBS—Vitrocell), glucose 0.4%, hemin 0.1 μ M, and penicillin-G 59 mg/L as previously described [47]. Metacyclic trypomastigote (MT) parasites were obtained by the metacyclogenesis protocol described at [48]. Briefly, epimastigotes in the exponential growth phase were diluted at 1×10^6 parasites/mL and cultivated for 5 days to reach the stationary phase. Then, parasites were resuspended at 5×10^8 parasites/mL in triatomine artificial urine medium (TAU, 190 mM NaCl, 17 mM KCl, 2 mM CaCl $_2$, 2 mM MgCl $_2$, 8 mM phosphate buffer pH 6.0) for 2 h at 28 °C. Next, parasites were diluted to 5×10^6 per

mL in TAU medium supplemented with 10 mM L-proline, 50 mM L-glutamate, 2 mM L-aspartate, and 10 mM glucose and maintained for 96 h at 28 °C. metacyclics were obtained from the supernatant for 72 h and purified by DEAE columns.

Nuclear tridimensional reconstruction

Image stacks previously obtained by focused ion beam scanning electron microscopy (FIB-SEM) [15] were automatically aligned using the StackReg plugin in the Fiji/ImageJ software package (<http://fiji.sc/wiki/index.php/Fiji>). Aligned images were saved as image sequence files and then imported into the 3D reconstruction program Free-D 1.15 [49]. Heterochromatin, euchromatin, and nucleolus were accomplished by manual classification and tracing the boundary contours of ultrastructures of interest. Visualization of the 3D shape of the cell nucleolus (red), heterochromatin (purple), and euchromatin (yellow) was achieved by representing the traces of these objects using Free-D 1.15. Nuclear 3D reconstructions were performed in triplicate using 70 to 130 slices for each cell.

FAIRE sample preparation

The FAIRE samples were prepared as described in [3] with some modifications. A total of 2×10^7 parasites (in biological duplicates) were fixed for 5 min with formaldehyde and directly added to the medium to a final concentration of 1%. Formaldehyde was quenched with 125 mM glycine, followed by extensive washing with PBS. Pellets were resuspended in 1 mL of cold lysis buffer TELT (50 mM Tris-HCl pH 8.0, 62.5 mM EDTA, 2.5 M LiCl, 4% Triton X-100) and sonicated with six cycles of 30 s with an interval of 1 min each (Output 4 and Duty 30 parameters) in a Tommy Ultrasonic Disruptor UD-201 apparatus. To prepare the input DNA control, 100 μ L of each sample was centrifuged at 15,000 g for 5 min at 4 °C. The supernatants were incubated with 1 μ L of RNaseA DNase-free for 30 min at 37 °C followed by treatment with 1 μ L of proteinase K for 1 h at 55 °C. The cross-links were reversed by overnight incubation at 65 °C. DNA was extracted with phenol:chloroform:isoamyl alcohol solution and precipitated with 3 M sodium acetate (pH 5.2) and 95% ethanol. To prepare FAIRE DNA, the remaining cell lysate was centrifuged at 15,000 for 5 min at 4 °C, and DNA was extracted as described above; however, RNaseA treatment, protein removal by proteinase K and cross-linking reversal were performed after DNA purification. The obtained DNA was quantified using a NanoDrop spectrophotometer (Thermo Fisher). A total of 500 ng of input DNA control was run on a 1% agarose gel to verify the sonication efficiency. DNA fragments were deep

sequenced using an Illumina NextSeq 500 (2×75 bp) at the Polyomics Facility (Glasgow University).

