

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA
DISSERTAÇÃO DE MESTRADO

WANESSA MOREIRA GOES

**EFEITO DA DELEÇÃO DO GENE QUE CODIFICA UMA PROTEÍNA DE LIGAÇÃO
À RNA SOBRE O TRANSCRIPTOMA DE EPIMASTIGOTAS DE
*TRYPANOSOMA CRUZI***

BELO HORIZONTE

2018

WANESSA MOREIRA GOES

**EFEITO DA DELEÇÃO DO GENE QUE CODIFICA UMA PROTEÍNA DE LIGAÇÃO
À RNA SOBRE O TRANSCRIPTOMA DE EPIMASTIGOTAS DE
*TRYPANOSOMA CRUZI***

Dissertação de Mestrado apresentada ao programa interunidades de Pós-Graduação em Bioinformática da Universidade Federal de Minas Gerais como requisito para obtenção do título de Mestre em Bioinformática.

Orientadora: Prof^a. Dra. Santuza Maria Ribeiro Teixeira
Co-Orientador: Dr. Fabiano Sviatopolk-Mirsky Pais

BELO HORIZONTE

2018

À minha mãe, Marivone Moreira dos Santos, pelo apoio incondicional e pela disposição em todos os momentos durante a realização deste trabalho.

AGRADECIMENTOS

À Prof^a. Dra. Santuza Maria Ribeiro Teixeira por ter me aceito como aluna, pela orientação em todas as etapas deste trabalho, pelo incentivo na busca do conhecimento, pela contribuição em minha formação profissional e pessoal e pelos bons momentos vividos dentro do laboratório e fora deste.

Ao Dr. Fabiano Sviatopolk-Mirsky Pais pela co-orientação durante o desenvolvimento deste trabalho, por toda a experiência em bioinformática transferida, pelo apoio, preocupação e incentivo durante todo o tempo.

Ao meu colega, Antonio Edson Rocha Oliveira que me deu uma enorme contribuição em todas as análises de bioinformática e que com muita paciência me ensinou vários aspectos da área de biologia molecular e bioinformática. Agradeço pela contribuição na escrita deste trabalho, pelo incentivo e humildade em sempre estar disposto a ajudar.

À Bruna Mattioly Valente pela ajuda durante todo o desenvolvimento deste trabalho, pelas dúvidas esclarecidas e pelos momentos vividos durante o período do mestrado.

Aos meus colegas de laboratório, Carlos Andrés Toro Acevedo, Gabriela Burle, Melissa, Nailma Silva Aprigio dos Santos, Rafael André Ferreira, Renata Barbosa Peixoto, Thaís e Viviane, pela companhia durante todo o período em que permaneci no laboratório.

Aos meus pais, Marivone Moreira dos Santos e Marcos Linhares Goes, pelo apoio incondicional, pelo incentivo e pela contribuição financeira e enriquecimento científico.

Ao meu irmão Allyson Moreira Goes pela contribuição na construção de algumas figuras deste trabalho.

À Francielle Resende de Oliveira pela ajuda na formatação e correção gramatical do trabalho, apoio emocional e por ter estado comigo em todas as horas.

Ao programa de Pós-Graduação em Bioinformática da UFMG, em especial à Sheila, Fernanda, Natália, Amanda e Tiago, pelo carinho, pela disponibilidade e pelo esforço para solucionar todos os problemas que apareceram durante o período do Mestrado.

À agência financiadora Fapemig pela bolsa, que me auxiliou durante a minha estadia em Belo Horizonte e na realização deste trabalho.

“Não posso dar a nenhum cientista de qualquer idade melhor conselho do que este:
a intensidade da convicção de que uma hipótese é verdadeira não tem nenhuma
relação com se é ou não verdadeira. ”

- Peter Medawar

“Por que esta magnífica tecnologia científica, a que economiza trabalho e torna a
vida mais fácil, nos dá tão pouca felicidade? A resposta é esta: simplesmente,
porque nós ainda não aprendemos a usá-la sabiamente. ”

- Albert Einstein

RESUMO

Trypanosoma cruzi, o agente etiológico da doença de Chagas é um protozoário que apresenta três formas durante o seu ciclo de vida, as quais são bioquimicamente e morfologicamente distintas e programadas para responder rapidamente às drásticas mudanças ambientais que esse parasita enfrenta em seus vários hospedeiros. Ao contrário de outros eucariotos, os genes que codificam proteínas neste protozoário são transcritos em pré-mRNAs policistônicos, os quais são processados em mRNAs maduros por meio de reações acopladas de "trans-splicing" e de poli-adenilação. Devido a isso, o controle da expressão gênica depende principalmente de mecanismos pós-transcricionais mediados por proteínas de ligação ao RNA (RBPs), que controlam os níveis *steady-state* e as taxas de tradução dos mRNAs. RNA-seq de epimastigotas, tripomastigotas e amastigotas, juntamente com buscas por motivos característicos de RBPs permitiu a identificação de uma proteína com níveis de expressão aumentados na forma epimastigota, quando comparados às formas amastigotas e tripomastigotas. Essa proteína foi nomeada TcRBP99. Experimentos de metaciclo gênese comparando epimastigotas do tipo selvagem (WT) e epimastigotas nas quais o gene que codifica TcRBP99 foi eliminado identificou um papel relacionado à diferenciação do parasito uma vez que um aumento na taxa de metaciclo gênese foi observado em parasitas nocautes. Com base nesses dados, o presente trabalho teve como objetivo analisar o efeito da deleção do gene que codifica a TcRBP99, no controle da expressão gênica de formas epimastigotas de *T. cruzi*, buscando identificar genes cuja expressão estaria associada à metaciclo gênese. Foram conduzidas análises de RNA-seq que comparam a expressão gênica de epimastigotas WT e epimastigotas de duas linhagens nocautes para essa RBP. Bibliotecas de cDNA foram sequenciadas na plataforma Illumina MiSeq e em seguida foram conduzidas análises de qualidade das *reads* geradas. Posteriormente, realizou-se mapeamento das *reads* contra o genoma completo de CL Brener, por meio da ferramenta TopHat2 e análises de expressão gênica diferencial utilizando os seguintes pacotes: EdgeR, limma e Deseq2. Para a seleção de genes diferencialmente expressos (DGE), foram definidos os parâmetros p-ajustado (p_{adj}) < 0,05 e mudança de expressão maior ou menor que 2 vezes. Por fim, regiões 3' UTR foram determinadas, possibilitando a identificação de motivos, por meio das ferramentas MEME e MAST. Como resultados, obtivemos em média, 3,6 milhões de *reads* com qualidade baseada no *phred score* igual ou maior que 20 para cada uma das sete bibliotecas, e mapeamento, em média, de 73,28% dessas *reads*. Análises de expressão diferencial revelaram 9 genes que apresentaram expressão reduzida em epimastigotas nocautes em comparação com WT: um deles codifica para a proteína associada à diferenciação, cujo mRNA está aumentado em epimastigotas WT comparadas com tripomastigotas e amastigotas. Dois genes da família RHS e um gene que codifica transportador de aminoácidos também tiveram a expressão diminuída na ausência da TcRBP99. Análises de motivos em regiões 3' UTRs desses mRNAs mostraram que há três motivos conservados, mas que nenhum está presente nas quatro UTRs, ao mesmo tempo. Os resultados obtidos neste trabalho sugerem que a TcRBP99, ao se ligar a mRNAs em formas epimastigotas causa a estabilização desses mRNAs, aumentando assim a expressão de genes responsáveis pela proliferação de epimastigotas, o que conseqüentemente leva a diminuição da taxa de diferenciação de epimastigotas para tripomastigotas metacíclicas.

Palavras-chave: *Trypanosoma cruzi*, RBP, RNA-seq, 3' UTR.