Bioinformatic analyses

Sequence read quality check, mapping, and normalization

The methods applied to analyze the FAIRE-seq data are shown in Additional file 1: Figure S1 C and described below. Raw sequencing files were quality checked with FASTQC [50]. A quality filter was not necessary (see Additional file 1: Figure S1 B). Reads were then mapped against the *Trypanosoma cruzi* Dm28c genome (downloaded from <https://tritrypdb.org/>—DB46, Dm28c 2018 version) using the GATK pipeline [51], which employs BWA [52] as the alignment algorithm. The GATK pipeline was executed until reaching the read deduplication step. To avoid ambiguous read mapping, the alignment files were filtered using a MAPQ (MAPing Quality) score of 10 using SAMtools [53]. MAPQ filtered alignment files were processed with multiBamSummary (binSize = 50) from deepTools 3.3.0 [54] to generate the scaling factors for normalization and the files necessary to perform PCA and Spearman correlation analysis (using plotPCA and plotCorrelation functions, respectively). The data were normalized using the RPGC (Reads Per Genome Coverage) method, present in bamCoverage function, applying the following scaling factors for each dataset: EpiR1: 2.3087, EpiR2: 4.0037, MetaR1: 0.5875, MetaR2: 0.6076, ControlR1: 0.6763, ControlR2: 0.6551. Replicates were merged into one bigwig file using the bigWigCompare function, calculating the mean signal for each replicate. The comparisons Epi vs. Control and Epi vs. Meta were performed with the bigWigCompare function using the parameters skipZeroOverZero, pseudocount 1 1, and operation log2. The datasets generated and/or analyzed during the current study are available in the NCBI BioProject repository under the accession code PRJNA763084 and BioSample codes SAMN21423261, SAMN21423262, SAMN21423263, SAMN21423264, SAMN21423265 and SAMN21423266.

Extragenomic feature annotation

Polycistronic regions (same strand CDS cluster region), dSSR and cSSR were annotated using the GFF file from *T. cruzi* Dm28c (downloaded from <https://tritrypdb.org/>—DB46, Dm28c 2018 version) and a custom Python script (available at: <https://github.com/alexranieri/annotatePolycistron>). Briefly, the script parses the GFF file, and a polycistron is considered for annotation if two or more CDSs of the same strand direction are found consecutively. The ID of each feature is incorporated in the polycistron annotation content, along with the number of features found. Polycistrons with only one feature were

also annotated. In a second analysis, the script processes the SSRs, and dSSR is considered when a region between the start of a negative-strand polycistron followed by the start of a positive-strand polycistron is found. cSSR is considered when a region between the end of a positive-strand polycistron followed by the end of a negative-strand polycistron is found. If none of those conditions are met, i.e., a positive-strand CDS polycistron is followed by another type of feature. The genomic region between them is annotated as an Intergenic Polycistron Region.

5' and 3' UTR annotations were obtained with UTRme [55] using the spliced-leader sequence (AACTAACGC TATTATTGATACAGTTTCTGTACTATATTG) from [56] and transcriptomic data from Smircich et al., 2015. tRNAscan-SE [57], with default parameters for eukaryotes, was used to identify anticodons and amino acids from tRNA sequences obtained from hierarchical cluster analysis.

Transcriptomic analysis

RNA-seq reads were preprocessed according to Smircich et al., 2015, with an additional step of sequencing adapter removal using cutadapt [58]. Then, the processed reads were mapped against the *T. cruzi* Dm28c (V46) genome using STAR [59]. For each CDS in the genome, transcripts per million (TPM) values were obtained from transcriptomic data using TPMCalculator [60]. RPGC-normalized tag counts (in log₂) were retrieved using annotatePeaks function from HOMER [61], passing the wig files, CDS coordinates in bed file and the parameters: -size 'given', -raw or -noadj (to avoid new normalization). Genes were classified into low, medium, or high expressed by ordering TPM values from highest to lowest and then divided into three groups containing the same elements. CDS IDs were correlated to obtain both TPM and RPGC tag count values. Statistical significance tests between value counts (TPM and RPGC) of expression classes were performed with the Wilcoxon–Mann–Whitney test.

Identification of open chromatin landscapes and visualization

To avoid plotting regions from small sequences, only features from the largest contigs were selected. Thus, the first 188 sequences of the genome were analyzed, which comprised 85% of the total number of bases. Genomic coordinates of each feature analyzed were obtained from the final GFF annotation (after using the annotatePolycistron Python script and UTRme) as bed files. Landscape profiles and heatmaps were plotted using the computeMatrix (with scale-regions and skipZeros options) and plotHeatmap (with hierarchical clustering) functions

from deepTools. The final GFF annotation, bed files, and RPGC-normalized data (as wig files) were used as input in IGV (Integrative Genomics Viewer) [62] for additional inspection and visualization.

Additional analysis

Gene Ontology (GO) analysis was performed by passing the feature IDs to TriTrypDB and then using the GO Enrichment tool. The obtained results were exported to REVIGO [63] for visualization. Repeat regions were obtained from <http://bioinformatica.fcien.edu.uy/cruzi/browser/downloads.php>.

Superposition of FAIRE and MNase-seq data

MNase-seq data were retrieved and processed according to our previous study [25]. The aligned bam files were RPGC-normalized using bamCoverage from deepTools, applying the following scaling factors for each sample: SAMN16241276: 1.0625, SAMN16241277: 0.9521, SAMN16241279: 0.9223, SAMN16241280: 1.0565, SAMN16241281: 1.0104, and SAMN16241282: 0.9837.

tRNA fluorescence in situ hybridization (FISH)

Parasites (10⁷) in the exponential growth phase were fixed with 4% paraformaldehyde in PBS for 20 min at room temperature. Pellets were resuspended in NH₄Cl under agitation for 10 min, washed with PBS, and adhered to slides with an adhesive frame previously treated with polylysine. Then, parasites were permeabilized with 0.2% Triton X-100 in PBS for 5 min, blocked and prehydrated with 2% BSA, 5X Denhardt, 4X SSC and 35% formamide for 2 h. A total of 1 ng per mL of fluorescent oligonucleotide was added to the hybridization solution: 1 mM EDTA, 1% BSA, 0.5 M NaHPO₄, 7% SDS, and heated at 75 °C for 5 min. For hybridization, a wet chamber was prepared with 50% formamide in 2X SSC, and the slides inside were incubated at 37 °C overnight. The slides were washed twice with 4X SSC, four times with 50% formamide and 2X SSC, three times with 2X SSC, three times with 1X SSC, and three times with 0.5X SSC. The slides were mounted with Vectashield plus DAPI solution (Vectors). Images were captured by an Olympus BX51 microscope, and the intensity of fluorescence was quantified by ImageJ with mean gray values.

Northern blot assays

Briefly, total RNA from 5 × 10⁷ parasites (epimastigotes and metacyclics forms) were extracted with TRIzol (Invitrogen) according to manufactures instructions. RNA was quantified with Qubit (Thermo). Three-hundred nanograms of RNAs or the equivalent of 7.6 × 10⁶ cells

(~2.7 µg and 0.3 µg of epimastigotes and metacyclics, respectively) were run in 15% denaturing polyacrylamide gel (6.5 M Urea, in 1 × TBE buffer) and transferred to a nylon membrane. The membrane was cross-linked in UV-light for 3 min and pre-hybridized with 5 mL of hybridization buffer (8X SSPE, 5X DENHART, 0.1% SDS and 50 µg Salmon sperm DNA) for 1 h at 30 °C. Next, the membrane was incubated with 20 pmol of biotin labeled probes (tRNA-Glu: 5'CAGGTGTTGTAACCGTTATAC CATATC3'; tRNA-Asp: 5'CGGGTATACTTACCACTA TACTACCGA3' and 5S: 5'CTTAACTTCACAAATCGG ACGGGAT3') overnight. The membrane was processed and the stained detected according to the instructions in the Chemiluminescent Nucleic Acid Detection Module Kit (Thermo). The chemiluminescence signal intensity was measured by ImageJ with mean gray values.

Abbreviations

PTU: Polycistronic transcription unit; dSSR: Divergent strand switch region; cSSR: Convergent strand switch region; TSS: Transcription start site; TTS: Transcription termination site; FAIRE: Formaldehyde-assisted isolation of regulatory elements; MNase: Micrococcal nuclease.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13072-022-00450-x>.