ABSTRACT

Trypanosoma cruzi, the etiological agent of Chagas' disease, is a protozoan that presents three forms during its life cycle, which are biochemically and morphologically distinct and programmed to respond quickly to the drastic environmental changes that this parasite faces in its various hosts. Unlike other eukaryotes, the genes encoding proteins in this protozoan are transcribed into polycistronic pre-mRNAs, which are processed into mature mRNAs by coupled trans-splicing and polyadenylation reactions. Because of this, control of gene expression depends primarily on post-transcriptional mechanisms mediated by RNA binding proteins (RBPs), which control steady-state levels and translation rates of mRNAs. RNA-seq of epimastigotes, trypomastigotes and amastigotes, together with searches for motifs characteristic of RBPs allowed the identification of a protein with increased levels of expression in the epimastigote form, when compared to amastigote and trypomastigote forms. This protein was named TcRBP99. Metacyclogenesis experiments comparing wild-type epimastigotes (WT) and epimastigotes in which the gene encoding TcRBP99 was deleted identified a role related to parasite differentiation since an increase in the rate of metacyclogenesis was observed in parasitic knockouts. Based on these data, the present work had the objective of analyzing the effect of the deletion of the gene encoding TcRBP99, in the control of the gene expression of epimastigote forms of *T. cruzi*, in order to identify genes whose expression would be associated with metacyclogenesis. RNA-seq analyzes were performed comparing the gene expression of WT epimastigotes and epimastigotes from two knockout lines for this RBP. cDNA libraries were sequenced on the Illumina MiSeq platform and then quality analyzes of the generated reads were conducted. Later, the reads were mapped against the complete genome of CL Brener, using the TopHat2 tool and differential gene expression analyzes using the following packages: EdgeR, limma and Deseq2. For the selection of differentially expressed genes (GDE), the parameters p-adjusted (padj) <0.05 and change of expression greater than or less than 2 times were defined. Finally, 3' UTR regions were determined, allowing the identification of motifs, using the MEME and MAST tools. As a result, we obtained, on average, 3.6 million reads with quality based on the phred score equal to or greater than 20 for each of the seven libraries, and an average mapping of 73.28% of these reads. Differential expression analyzes revealed 9 genes that showed reduced expression in epimastigotes knockouts compared to WT: one of them codes for the differentiation-associated protein, whose mRNA is increased in WT epimastigotes compared to trypomastigotes and amastigotes. Three RHS family genes and one amino acid transporter coding gene also had diminished expression in the absence of TcRBP99. Analysis of motifs in 3' UTRs regions of these mRNAs showed that there are three conserved motifs, but none are present in the four UTRs at the same time. The results obtained in this work suggest that the TcRBP99, when binding to mRNAs in epimastigote forms causes the stabilization of these mRNAs, thus increasing the expression of genes responsible for the proliferation of epimastigotes, which consequently leads to a decrease in the rate of epimastigote differentiation for trypomastigotes metacyclic.

Key-words: *Trypanosoma cruzi*, RBP, RNA-seq, 3' UTR.

LISTA DE FIGURAS

Figura 1. Distribuição de casos da doença de Chagas relatados entre os anos de 2010-2013.	2
Figura 2. Ciclo de vida de <i>T. cruzi</i> no hospedeiro humano.	4
Figura 3. Manifestações clínicas da doença de Chagas.	7
Figura 4. Proteínas de ligação a RNA (RBPs) com expressão aumentada em epimastigotas.	22
Figura 5. Esquema representativo da técnica de RNA-Seq.	30
Figura 6. <i>Pipeline</i> para análise de 3' UTR de <i>Trypanosoma cruzi</i>	39
Figura 7. Níveis de expressão do alelo TcCLB.506739 do clone CL Brener nas diferentes formas de vida de <i>T. cruzi</i>	42
Figura 8. Alinhamento local das sequências correspondentes aos alelos TcCLB.506739.99 e TcCLB.510819.119 de <i>T. cruzi</i>	44
Figura 9. Alinhamento múltiplo das sequências dos genes sintênicos a TcRBP99 depositados no banco de dados TritrypDB.	45
Figura 10. Análise de Componente Principal (PCA) das bibliotecas de cDNA.	47
Figura 11. <i>HeatMap</i> das bibliotecas de cDNA.	47
Figura 12. Gráficos de qualidade por base de <i>reads</i> brutas provenientes de epimastigotas nocautes para TcRBP99.	49
Figura 13. Gráficos de qualidade por base de <i>reads</i> brutas provenientes de epimastigotas WT.	50
Figura 14. Gráficos de qualidade por base de <i>reads</i> trimadas provenientes de epimastigotas nocautes para TcRBP99.	51
Figura 15. Gráficos de qualidade por base de <i>reads</i> trimadas provenientes de epimastigotas WT.	52
Figura 16. Diagrama de Veen - Quantidade de genes diferencialmente expressos identificados com os pacotes EdgeR, limma e Deseq2.	56
Figura 17. Níveis de expressão gênica em <i>T. cruzi</i>	57
Figura 18. Genes diferencialmente expressos em epimastigotas nocautes comparadas com epimastigotas WT, segundo parâmetros de padj e LogFoldChange (FC) pré-estabelecidos.	58
Figura 19. Genes diferencialmente expressos em epimastigotas nocautes comparadas com epimastigotas WT estatisticamente significativos (padj < 0,05). ...	60
Figura 20. Níveis de expressão de possíveis genes alvos da TcRBP99 ao longo do ciclo de vida <i>T. cruzi</i>	61

Figura 21. <i>HeatMap</i> de 8 membros da família multigênicas dos transportadores de aminoácidos gerado a partir de dados de expressão de epimastigotas WT e nocautes.	62
Figura 22. <i>HeatMap</i> dos 17 membros da família multigênica das RHS gerado a partir de dados de expressão das formas epimastigotas, amastigotas coletadas 60 e 96 horas após a infecção e tripomastigotas (BELEW et al., 2017).	63
Figura 23. <i>HeatMap</i> de 6 membros da família multigênicas das proteínas associadas à diferenciação gerado a partir de dados de expressão epimastigotas WT e nocautes.	64
Figura 24. <i>HeatMap</i> dos 13 membros da família multigênicas das proteínas associadas à diferenciação gerado a partir de dados de expressão das formas epimastigotas, amastigotas coletadas 60 e 96 horas após a infecção e tripomastigotas (BELEW et al., 2017).	65
Figura 25. <i>HeatMap</i> de 26 membros da família multigênicas das RHS gerado a partir de dados de expressão de epimastigotas WT e nocautes.	66
Figura 26. <i>HeatMap</i> dos 524 membros da família multigênicas das RHS gerado a partir de dados de expressão das formas epimastigotas, amastigotas coletadas 60 e 96 horas após a infecção e tripomastigotas (BELEW et al., 2017).....	67
Figura 27. Visualização do mapeamento de <i>reads</i> de epimastigotas, amastigotas 96 horas pós infecção de células HFF e <i>reads</i> cuja extremidades foram aparadas após a presença de sequências de mais de 4 adeninas ou 4 timinas, correspondentes as duas primeiras bibliotecas citadas.	71
Figura 28. Visualização do mapeamento de <i>reads</i> de epimastigotas, amastigotas 96 horas pós infecção de células HFF e <i>reads</i> cuja extremidades foram aparadas após a presença de sequências de mais de 4 adeninas ou 4 timinas, correspondentes as duas primeiras bibliotecas citadas.	72
Figura 29. Visualização do mapeamento de <i>reads</i> de epimastigotas, amastigotas 96 horas pós infecção de células HFF e <i>reads</i> cuja extremidades foram aparadas após a presença de sequências de mais de 4 adeninas ou 4 timinas, correspondentes as duas primeiras bibliotecas citadas.	73
Figura 30. Árvore filogenética de sequências nucleotídicas das regiões 3' UTRs de mRNAs cuja expressão foi possivelmente diminuída pela ação da TcRBP99.	74
Figura 31. Árvore filogenética de sequências nucleotídicas das regiões 3' UTRs de mRNAs cuja expressão foi possivelmente diminuída pela ação da TcRBP99.	75
Figura 32. Alinhamento das sequências correspondentes às regiões CDS de mRNAs cuja expressão foi possivelmente diminuída pela ação da TcRBP99.	76
Figura 33. Árvore filogenética das regiões CDS de mRNAs cuja expressão foi possivelmente diminuída pela ação da TcRBP99.	77
Figura 34. Logo dos motivos descobertos entre as sequências de regiões 3' UTR dos possíveis mRNAs alvos da TcBP99.	78
Figura 35. Visualização de ocorrências dos motivos descobertos pela ferramenta MEME.	78

Figura 36. Logo dos motivos considerados na análise de alinhamento realizados através da ferramenta MAST.	79
Figura 37. Diagrama dos motivos preditos presentes nas regiões 3' UTRs dos possíveis mRNAs de alvos da TcRBP99.	79
Figura 38. Epimastigotas do tipo selvagem <i>versus</i> epimastigotas nocautes para a TcRBP99.	91

LISTA DE TABELAS

Tabela 1. Principais características dos alelos que codificam a TcRBP99.	43
Tabela 2. Resultados do sequenciamento das bibliotecas de cDNA de duas linhagens de epimastigotas nocautes e epimastigotas WT.	46
Tabela 3. Número total de <i>reads</i> mapeadas contra o genoma <i>Esmeraldo-like</i> e <i>Non Esmeraldo-like</i> de <i>T. cruzi</i>	53
Tabela 4. Contagem de <i>reads</i> mapeadas para cada gene do genoma <i>Esmeraldo-like</i> e <i>Non Esmeraldo-like</i> de <i>T. cruzi</i> de acordo com o arquivo GFF.	54
Tabela 5. Genes com expressão aumentada em epimastigotas nocaute para TcRBP99 comparado com epimastigotas <i>wild-type</i>	58
Tabela 6. Genes com expressão diminuída em epimastigotas nocaute para TcRBP99 comparado com epimastigotas <i>wild-type</i>	59
Tabela 7. Relação entre regiões codificadoras e regiões 3' UTRs dos possíveis genes alvos de regulação exercida pela TcRBP99.	70
Tabela 8. <i>Script</i> utilizado para a realização de análise de expressão diferencial.	115 e 116