Additional file 1: Figure S1. A. Data summary for sequenced samples (EPI – epimastigote; META – metacyclic trypomastigote) and control for each replicate and read pairs. B. FastQC mean quality scores of sequenced samples. C. Scheme of the pipeline used to explore the FAIRE-seq data. D. Number of mapped reads against the *Trypanosoma cruzi* Dm28c genome for each dataset (biological replicates). Blue bars represent total dataset reads; orange bars indicate the mapped reads; gray bars show the remaining mapped reads with MAPQ scores above 10. The percentage is relative to the total reads of each dataset (blue bars). Spearman correlation (E) and PCA (F) of read counts for each dataset (biological replicates) from epimastigote (EPI), metacyclic trypomastigote (metacyclic) life forms and control samples generated with deepTools. **Figure S2.** A. Representative IGW snapshot of FAIRE-seq data mapped against the *T. cruzi* Dm28c genome (contig PRFA01000010). The data were normalized using RPGC (see Materials and Methods in main text). Epimastigote biological replicates (EPI-1 and EPI-2) are shown in red, and metacyclic trypomastigote biological replicates (META-1 and META-2) are shown in blue. Datasets not filtered with MAPQ score 10 are labeled as "WithoutQ10". At the genomic features, repeat regions are represented in yellow. The dashed rectangle highlights a long repeat region and its low coverage upon MAPQ10 removal. B. IGW snapshots of two contigs comparing the raw mapped FAIRE read profile to the removal of multimapped reads (Q10 - arrows) in epimastigotes and metacyclic trypomastigotes. Core and disruptive genomic compartments are shown as red and green tracks, respectively. **Figure S3.** A. Superposition of FAIRE-seq (red) and MNase-seq (blue) in epimastigotes. The y-axis represents RPGC normalized tag counts for FAIRE-seq on the right side and MNase-seq on the left side. The profile CDSs transcribed by RNA Pol II is shown. B. Comparison of core and disruptive compartments of MNase-seq data using RPGC counts in epimastigotes (red) and trypomastigotes (blue). The median values for epimastigotes were 2.04 and 1.09 for the core and disruptive compartments, respectively; for trypomastigote, the median values were 2.13 and 1.14. *** Wilcoxon-Mann-Whitney test with p-value = 0.001. C. RPGC tag counts at multigenic family genes and other CDSs transcribed by RNA Pol II for epimastigotes (red) and metacyclic trypomastigotes (blue). D. Summary plots for virulence factors

in epimastigotes and metacyclics. *** Wilcoxon-Mann-Whitney test with p-value = 0.001. **Figure S4.** A. Landscape profile of cSSR (blue) and dSSR (green) showing median values in RPGC log2 ratio of epimastigotes/metacyclics (left). Start and End represent the first and last bases of each feature, respectively. Boxplots showing RPGC counts for dSSR (upper) and cSSR (lower) for epimastigotes (red) and metacyclics (blue). *** Wilcoxon-Mann-Whitney test with p-value = 0.001. B. The percentage of bases of multifamily genes for each polycistron according to clustering is shown in Figure 4D. In B, clusterization was performed considering 1 kb upstream and downstream from the polycistron, while in C, clusterization considered only the PTU. Statistical analysis was performed using the Wilcoxon-Mann-Whitney test (*** p-value = 0.001; * p-value = 0.05). D. Landscape profiles of all CDSs (upper) and clustering based on the RPGC log2 ratio of epimastigotes/metacyclics (lower). **Figure S5.** Genes classified as high, medium and low expressed based on TPM values for epimastigote (red) and metacyclic trypomastigote (blue) life forms. *** Wilcoxon-Mann-Whitney test with p-value = 0.001. **Figure S6.** A. IGW snapshots of FAIRE-seq enrichment in epimastigotes (duplicates – red) compared to metacyclics (duplicates – blue) in three different loci. tDNAs are highlighted in yellow rectangles. B. Number of tRNA genes in each cluster. The tRNA genes grouped in each cluster after hierarchical clustering, as shown in Figure 6C, were counted and plotted in column graphics according to R group classification of the amino acids. C. Representative scheme of tRNA gene localization at divergent, convergent and codirectional PTUs. The arrows indicate the sense of transcription in each PTU. D. Quantification of tRNA-FISH UUC (Glu), GUC (Asp) in epimastigote (blue) and metacyclic (green) forms. Mean gray values represent intensity per area. Unpaired T-test: ** (p-value > 0.05), and *** (p-value < 0.0001). E. Quantification of tRNA-Northern blot UUC and GUC in Epi (blue) and MT (green) forms. The expression levels are represented by intensity per area (Mean gray values). Unpaired T-test: * (p-value < 0.05), and *** (p-value < 0.0001). ns, non-significant.