LISTA DE SIGLAS E ABREVIATURAS

BAC.....	do inglês <i>Bacterial Artificial Chromosome</i>
BAM.....	do inglês <i>Binary Alignment/Map</i>
cDNA.....	DNA complementar
CDS.....	do inglês <i>Coding DNA Sequence</i>
DTU.....	do inglês <i>Discrete Typing Units</i>
EST.....	do inglês <i>Expressed Sequence Tag</i>
FDR.....	do inglês <i>False Discovery Rate</i>
GDE.....	Genes Diferencialmente Expressos
GFF.....	do inglês <i>General Feature Format</i>
HFF.....	do inglês <i>Human Foreskin Fibroblast</i>
HTML.....	do inglês <i>Hypertext Markup Language</i>
IGV.....	do inglês <i>Integrative Genomics Viewer</i>
LogFC.....	do inglês <i>LogFoldChange</i>
LTR.....	do inglês <i>Non-long Terminal Repeat</i>
MAST.....	do inglês <i>Motif Alignment & Search Tool</i>
MEME.....	do inglês <i>Multiple EM for Motif Elicitation</i>
NGS.....	do inglês <i>Next Generation Sequencing</i>
ORF.....	do inglês <i>Open Reading Frame</i>
ORF Finder.....	do inglês <i>Open Reading Frame Finder</i>
PAD.....	Proteína associada à diferenciação
PCA.....	do inglês <i>Principal Component Analysis</i>
PCR.....	do inglês <i>Polymerase Chain Reaction</i>
PNG.....	do inglês <i>Portable Network Graphics</i>
PFGE.....	do inglês <i>Pulsed-field Gel Electrophoresis</i>
RBP.....	do inglês <i>RNA Binding Protein</i>
RNAPII.....	RNA Polimerase II
RNAi.....	RNA interferente
RNP.....	Ribonucleoproteína
RRM.....	do inglês <i>RNA Recognition Motif</i>
SAS.....	do inglês <i>Splice-Acceptor Site</i>
SAM.....	do inglês <i>Sequence Alignment/Map</i>

SL..... do inglês *Spliced Leader*
TMM.....do inglês *Trimmed mean of M-values*
TritypDB.....do inglês *TritypDataBase*
URE.....do inglês *U-rich elements*
TS.....Trans-sialidase
UTR.....do inglês *Untranslated Regions*
XML.....do inglês *Extensible Markup Language*
WGS.....do inglês *Whole-Genome Shotgun*
WT.....do inglês *Wild-type*

SUMÁRIO

1. INTRODUÇÃO	1
1.1 <i>Trypanosoma cruzi</i> e a Doença de Chagas	1
1.1.1 Aspectos gerais da doença: epidemiologia e transmissão	1
1.1.2 Aspectos clínicos da doença	5
1.2 Genoma nuclear do <i>Trypanosoma cruzi</i>	8
1.3 Transcrição em Tripanossomatídeos e a regulação da expressão gênica	11
1.4 Proteína de ligação ao RNA como reguladores chaves da expressão gênica em <i>T. cruzi</i>	14
1.5 Estudos do transcriptoma nas diferentes fases do ciclo de vida do <i>T. cruzi</i>	20
1.6 Análise de expressão diferencial e suas implicações	23
2. JUSTIFICATIVA	26
3. OBJETIVOS	28
3.1 Objetivo Geral	28
3.2 Objetivos específicos	28
4. METODOLOGIA	29
4.1 Isolamento de RNA total e construção de bibliotecas	29
4.2 Análise da qualidade das <i>reads</i>	31
4.3 Trimagem e filtragem de <i>reads</i> com bases de baixa qualidade	31
4.4 Mapeamento das <i>reads</i>	32
4.5 Contagem das <i>reads</i> por gene	33
4.6 Análise das contagens e expressão diferencial	34
4.7 Mapeamento e alinhamento das regiões 3' UTR de genes diferencialmente expressos	36
4.8 Alinhamento múltiplo local de regiões 3' UTR de genes alvos da TcRBP99 e comparação com as CDS correspondentes	39
4.9 Predição de motivos na região 3' utr	40
5. RESULTADOS	42
5.1 Identificação dos níveis de expressão e análises <i>in silico</i> da TcRBP99	42
5.2 Análise de <i>reads</i> obtidas no sequenciamento de bibliotecas de cDNA de epimastigotas WT e nocautes para a TcRBP99	45
5.3 Mapeamento e contagem de <i>reads</i> alinhadas ao genoma do clone CL Brener de <i>T. cruzi</i>	53
5.4 Análise de expressão diferencial, identificação de possíveis mRNA alvos da TcRBP99 e comparação de pacotes utilizados	55
5.5 Determinação das regiões 3' UTRs de mRNAs possíveis alvos da TcRBP99	68
5.6 Alinhamento múltiplo local das regiões 3' UTRs	74
5.7 Sequências ricas em uracila e citosina presentes na região 3' UTRs	77
6. DISCUSSÃO	81
7. PERSPECTIVAS	93
REFERÊNCIAS BIBLIOGRÁFICAS	94
APÊNDICE	115

1. INTRODUÇÃO

1.1 *Trypanosoma cruzi* e a Doença de Chagas

1.1.1 Aspectos gerais da doença: epidemiologia e transmissão

A doença de Chagas ou Tripanossomíase Americana, causada pelo *Trypanosoma cruzi*, foi descoberta há mais de 100 anos pelo cientista Carlos Justiniano Ribeiro Chagas. Considerada um importante problema de saúde pública, a doença de Chagas é classificada como uma doença negligenciada e apresenta-se de forma endêmica na América Latina (principalmente na Argentina, Belize, República Bolivariana da Venezuela, Brasil, Chile, Colômbia, México, Costa Rica, Equador, El Salvador, Guiana Francesa, Guatemala, Guiana, Honduras, Nicarágua, Panamá, Paraguai, Peru, Bolívia, Suriname e Uruguai), mas existem também registros de casos na região sul dos Estados Unidos (RASSI JR et al., 2012; WHO, 2015a).

De acordo com PAN AMERICAN HEALTH ORGANIZATION e WORLD HEALTH ORGANIZATION, (2017) existem cerca de 6 a 7 milhões de pessoas infectadas pelo *T. cruzi*, e uma média de 30 mil novos casos surgem por ano. Há ainda, uma estimativa de 70 milhões de pessoas, nas Américas, vivendo em área de risco. Além disso, cerca de 14 mil mortes são reportadas por ano e 8.000 recém-nascidos adquirem a doença ainda na gestação (WHO, 2017). A distribuição global de ocorrências da doença de Chagas é ilustrada na Figura 1.

O *T. cruzi* é um parasita intracelular obrigatório que apresenta quatro formas de vida: epimastigota, tripomastigota metacíclica, amastigota e tripomastigota extracelular, sendo a primeira, a forma replicativa que se desenvolve exclusivamente no hospedeiro invertebrado do parasita. Ao contrário, as amastigotas representam a forma replicativa que é encontrada no interior de células nucleadas de hospedeiros vertebrados infectados. Por último, a forma flagelada tripomastigota representa o parasita em dois momentos ao longo do ciclo de vida. No primeiro, quando originado a partir da diferenciação de epimastigotas, no inseto vetor, recebendo o nome de tripomastigotas metacíclicas, sendo esta a forma replicativa e infectiva para o hospedeiro vertebrado. Um segundo tipo de tripomastigotas surgem no hospedeiro

mamífero pela diferenciação de amastigota intracelulares em tripomastigotas extracelulares, sendo esta infectiva para o inseto triatomíneo (BERN, 2015; CENTERS FOR DISEASE CONTROL AND PREVENTION, 2017).

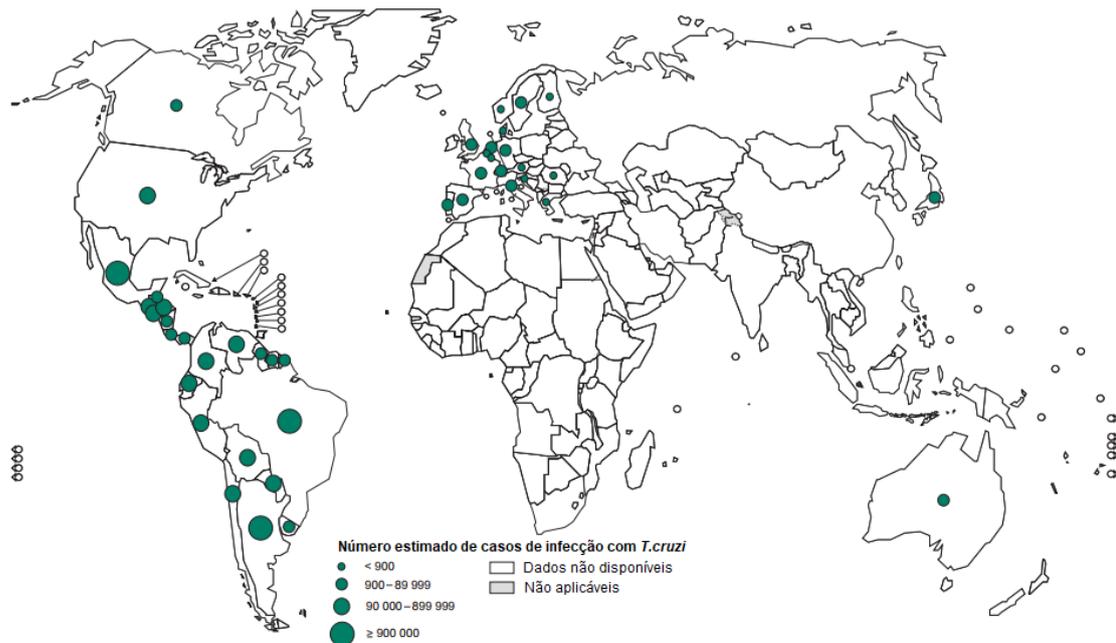


Figura 1. Distribuição de casos da doença de Chagas relatados entre os anos de 2010-2013.

Fonte: *World Health Organization* (2015).

Considerada uma antropozoonose, a transmissão do *T. cruzi* para o ser humano e outras espécies de mamíferos (principalmente roedores e marsupiais) acontece comumente por intermédio de fezes de insetos triatomíneos contaminadas com a forma infectiva da doença (GROOM et al., 2017; NGUYEN; WASEEM, 2017; NOIREAU et al., 2009). Triatomíneos hematófagos, pertencentes, sobretudo ao gênero *Triatoma* (espécie *Triatoma infestans*, popularmente designado de “chupão” ou “barbeiro”, *Triatoma sanguisuga* e *Triatoma gerstaeckeri* – sul dos EUA, *Triatoma dimidiata* – México, América do Sul e Central), que apresenta prevalência na América do Sul, depositam suas fezes e urina na pele do hospedeiro durante ou após o repasto sanguíneo (NEVES et al., 2011; NGUYEN; WASEEM, 2017). Em outros milhares de casos, a transmissão do *T. cruzi* ocorre por meio de triatomíneos pertencentes aos gêneros *Rhodinus* e *Panstrongylus* (STANAWAY; ROTH, 2015).

O *Triatoma sp* está adaptado a áreas rurais peri-domiciliares e às condições de

baixa umidade (secas), cenário típico de países da América Central e América do Sul. Habitações de má qualidade, por sua vez, atuam como alojamento para o inseto, o que propicia uma situação favorável ao ataque ao hospedeiro (CHAUDHURI; CHAUDHURI, 2004; NOIREAU et al., 2009; TEIXEIRA et al., 2006). Acreditava-se que somente pessoas que viviam em zonas rurais estavam sujeitas à infecção pelo *Trypanosoma*, mas, devido à migração destas para áreas urbanas ou para outros países e a migração de pessoas doentes que vivem na zona urbana para países ou áreas não endêmicas, onde existe a presença do vetor, a doença de Chagas continua sendo disseminada, configurando-se como uma doença também re-emergente, em constante expansão de sua distribuição geográfica (CONNERS et al., 2017; MACKEY et al., 2014; MACKEY; LIANG, 2012; RASSI JR et al., 2012).

Após o repasto sanguíneo, as formas tripomastigotas metacíclicas misturadas às fezes ou urina do inseto vetor atingem a corrente sanguínea. A coceira provocada pela picada do inseto facilita a entrada do parasita no organismo, através do local da picada ou mesmo atravessando a membrana mucosa intacta dos olhos, nariz e boca, por exemplo, ou ainda, por meio de feridas e cortes na pele. Não obstante, também é possível que a transmissão da doença ocorra por meio de transfusão sanguínea, transplante de órgãos e de medula óssea. Transmissão vertical, via placenta, nos quais bebês nascem infectados de mães com a mesma doença, ou ainda, em poucos casos, por acidentes em laboratórios de pesquisa, através da manipulação do parasita (BERN, 2015; BUEKENS et al., 2013; HERWALDT, 2001; HOFFLIN et al., 1987; HOWARD et al., 2014; KINOSHITA-YANAGA et al., 2009).

Uma vez no organismo do hospedeiro, *T. cruzi* diferencia-se nas demais formas de vida no hospedeiro vertebrado e posteriormente no hospedeiro invertebrado, ocorrendo variações morfológicas e metabólicas, de acordo com as necessidades para adaptação a cada hospedeiro. Na Figura 2 está representado o ciclo de vida do *T. cruzi*. Tripomastigotas metacíclicas, ao serem depositadas na pele do hospedeiro vertebrado atingem a corrente sanguínea por uma das vias de entrada (1) e invadem células nucleadas (2). Tripomastigotas diferenciam-se em amastigotas intracelulares e se replicam por fissão binária (3). Após 4 ou 5 dias, as amastigotas se transformam em tripomastigotas, promovendo a ruptura das células infectadas, ficando estas, por fim, livres na corrente sanguínea (4). O ciclo tem continuidade com

a infecção de outros vetores (5), que porventura se alimentam do hospedeiro ou ainda, através da infecção de novas células, dando início a novos ciclos de replicação. Já no hospedeiro invertebrado, as tripomastigotas atingem o intestino médio, diferenciam-se em epimastigotas (6) e passam por um processo de replicação (7). Por sua vez, estas migram para o intestino grosso, onde se diferenciam novamente, dando origem às tripomastigotas metacíclicas infectantes, as quais são excretadas juntamente com as fezes do vetor, reiniciando o ciclo (8).

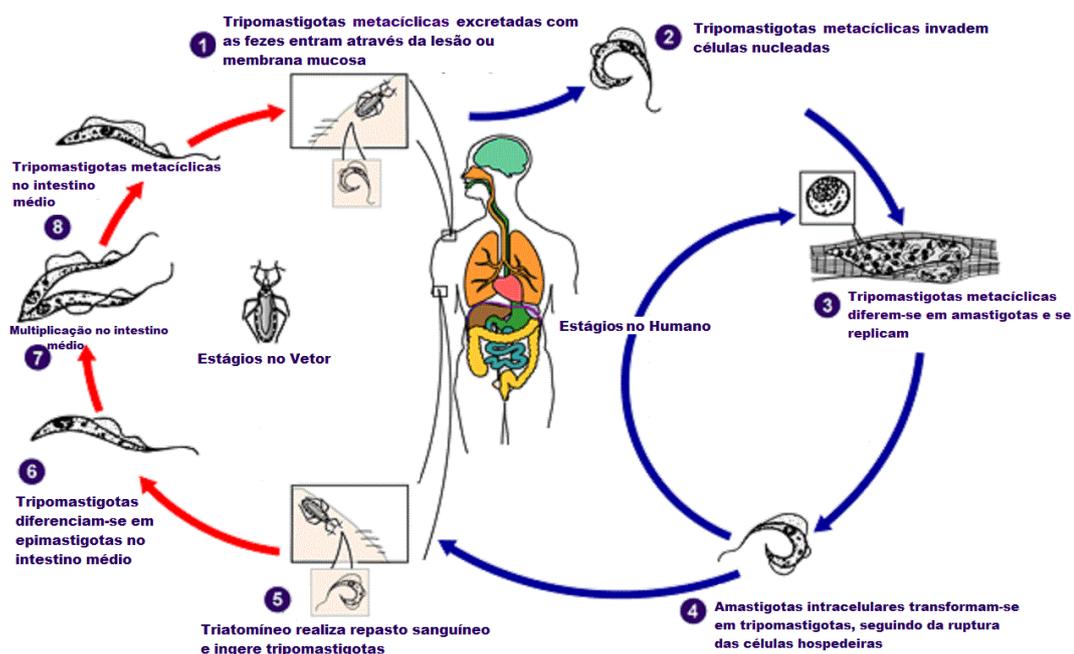


Figura 2. Ciclo de vida de *T. cruzi* no hospedeiro humano.

Fonte: *Centers for Disease Control and Prevention* - CDC (2015), adaptada pelo autor.