Additional file 2: Table S1. Repetitive regions at *T. cruzi* genome.

Table S2. Log2 RPGC at core and disruptive compartment. **Table S3.** RPGC at the indicated features in epimastigotes and metacyclics (Figure 4C). **Table S4.** RPGC at the indicated features in epimastigotes and metacyclics (Figure 4C). **Table S4.** Hierarchical clustering output from PTUs (with or without SSRs) and tDNA.

Additional file 3: Videos S1–S4. 3D reconstruction of euchromatin (yellow), heterochromatin (purple) and nucleolus (red) in epimastigotes, intermediate 1 and 2, and metacyclics. <https://repositorio.butantan.gov.br/handle/butantan/3965>.

Acknowledgements

We thank Karin Cruz and Ivan Novaski Avino for their technical assistance. We thank Cibele Masotti for reading this manuscript and providing critical comments, Dr. Jose Patané for initial statistical analysis, Dr. David Silva and Daniel Ohara for bioinformatics support, and Dr. Natalia Karla Bellini for metacyclics purification experiments.

Author contributions

JPCC, MCE and SB conceived the original idea. SB, JNR, HMCS and LL performed the wet experiments. ARL processed and analyzed the computational data. JNR, HMCS, HGS, ARJ, LL and JPCC analyzed the data. FH, PAG, MCM, CM, MCE and AMS helped supervise the project. JPCC supervised the project. JPCC and ARL wrote the manuscript. All authors provided critical feedback and helped shape the research, analysis and manuscript.

Funding

This work was supported by fellowships from FAPESP and by grants (#18/15553-9 (JPCC), #18/14432-3 (AMS), #13/07467-1 (CeTICS), #19/19690-3 (ARJL), # 19/19834-5 (HMCS), # 20/02708-4 (HGS) from the Sao Paulo Research Foundation (FAPESP), by the Serrapilheira Institute (grants Serra-1709-16865 and Serra-1709-16607), and Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ). ARJL and HMCS were supported by fellowships from FAPESP; SBP was supported by a fellowship from the Serrapilheira Institute. MCE has a fellowship from the National Council for Scientific and Technological Development (CNPq). The sponsors have no role in the design of the

study and collection, analysis, and interpretation of data or in writing the manuscript.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the NCBI BioProject repository under the accession code PRJNA763084 and BioSample codes SAMN21423261, SAMN21423262, SAMN21423263, SAMN21423264, SAMN21423265 and SAMN21423266.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Laboratório de Ciclo Celular, Instituto Butantan, São Paulo, SP, Brazil. ²Centro de Toxinas, Resposta Imune E Sinalização Celular (CeTICS), Instituto Butantan, São Paulo, Brazil. ³Departamento de Microbiologia, Universidade Federal de São Paulo, Escola Paulista de Medicina, Imunologia E Parasitologia, São Paulo, SP, Brazil. ⁴Laboratório de Ultraestrutura Celular Hertha Meyer, Instituto de Biofísica Carlos Chagas Filho, Universidade Federal Do Rio de Janeiro, IBCCF, CCS, UFRJ, Cidade Universitária, Rio de Janeiro, RJ, Brazil. ⁵Centro Nacional de Biologia Estrutural E Bioimagem, Rio de Janeiro, RJ, Brazil. ⁶Centro de Oncologia Molecular, Hospital Sírio Libanês, São Paulo, SP, Brazil. ⁷Instituto Carlos Chagas, Fiocruz, Curitiba, PR, Brazil. ⁸Universidade de São Paulo, São Paulo, SP, Brazil.

Received: 25 November 2021 Accepted: 18 April 2022

Published: 1 June 2022

References

- Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res*. 2011;21(3):381–95.
- Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, et al. Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature*. 2011;471(7339):480–5.
- Simon JM, Giresi PG, Davis IJ, Lieb JD. Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. *Nat Protoc*. 2012;7(2):256–67.
- Fong N, Saldi T, Sheridan RM, Cortazar MA, Bentley DL. RNA Pol II dynamics modulate co-transcriptional chromatin modification, CTD phosphorylation, and transcriptional direction. *Mol Cell*. 2017;66(4):546–557.e3.
- Giresi PG, Kim J, McDaniel RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res*. 2007;17(6):877–85.
- Clayton C. Regulation of gene expression in trypanosomatids: living with polycistronic transcription. *Open Biol*. 2019;9:1–24.
- Wedel C, Förstner KU, Derr R, Siegel TN. GT-rich promoters can drive RNA pol II transcription and deposition of H2A.Z in African trypanosomes. *EMBO J*. 2017;36(17):2581–94.
- Respuela P, Ferella M, Rada-iglesias A. Histone acetylation and methylation at sites initiating divergent polycistronic transcription in *Trypanosoma cruzi* * □. *J Biol Chem*. 2008;283:15884–92.
- Siegel TN, Hekstra DR, Kemp LE, Figueiredo LM, Lowell JE, Fenyo D, et al. Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. *Genes Dev*. 2009;23(9):1063–76.
- Wright JR, Siegel TN, Cross GAM. Histone H3 trimethylated at lysine 4 is enriched at probable transcription start sites in *Trypanosoma brucei*. *Mol Biochem Parasitol*. 2010;172:141–4.
- Reynolds D, Hofmeister BT, Cliffe L, Alabady M, Siegel TN, Schmitz RJ, et al. Histone H3 variant regulates RNA polymerase II transcription termination and dual strand transcription of siRNA loci in *Trypanosoma brucei*. *PLoS Genet*. 2016;12:1–25.
- Cliffe LJ, Siegel TN, Marshall M, Cross GAM, Sabatini R. Two thymidine hydroxylases differentially regulate the formation of glucosylated DNA at regions flanking polymerase II polycistronic transcription units throughout the genome of *Trypanosoma brucei*. *Nucl Acids Res*. 2010;38(2):3923–35.
- Kaufer A, Ellis J, Stark D, Barratt J. The evolution of trypanosomatid taxonomy. *Parasites Vectors*. 2017;10:1–17.
- Jimenez V. Dealing with environmental challenges: mechanisms of adaptation in *Trypanosoma cruzi*. *Res Microbiol*. 2014;165:155–65.
- Gonçalves CS, Ávila AR, De Souza W, Motta MCM, Cavalcanti DP. Revisiting the *Trypanosoma cruzi* metacyclogenesis: morphological and ultrastructural analyses during cell differentiation. *Parasites Vectors*. 2018;11:1–14.
- Souza W. Basic cell biology of *Trypanosoma cruzi*. *Curr Pharm Des*. 2005;8(4):269–85.
- Elias MCQB, Marques-Porto R, Freymüller E, Schenkman S. Transcription rate modulation through the *Trypanosoma cruzi* life cycle occurs in parallel with changes in nuclear organisation. *Mol Biochem Parasitol*. 2001;112(1):79–90.