Atualmente, os casos de transmissão vetorial têm diminuído em muitos países, devido aos programas de controle do vetor (SHIKANAI-YASUDA; CARVALHO, 2012), ao mesmo tempo em que a transmissão oral vem ganhando destaque nas formas de transmissão da doença de Chagas, devido a surtos causados em países da América Latina (ALARCÓN DE NOYA et al., 2010; COURA, 2006; SILVA-DOS-SANTOS et al., 2017; XAVIER et al., 2014). Na Amazônia brasileira, esta é a mais importante forma de transmissão, representando 70% dos casos em um período que vai de 2000 a 2010, porém, os primeiros casos de transmissão oral datam de 1965, tendo sido

relatado um caso em Teotonia, no Rio Grande do Sul (FERREIRA et al., 2001; NÓBREGA et al., 2009; PEREIRA et al., 2009; SÁNCHEZ; RAMÍREZ, 2013; SHIKANAI-YASUDA; CARVALHO, 2012). A transmissão por via oral está associada à ingestão de alimentos e bebidas contaminadas com o vetor infectado ou o protozoário, podendo ser citados como exemplos o extrato de cana-de-açúcar, polpa de açaí, suco de goiaba, babaçu e vinho de palma (BENCHIMOL BARBOSA, 2006; COURA et al., 1997; DIAS et al., 2008; MARCILI et al., 2009b; SÁNCHEZ; RAMÍREZ, 2013; SHIKANAI-YASUDA; CARVALHO, 2012; TOSO et al., 2011). Devido à alta carga parasitária, a infecção a partir de alimentos e bebidas são em geral mais graves do que quando transmitidos por via cutânea, (EICKHOFF et al., 2013). Finalmente, a transmissão por transfusão de sangue chega a ser a segunda via de transmissão mais importante da doença de Chagas. Em países altamente industrializados, como EUA, Canadá, Espanha e Suíça, essa via é responsável pela maioria dos casos, pois ocorrem constantes movimentos migratórios de latino-americanos. (GRANT et al., 1989; JACKSON et al., 2009; JACKSON et al., 2014; KESSLER et al., 2013; PÉREZ-MOLINA et al., 2011; RODRIGUEZ-GUERINEAU et al., 2014).

1.1.2 Aspectos clínicos da doença de Chagas e tratamento

A fase aguda da doença de Chagas é caracterizada pela presença do parasita no sangue do hospedeiro, apesar de a carga parasitária ser baixa, ainda assim, é possível detectar tripomastigotas de *T. cruzi* por microscopia direta de uma amostra sanguínea ou por reação em cadeia da polimerase (PCR) em receptores de transplante de órgãos ou após exposição acidental (NGUYEN; WASEEM, 2017; RASSI JR et al., 2010; STANAWAY; ROTH, 2015). Porém, na maioria dos pacientes portadores da doença, a mesma não é percebida, pois 95% dos casos são assintomáticos ou apresentam sintomas leves, e nos 5% restantes, os sintomas incluem febre e mal-estar, que duram de 4-8 semanas (LARANJA et al., 1956; STANAWAY; ROTH, 2015; WEGNER; ROHWEDDER, 1972). Outros sintomas incluem dores musculares e articulares, diarreia, edema (local ou generalizado), distúrbios respiratórios, aumento dos gânglios linfáticos, baço e fígado, dificuldade de respirar e coma (TEIXEIRA et al., 2006; WHO, 2015b). Sinais visíveis, como por

exemplo, lesões na pele (chagoma, no local de entrada do parasita) ou o inchaço arroxeadado de uma pálpebra (sinal de Romaña) podem ser notados em menos da metade dos casos (DNDI, 2015; RASSI JR et al., 2010). Este estágio da doença é responsável por menos de 5% das mortes, que ocorrem na maioria das vezes devido a complicações como a miocardite ou meningoencefalite (TEIXEIRA et al., 2006).

A fase aguda da doença de Chagas segue para uma fase crônica intermediária, que continua a ser assintomática, soropositiva para *T. cruzi* e sem nenhum acometimento visceral, sendo este, portanto, o maior grupo em que se enquadram os portadores da doença (COURA, 1975; DIAS, 1989; PÉREZ-MOLINA; MOLINA, 2017; STANAWAY; ROTH, 2015). Esta condição assintomática pode durar de 10 a 30 anos após a infecção, facilitando a transmissão do parasita por transfusões ou transmissão vertical, devido ao desconhecimento acerca de estar ou não portando a doença (DNDI, 2015). Contudo, cerca de um terço da população evolui para um quadro sintomático da fase crônica que se caracteriza por complicações, devido ao fato de os parasitas estarem alojados de forma latente, principalmente no coração e nos músculos digestivos, levando a quadros graves de cardiopatia chagásica crônica (14% - 45%), que leva a eventos de insuficiência cardíaca, eventos de arritmias ou até mesmo morte súbita, por atingirem principalmente o sistema de condução e o miocárdio. Além disso, parasitas alojados nos músculos digestivos podem provocar alargamento do esôfago (megaesôfago) ou cólon (megacólon), em uma proporção que varia de 10% a 21% (DE OLIVEIRA et al., 1998; TEIXEIRA et al., 2006; WHO, 2015b). Alterações neurológicas também são sintomas que podem ser observados neste estágio da doença, podendo provocar a morte, anos depois de o paciente ter sido infectado, pois progressivamente o sistema nervoso é destruído (WHO, 2015b).

As manifestações clínicas agudas e crônicas mais comuns na doença de Chagas estão ilustradas na Figura 3.

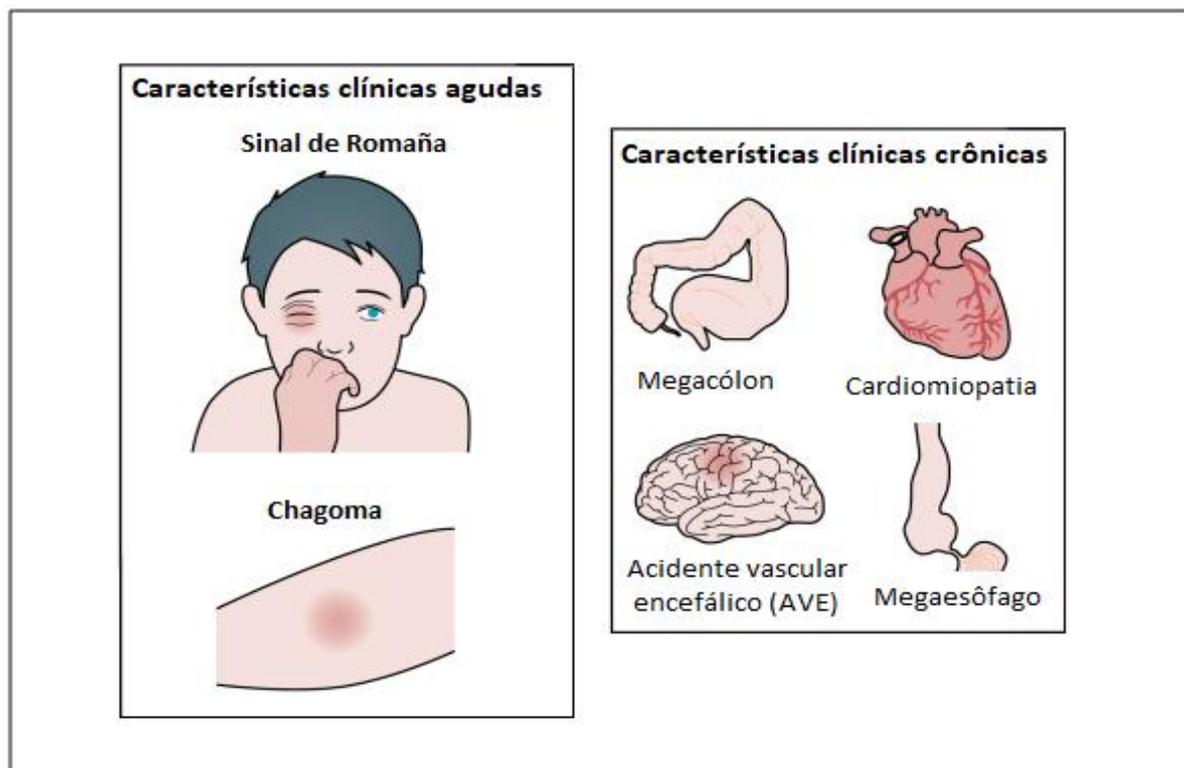


Figura 3. Manifestações clínicas da doença de Chagas. (A) Sinal de Romãña e chagoma, decorrentes da picada de insetos triatomíneos e de edema inflamatório unilateral, respectivamente, são sintomas característicos da fase aguda da doença de Chagas. (B) Megacólon e megaesôfago resultantes de eventos de dilatação, juntamente com a destruição das fibras miocárdicas (cardiomiopatia chagásica) e raramente AVE caracterizam a fase crônica da doença.

Fonte: PÉREZ-MOLINA e MOLINA, (2017), adaptada pelo autor.

Atualmente, para o tratamento contra a doença de Chagas são utilizados apenas dois fármacos: Nifurtimox e Benzonidazol. A cura parasitológica fornecida por ambos é, porém, limitada, principalmente em pacientes em estado crônico, onde resistência a estas drogas já foi relatada (BUCKNER et al., 1998; LESCURE et al., 2010; RASSI JR et al., 2010). Outro fator limitante é o fato de o tratamento medicamentoso ser recomendado apenas para pacientes que ainda estejam na fase aguda da doença, para aqueles que contraíram de maneira congênita ou para aqueles que sofreram reativação da doença, que ocorre devido ao uso de fármacos utilizados no tratamento de pacientes com HIV, como corticosteroides (ex. prednisona) e outros fármacos imunossupressoras. Crianças infectadas e adultos de até 18 anos que apresentam a doença crônica também são passíveis de tratamento (CORDOVA et al., 2008; LESCURE et al., 2010; RASSI JR et al., 2010).

Devido ao fato de ser categorizada como uma doença negligenciada, os esforços de pesquisa e desenvolvimento de novas vacinas e terapias anti-chagásicas tem-se revelado como um processo extremamente lento. Isso leva a um quadro ainda mais alarmante: não há nenhuma vacina eficaz sendo comercializada ou novos fármacos aprovados para o uso clínico (PINTO et al., 2013). Para fazer avançar as pesquisas nessa direção, são, portanto, imprescindíveis estudos moleculares sobre diferentes aspectos da biologia desse parasito, os quais têm como base os dados sobre o seu genoma.

1.2 Estudos sobre o genoma do *Trypanosoma cruzi*

O *T. cruzi* é um protozoário eucariótico, unicelular e flagelado pertencente a ordem Kinetoplastida, contendo DNA mitocondrial (kDNA) e DNA nuclear. Assim como os demais membros da família Tripanossomatidae, seu genoma nuclear é organizado em longos *clusters* de genes poliscistrônicos (PGCs), ou seja, dezenas a centenas de genes codificadores de proteínas presentes na mesma fita de DNA são transcritos em uma mesma molécula de pré-mRNA. A ausência de íntrons também é uma característica marcante da organização do genoma nuclear desses parasitas, tendo sido descritos eventos de *cis-splicing* apenas para o gene que codifica a polimerase poly(A) (MAIR et al., 2000). Como consequência dessa organização, os genes são densamente agrupados, separados apenas por pequenas regiões intergênicas ricas em sequências de polipirimidinas, as quais atuam como sinalizadores para os dois processos importantes que ocorrem no processamento do mRNA: *trans-splicing* e poliadenilação (LIANG et al., 2003; ZWIERZYNSKI et al., 1989), os quais serão descritos em detalhes na próxima sessão.

O conhecimento adquirido acerca da estrutura do genoma teve início em 1994, com o Projeto Genoma do *T. cruzi* (DEGRAVE et al., 1997). Todas as análises realizadas neste projeto foram baseadas no clone CL Brener (BRENER; CHIARI, 1963). O fato de *T. cruzi* ser constituído de uma população extremamente heterogênea, apresentando alta diversidade genética e fenotípica intraespecífica culminou na classificação deste parasita em seis DTUs (do inglês - *Discrete Typing Units*) denominadas TcI ao TcVI que podem ser identificados por marcadores moleculares ou imunológicos comuns (MARCILI et al., 2009a; TIBAYRENC, 1998;

ZINGALES et al., 2009). Além disso, *T. cruzi* apresenta propriedades biológicas e características moleculares distintas, o que reforçou a necessidade de escolha de um clone para servir como modelo de estudo (DVORAK, 1984; SOUTO; ZINGALES, 1993; TIBAYRENC; AYALA, 1988; ZINGALES; COLLI, 1985).

O clone CL Brener, derivado da cepa CL e pertencente a DTU VI na classificação de *T. cruzi* (ZINGALES et al., 2009) possui algumas características biológicas que foram decisivas para que esta fosse escolhida como modelo: ter sido isolado a partir de *Triatoma infestans* (BRENER; CHIARI, 1963); ser infectivo para camundongos (BRENER, 1977); possuir tropismo por células cardíacas e musculares (MELO; BRENER, 1978); mostrar uma fase aguda clara em humanos infectados acidentalmente; e possuir baixa resistência às drogas usadas clinicamente na doença de Chagas (FILARDI; BRENER, 1987; ZINGALES et al., 1997) .

Laboratórios pertencentes à rede do Projeto Genoma desenvolveram vários trabalhos que foram importantes para o início da compreensão a respeito do genoma de *T. cruzi*. Foram identificadas *tags* de sequências transcritas (ESTs) e sítios de sequências com *tags* (STs) a partir de cromossomos isolados (BRANDÃO et al., 1997; CERQUEIRA et al., 2005), análise do cariótipo nuclear do parasita (CANO et al., 1995; HENRIKSSON et al., 1990; HENRIKSSON et al., 1996; HERNANDEZ-RIVAS; SCHERF, 1997; SANTOS et al., 1997) e a construção de mapas genéticos e físicos (CERQUEIRA et al., 2005; KOLEV et al., 2010; PORCEL et al., 2000). Contudo, a elucidação da estrutura completa do genoma de *T. cruzi* somente foi possível após a publicação da montagem de seu genoma, em 2005 (EL-SAYED et al., 2005a), concomitantemente à publicação dos genomas de *Trypanosoma brucei* (*T. brucei*) e *Leishmania major* (*L. major*). O mesmo foi sequenciado pela primeira vez através da técnica *whole-genome shotgun* (WGS), com uma cobertura de 14 vezes, o que revelou a natureza híbrida diplóide do clone CL Brener, possivelmente originado a partir de um evento de fusão entre genomas de um membro de DTU II e um membro de DTU III. Foram obtidos um total de 8740 *contigs* que possibilitaram a construção parcial de 5489 *scaffolds*, totalizando 67 Mb (EL-SAYED et al., 2005a). Comparações pós-montagem de *reads* obtidas da cepa Esmeraldo (DTU II) com os *contigs* do clone CL Brener possibilitaram a distinção dos dois haplótipos, que mostraram ter 30.5 Mb compartilhados entre eles. Além disso, os dois haplótipos apresentaram um alto nível

de sintenia e conservação, principalmente para genes *housekeeping*, que chegam a ter 97.8% de similaridade (EL-SAYED et al., 2005a).

Considerando o genoma haplóide de *T. cruzi* foram identificados cerca de 12.000 genes. Um total de 22.570 genes codificadores de proteínas foi também descrito para os dois haplótipos, em que 6.159 correspondem ao haplótipo II e 6.043 ao outro haplótipo. Outras 10.368 sequências analisadas não puderam ser atribuídas a nenhum dos dois haplótipos. Outro dado relevante é o alto conteúdo repetitivo, que supera 50% do genoma e engloba grandes famílias gênicas, as quais codificam para proteínas de superfície (MASPs foram identificadas como sendo uma família multigênica que apresenta 1300 cópias; trans-sialidases (TS), mucinas e proteases gp63), retrotransposons e repetições subteloméricas, responsáveis pelo polimorfismo de tamanho em cromossomos homólogos (EL-SAYED et al., 2005a; HENRIKSSON et al., 1996).

O resultado final da montagem não possibilitou uma anotação completa ou mesmo com elevada acurácia de todo o genoma. Esse fato é devido às regiões com grande conteúdo repetitivo, em especial regiões relacionadas à famílias multigênicas codificadoras de proteínas de superfície, o que impossibilita uma montagem precisa para essas regiões. Por isso, outros trabalhos foram desenvolvidos a fim de melhorar o que havia sido feito e ampliar a compreensão acerca do genoma de *T. cruzi* e de outras espécies de tripanossomatídeos, com as quais apresenta um alto grau de sintenia, como o genoma de *T. brucei* e *Leishmania major* (EL-SAYED et al., 2005b).