- Ferreira LRP, Dossin FDM, Ramos TC, Freymüller E, Schenkman S. Active transcription and ultrastructural changes during *Trypanosoma cruzi* metacyclogenesis. *An Acad Bras Cienc*. 2008;80:157–66.
- Smircich P, Eastman G, Bispo S, Duhagon MA, Guerra-Slompo EP, Garat B, et al. Ribosome profiling reveals translation control as a key mechanism generating differential gene expression in *Trypanosoma cruzi*. *BMC Genom*. 2015. <https://doi.org/10.1186/s12864-015-1563-8>.
- Chung J, Rocha AA, Tonelli RR, Castilho BA, Schenkman S. Eukaryotic initiation factor 5A dephosphorylation is required for translational arrest in stationary phase cells. *Biochem J*. 2013;451(2):257–67.
- Tonelli RR, da Silva Augusto L, Castilho BA, Schenkman S. Protein synthesis attenuation by phosphorylation of eIF2 α is required for the differentiation of *Trypanosoma cruzi* into infective forms. *PLoS ONE*. 2011;6(11):e27904.
- Berná L, Rodríguez M, Chiribao ML, Parodi-Talice A, Pita S, Rijo G, et al. Expanding an expanded genome: long-read sequencing of *Trypanosoma cruzi*. *Microb Genom*. 2018;13(1):531.
- Berná L, Chiribao ML, Greif G, Rodríguez M, Alvarez-Valin F, Robello C. Transcriptomic analysis reveals metabolic switches and surface remodeling as key processes for stage transition in *Trypanosoma cruzi*. *PeerJ*. 2017;2017:1–32.
- De Jesus TCL, Nunes VS, Lopes MDC, Martil DE, Iwai LK, Moretti NS, et al. Chromatin proteomics reveals variable histone modifications during the life cycle of *Trypanosoma cruzi*. *J Proteome Res*. 2016;15(6):2039–51.
- Lima ARJ, de Araujo CB, Bispo S, Patané J, Silber AM, Elias MC, et al. Nucleosome landscape reflects phenotypic differences in *Trypanosoma cruzi* life forms. *PLoS Pathog*. 2021;17(11):e1009272.
- Rosón JN, de Oliveira Vitarelli M, Costa-Silva HM, Pereira KS, da Silva Pires D, de Sousa Lopes L, et al. Histone H2B.V demarcates strategic regions in the *Trypanosoma cruzi* genome, associates with a bromodomain factor and affects parasite differentiation and host cell invasion. *bioRxiv*. 2021;2021.06.08.447515.
- Kolev NG, Franklin JB, Carmi S, Shi H, Michaeli S, Tschudi C. The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. *PLoS Pathog*. 2010;6(9):e1001090.
- Martínez-Calvillo S, Yan S, Nguyen D, Fox M, Stuart K, Myler PJ. Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. *Mol Cell*. 2003;11(5):1291–9.
- Garcia-Silva MR, Frugier M, Tosar JP, Correa-Dominguez A, Ronalte-Alves L, Parodi-Talice A, et al. A population of tRNA-derived small RNAs is actively produced in *Trypanosoma cruzi* and recruited to specific cytoplasmic granules. *Mol Biochem Parasitol*. 2010;171(2):64–73.
- Falk M, Feodorova Y, Naumova N, Imakaev M, Lajoie BR, Leonhardt H, et al. Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature*. 2019;118:098002.
- Fortin JP, Hansen KD. Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. *Genome Biol*. 2015;16(1):1–23.
- Nagy G, Dániel B, Jónás D, Nagy L, Barta E. A novel method to predict regulatory regions based on histone mark landscapes in macrophages. *Immunobiology*. 2013;218(11):1416–27.