WEATHERLY et al. (2009) utilizou *contigs* e *scaffolds* de *T. cruzi*, previamente montados, e com base em mapas de sintenia gerados com a espécie de *T. brucei* e em extremidades de sequências de *Bacterial Artificial Chromosome* (BAC), que possuíam sequências presentes em *T. cruzi*, foi possível obter cromossomos completos e determinar a homologia entre os dois haplótipos para cada cromossomo de *T. cruzi*. Foram montados um total de 41 pares de cromossomos, com tamanho variando de 78kb a 2,4Mb, os quais representavam 90% (21133 de 23216 genes) dos genes anotados (famílias multigênicas foram retiradas das análises), resultado este que divergiu do que havia sido previamente obtido por análise de eletroforese de gradiente de campo pulsado (PFGE), que estimou um total de 55 cromossomos para o clone CL Brener (BRANCHE et al., 2006). Resultados de outro estudo corroboram

com os tamanhos aproximados obtidos para cada cromossomo, os quais permanecem estáveis ao longo de todas as fases do ciclo celular e durante a diferenciação (ENGMAN et al., 1987; GALINDO et al., 2007).

1.3 Transcrição em Tripanossomatídeos e regulação da expressão gênica

O processo pelo qual se obtém moléculas de RNA a partir de uma molécula de DNA molde é chamado de transcrição, onde nos tripanossomatídeos ocorre de forma policistrônica e não monocistrônica, como ocorre para a maioria dos organismos eucariotos (PALENCHAR; BELLOFATTO, 2006). Os *clusters* de genes que codificam proteínas funcionalmente não relacionadas são co-transcritos por RNAPII, produzindo pré-mRNAs (transcritos policistrônicos), de forma convergente ou divergente, ou seja, são transcritos em direção aos telômeros ou partindo deles (MARTÍNEZ-CALVILLO et al., 2003, 2004; VANHAMME; PAYS, 1995). Com exceção do gene que codifica a molécula de *spliced-leader* (SL), nenhum promotor foi identificado como iniciador da transcrição e poucos fatores de transcrição foram descritos até então (CRIBB et al., 2010; CRIBB; SERRA, 2009).

mRNAs maduros são gerados a partir de unidades policistrônicas que são processadas em unidades monocistrônicas por duas reações de clivagem simultâneas. *Trans-splicing* adiciona uma sequência de 39-nt (mini-éxon ou SL) capeada (7-metilguanossina), guiado por um dinucleotídeo AG *upsteam*, na extremidade 5' de praticamente todos os mRNAs. Já na extremidade 3' é adicionada uma sequência poli(A), por um processo denominado poliadenilação (HUMMEL et al., 2000; PARSONS et al., 1984; SIEGEL et al., 2005; TEIXEIRA; DAROCHA, 2003; ZWIERZYNSKI et al., 1989). Sequências ricas em pirimidina localizadas entre duas regiões codificadoras adjacentes direcionam ambos os processos e rege o processamento de RNA co-transcricional, uma vez que a seleção de um sítio de *trans-splicing* para um gene influencia a escolha de um sítio de poliadenilação para o gene *upstream* (LEBOWITZ et al., 1993; LIANG et al., 2003).

Os processos de maturação do mRNA determinam o comprimento das regiões 3' e 5' UTRs (do inglês, *untranslated regions* - UTRs). Através do mapeamento de ESTs e cDNAs contendo sequências de SL e/ou cauda poli(A) em regiões intergênicas

constatou-se que o tamanho médio de regiões 5' UTR é de 35 nt e de regiões 3' UTR é de 264 nt. Da mesma forma, a distância média entre o trato de poli(PY) e o sítio acceptor de SL é de 18 nt, já para o sítio de adição da cauda poli(A) e o motivo poli(PY) mais próximo a distância média é de 41 nt (CAMPOS et al., 2008). Além disso, a sequência SL é adicionada apenas 30-100 nucleotídeos *upstream* do codón ATG de iniciação da tradução, sendo assim os mRNA de *T. cruzi* contêm 5' UTR curtos, comparadas com a região 3' UTR, que pode chegar a ter mais de 2000 nucleotídeos de comprimento, reflexo da grande quantidade de sítios de poliadenilação existentes para cada mRNA (BRANDÃO; JIANG, 2009; SIEGEL et al., 2010).

Outra consequência da transcrição policistrônica é a produção de mRNA em níveis iguais para todos os genes que pertencem ao mesmo *cluster*. Porém, devido as constantes mudanças que o *T. cruzi* sofre ao longo de seu ciclo de vida, as necessidades metabólicas variam e por esse motivo, o parasita desenvolveu mecanismos que regulam a expressão gênica estágio-específica. Os principais mecanismos regulatórios ocorrem após a transcrição, podendo influenciar o processamento, a estabilidade e a capacidade de tradução dos mRNAs (HAILE; PAPADOPOULOU, 2007).

Diferentes trabalhos à respeito do controle pós-transcricional da expressão gênica começaram a surgir a partir da década de 80, mostrando que eventos de *trans-splicing* e *splicing* alternativo exercem papéis determinantes na abundância de determinado gene ao longo do ciclo de vida de tripanossomatídeos (GUPTA et al., 2013; TEIXEIRA; et al., 1999). NILSSON et al., (2010) analisaram padrões de *splicing* em *T. brucei* de formas sanguíneas e procíclicas, identificando mais de 2500 eventos de *splicing* alternativo (mais de 85% das transcrições possuem variantes), alguns sendo regulados de forma estágio-específica. Este resultado corrobora os dados apresentados por KOLEV et al. (2010), onde apenas 11% do processamento de pré-mRNAs são direcionados por um único sítio de *trans-splicing*. Em *T. cruzi* o gene LYT1 sofre a ação do mesmo mecanismo, o que resulta em dois produtos de proteínas diferentes e um terceiro mRNA que codifica uma proteína truncada (BENABDELLAH et al., 2007; MANNING-CELA et al., 2002).

Os níveis de equilíbrio de diferentes mRNAs são controlados por taxas de decaimento específicas e dependem basicamente de três fatores: sinais *cis splicing*,

fatores que atuam em *trans* e mecanismos que interferem no *turnover* ou degradação de mRNAs (CLAYTON; SHAPIRA, 2007). Brandão e Jiang (2009) mostraram que existem sequências repetitivas (motivos) distribuídas ao longo de regiões 5' e 3' UTR, nesta última em maior proporção (90%). Esses motivos de sequência (elementos *cis*) desempenham um importante papel na regulação da expressão de vários genes independentes, atuando como sítios de ligação de proteínas (elementos *trans*), as quais podem atuar em mRNAs alvos, estabilizando-os ou desestabilizando-os, influenciando na meia vida dos mesmos ou ainda contribuindo para sua degradação ou melhora na eficiência do processo de tradução (DE GAUDENZI et al., 2003; HAILE; PAPADOPOULOU, 2007; HENDRIKS; MATTHEWS, 2007; NOZAKI; CROSS, 1995; WU et al., 2006).

A identificação e análise de motivos presentes em regiões 3' UTR da maioria dos mRNAs mostrou que os mesmos são conservados quando relacionados a um mesmo ligante. BARTHOLOMEU et al. (2002), ao deletar 369 nt *downstream* do gene que codifica para tubulina (sequência rica em elementos contendo poly(G) e poli(T)) observou que a expressão de mRNAs para este gene aumentou de 3 a 6 vezes em epimastigotas. JACKSON (2010) realizou a análise da diversidade estrutural de regiões 3' UTR de genes de *L. infantum*, *L. braziliensis*, *L. seymouri* e *Crithidia* sp, o que indicou a redundância dessas sequências e também diferenciação funcional, uma vez que apresentaram identidade > 40% para mais de 300 pb. Posteriormente foram identificados motivos ricos em uracila (URE), com comprimento de 100 nt, por meio da expressão de genes repórteres de *T. brucei*, os quais contribuem para a estabilidade dos respectivos transcritos nas formas procíclicas e instabilidade dos mesmo em formas sanguíneas (HAILE; PAPADOPOULOU, 2007). Em *T. cruzi* o mesmo evento foi observado para diferentes transcritos, fato que regula a meia-vida destes ao interagirem principalmente com proteínas de ligação ao RNA (RBPs), reforçando a importância da modulação da estabilidade de mRNAs (D'ORSO et al., 2003).

O papel de regiões 3' UTR e a identificação de elementos regulatórios continuam sendo alvos de diversos estudos. A identificação de motivos em mRNAs da glicoproteína GP82 são responsáveis pela regulação estágio-específica dessa molécula em tripomastigotas metacíclicas (BAYER-SANTOS et al., 2012). Outro

estudo, baseado na transfecção de epimastigotas e amastigotas com vetores contendo genes reporteres da luciferase flanqueados por sequências 3' UTR e região intergênica dos genes de alfa-tubulina e amastina (aumentados em epimastigota e amastigota, respectivamente) mostraram o papel dessas regiões na regulação estágio-específica de ambos os genes analisados. Aumento na atividade da luciferase foi detectado em epimastigotas e amastigotas transfectadas com o vetor contendo sequências de alfa-tubulina e amastina, respectivamente (ARAÚJO et al., 2011).