33. Abuin G, Freitas LHG, Colli W, Alves MJM, Schenkman S. Expression of trans-sialidase and 85-kDa glycoprotein genes in *Trypanosoma cruzi* is differentially regulated at the post-transcriptional level by labile protein factors. *J Biol Chem*. 1999;274(19):13041–7.
34. Jäger AV, Muiá RP, Campetella O. Stage-specific expression of *Trypanosoma cruzi* trans-sialidase involves highly conserved 3′ untranslated regions. *FEMS Microbiol Lett*. 2008;283(2):182–8.
35. Tillo D, Hughes TR. G+C content dominates intrinsic nucleosome occupancy. *BMC Bioinform*. 2009. <https://doi.org/10.1186/1471-2105-10-442>.
36. Ekanayak D, Sabatini R. Epigenetic regulation of polymerase II transcription initiation in *Trypanosoma cruzi*: Modulation of nucleosome abundance, histone modification, and polymerase occupancy by α -linked thymine DNA glucosylation. *Eukaryot Cell*. 2011;10(11):1465–72.
37. Ekanayake DK, Cipriano MJ, Sabatini R. Telomeric co-localization of the modified base J and contingency genes in the protozoan parasite *Trypanosoma cruzi*. *Nucl Acids Res*. 2007;35(19):6367–77.
38. Lombrana R, Álvarez A, Fernández-Justel JM, Almeida R, Poza-Carrión C, Gomes F, et al. Transcriptionally driven DNA replication program of the human parasite *leishmania major*. *Cell Rep*. 2016;16(6):1774–86.
39. Marchetti MA, Tschudi C, Silva E, Ullu E. Physical and transcriptional analysis of the *Trypanosoma brucei* genome reveals a typical eukaryotic arrangement with close interspersed of RNA polymerase II- and III-transcribed genes. *Nucl Acids Res*. 1998;26(15):3591–8.
40. O’Mullane L, Eperon IC. The pre-mRNA 5′ cap determines whether U6 small nuclear RNA succeeds U1 small nuclear ribonucleoprotein particle at 5′ splice sites. *Mol Cell Biol*. 1998;18(12):7510–20.
41. Padilla-Mejía NE, Florencio-Martínez LE, Figueroa-Angulo EE, Manning-Cela RG, Hernández-Rivas R, Myler PJ, et al. Gene organization and sequence analyses of transfer RNA genes in trypanosomatid parasites. *BMC Genom*. 2009. <https://doi.org/10.1186/1471-2164-10-232>.
42. Male G, Von Appen A, Glatt S, Taylor NMI, Cristovao M, Groetsch H, et al. Architecture of TFIIIC and its role in RNA polymerase III pre-initiation complex assembly. *Nat Commun*. 2015. <https://doi.org/10.1038/ncomms8387>.
43. Van Bortle K, Corces VG. tDNA insulators and the emerging role of TFIIIC in genome organization. *Transcription*. 2012;3(6):277–84.
44. Vasquez JJ, Hon CC, Vanselow JT, Schlosser A, Siegel TN. Comparative ribosome profiling reveals extensive translational complexity in different *Trypanosoma brucei* life cycle stages. *Nucl Acids Res*. 2014;42(6):3623–7.
45. Brackley CA, Romano MC, Thiel M. The dynamics of supply and demand in mRNA translation. *PLoS Comput Biol*. 2011;7(10):e1002203.
46. Shukla A, Bhargava P. Regulation of tRNA gene transcription by the chromatin structure and nucleosome dynamics. *Biochim Biophys Acta Gene Regul Mech* [Internet]. 2018;1861:295–309. <https://www.sciencedirect.com/science/article/pii/S1874939917302572>
47. Camargo EP. Growth and differentiation in *Trypanosoma cruzi*. I. Origin of metacyclic trypanosomes in liquid media. *Rev Inst Med Trop Sao Paulo*. 1964;12:93–100.
48. Contreras VT, Salles JM, Thomas N, Morel CM, Goldenberg S. In vitro differentiation of *Trypanosoma cruzi* under chemically defined conditions. *Mol Biochem Parasitol*. 1985;16(3):315–27.
49. Andrey P, Maurin Y. Free-D: An integrated environment for three-dimensional reconstruction from serial sections. *J Neurosci Methods*. 2005;145:233–44.
50. Andrews S. FastQC: A quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. 2010.
51. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From fastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinform*. 2013. <https://doi.org/10.1002/0471250953.bi1110s43>.
52. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
53. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
54. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucl Acids Res*. 2016;44(W1):W160–5.
55. Radio S, Fort RS, Garat B, Sotelo-Silveira J, Smircich P. UTRme: a scoring-based tool to annotate untranslated regions in trypanosomatid genomes. *Front Genet*. 2018. <https://doi.org/10.3389/fgene.2018.00671>.
56. McCarthy-Burke C, Taylor ZA, Buck GA. Characterization of the spliced leader genes and transcripts in *Trypanosoma cruzi*. *Gene*. 1989;82(1):177–89.
57. Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucl Acids Res*. 2005;33:W686–9.
58. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011;17(1):10.
59. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* [Internet]. 2012/10/25. Oxford University Press; 2013;29:15–21. <https://www.ncbi.nlm.nih.gov/pubmed/23104886>
60. Alvarez RV, Pongor LS, Mariño-Ramírez L, Landsman D. TPMCalculator: one-step software to quantify mRNA abundance of genomic features. *Bioinformatics*. 2019;35(11):1960–2.
61. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576–89.
62. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14(2):178–92.
63. Supek F, Bošnjak M, Škunca N, Šmuc T. Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*. 2011;6(7):e21800.

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