1.4 Proteínas de ligação ao RNA como reguladores da expressão gênica em *T. cruzi*

No mesmo ano da publicação do genoma de *T. cruzi*, um trabalho referente ao proteoma do parasita também foi publicado (ATWOOD et al, 2005). As quatro formas de vida tiveram seu proteoma analisado pela técnica de espectrometria de massas, o que resultou na identificação de 2748 proteínas aproximadamente, das quais 1168 correspondem à grupos de proteínas que já haviam sido anotadas. Curiosamente, várias proteínas identificadas apresentavam expressão diferencial entre as formas de vida de *T. cruzi*, sendo 248 presentes apenas em umas das quatro formas de vida. Em 2010, ALVES et al. caracterizou a formação de complexos de proteínas e mRNAs (mRNPs) em epimastigotas do clone virulento Dm28 (pertencente ao DTU I) de *T. cruzi*, utilizando também a técnica de espectrometria de massa. Esses estudos, juntamente com a montagem do genoma impulsionaram novos trabalhos que objetivam caracterizar diferentes proteínas, como por exemplo, aquelas envolvidas na regulação gênica.

As proteínas de ligação a RNA, ou RBPs, atuam como um dos principais elementos de regulação pós-transcricional. Ao se ligarem à mRNAs formando um complexo ribonucleoproteico ou mesmo à outras proteínas, cuja ação final resulte no controle da expressão, as RBPs estabilizam ou desestabilizam seus alvos, em respostas a diferentes sinais (ALVES; GOLDENBERG, 2016). Dessa forma, RBPs podem desempenhar um papel chave em processos de *splicing* (DE GAUDENZI et al., 2016a), transporte de mRNA, direcionando sua localização na célula (KISHORE et al., 2010), modulação do processo de tradução (EDWARDS, 2015) e decaimento

dos níveis de mRNA (RUIZ-ECHEVARRÍA; PELTZ, 2000).

Com o avanço de ferramentas de bioinformática e de técnicas de alto desempenho, principalmente proteômicas, foi possível identificar todas as RBPs e caracterizar várias delas, assim como sequências envolvidas no processamento de mRNAs (BELEW et al., 2017; BENZ et al., 2005; CAMPOS et al., 2008; DALLAGIOVANNA et al., 2005; DUHAGON et al., 2001; SMITH et al., 2008). Nesse sentido, estruturas modulares foram preditas para diferentes RBPs, mostrando que as mesmas são compostas por repetições de vários domínios básicos que se combinam de diferentes formas de modo a estabelecer a função desempenhada pela respectiva RBP (LUNDE et al., 2007). Os diferentes arranjos conferem afinidade e especificidade para uma grande diversidade de mRNAs alvos e medeiam interações proteína-proteína, revelando a alta versatilidade funcional desse grupo de proteínas (ARAÚJO; TEIXEIRA, 2011). Mas, o conhecimento acerca do papel das RBPs no controle dos níveis de expressão gênica ainda é limitado, havendo a necessidade de mais estudos para elucidar o completo mecanismo de ação das mesmas, bem como a caracterização daquelas as quais nenhum estudo foi realizado (KRAMER; CARRINGTON, 2011).

Os principais domínios identificados nas RBPs são: RRM (do inglês *RNA recognition motif*), PUF, dedo de zinco CCCH, Alba, KH, dsRBD, PAZ, PIWI e domínio S1 (descritos por LUNDE et al., 2007). O primeiro é o mais abundante e melhor caracterizado em RBPs (AFROZ et al., 2015), tendo sido encontrado em todas as fases de vida de *T. cruzi* (MARIS et al., 2005). Compreende aproximadamente 80-90 aminoácidos, cuja sequência é evolutivamente conservada, caracterizando dois motivos, denominados RNP1 e RNP2, e consiste em quatro cadeias- β antiparalelas, que formam uma folha β que é empacotada contra duas hélices α , adotando a conformação típica de $\beta 1\alpha 1\beta 2\beta 3\alpha 2\beta 4$ (OUBRIDGE et al., 1994). Além disso, dois ou mais RRM's podem ser combinados na mesma molécula para reconhecer elementos *cis* presentes nas UTRs de mRNAs (2 a 8 nucleotídeos), com maior afinidade e especificidade de sequência, atuando desde a biogênese do mRNA até o processo de *splicing* alternativo (CLÉRY; ALLAIN, 2012; revisado por PÉREZ-CAÑADILLAS; VARANI, 2001). Como exemplo de RBPs que contém este domínio pode ser citada a TcRBP19, uma proteína de 17-kDa que se liga preferencialmente à regiões poli(C),

sendo expressa a partir de um gene de cópia única apenas em baixos níveis, em amastigotas (PÉREZ-DÍAZ et al., 2007). Experimentos *in vitro* de superexpressão da TcRBP19 resultaram na redução da taxa de diferenciação durante a metaciclogênese e conseqüentemente na diminuição de células VERO (células epiteliais provenientes de rins de Macaco Verde Africano) e HeLa (tipo de célula imortal obtida a partir de um câncer cervical) infectadas com amastigotas do clone Dm28 de *T. cruzi* (PÉREZ-DÍAZ et al., 2012).

Outra RBP identificada, que apresenta o domínio RRM é a TcRBP28, uma proteína que contém motivos repetitivos ricos em Ala-Lys-Pro na sua porção N-terminal e que é reconhecida por anticorpos presentes no soro de pacientes com doença de Chagas (PAIS et al., 2008). Apesar de não ter função conhecida, sabe-se que a TcRBP28 é homóloga a duas RBPs já caracterizadas em *T. brucei*: *Tbp34* e *Tbp37*, as quais estão envolvidas em vias de exportação nuclear (GOMES et al., 2004; HELLMAN et al., 2007).

A família das TcUBPs consiste de pelo menos seis membros (TcUBP1-TcUBP6), identificados na base de dados de Trypanossomas do Centro Sanger, que apresentam um único domínio RRM comum (conservado ao nível dos motivos RNP1 e RNP2), o qual pode estabelecer relações estruturais com domínios auxiliares, permitindo que estas proteínas participem de interações com outras proteínas (DE GAUDENZI et al., 2003). Foi visto também, que possuem um padrão de expressão diferencial ao longo do ciclo de vida de *T. cruzi* e variação na preferência por motivos presentes na 3' UTR de mRNAs, contudo existe uma tendência de ligação à motivos *cis* poli(U) e poli(G). Análises de *Northern blot* mostrou que TcUBP1 e TcUBP6 são mais expressas em amastigotas e tripomastigotas, ao contrário de TcUBP2, TcUBP3 e TcUBP4 que são mais expressas em epimastigotas comparado aos outros estágios de vida. Já TcUBP5 apresenta-se diferencialmente expressa apenas em tripomastigotas (DE GAUDENZI et al., 2003).

Estudos individuais para UBP1 e UBP2 mostraram que estas fazem parte de um mesmo complexo ribonucleoproteico (mRNA), que se liga à motivos ricos em uracila (D'ORSO; FRASCH, 2002). Outro resultado mostra que TcUBP1 se liga a uma sequência de 44 nucleotídeos rica em AU na 3' UTR de mRNAs de TcSMUG (gene que codifica um membro da família das mucinas) e em sequências ricas em GU,

causando um efeito de desestabilização do alvo (D'ORSO; FRASCH, 2001). Contudo, o transcrito dicistrônico que codifica para estas duas proteínas tem seu processamento influenciado pela TcDRBD4/PTB2, uma proteína homóloga ao regulador de *splicing* humano, que se liga ao trato de polipirimidina, diminuindo a expressão de ambas, por diminuir a taxa de *trans-splicing* (DE GAUDENZI et al., 2016b).

Outra RBP que apresenta domínio RRM é a TcPABP1, que se liga a cauda poli(A) de mRNA alvos, caracterizada pela primeira vez em 1994. Análises de *Northern* e *Southern blot*, respectivamente, mostraram que tanto o mRNA, quanto a proteína são constitutivamente expressos durante todos os estágios de vida de *T. cruzi* (BATISTA et al., 1994). PABP1 e PABP2, proteínas homólogas em *T. brucei* foram localizadas em diferentes grânulos de ribonucleoproteínas (RNPs) que são semelhantes aos corpos de processamento (*p-bodies*). Esse resultado indica uma função associada ao armazenamento e decaimento dos níveis de mRNAs alvos disponíveis para a tradução, em resposta a inibição da tradução ou do *trans-splicing* (KRAMER et al., 2013).

Anos depois proteínas nomeadas TcRBP40 e TcRBP19, do grupo das RBPs que possuem domínio RRM, foram caracterizadas. RBP40 é uma proteína de 14,3 kDa para a qual 148 genes-alvos foram identificados. Essas proteínas apresentam um único domínio RRM que é capaz de reconhecer elementos ricos em AG na 3' UTR de mRNA. Variações na localização dessa proteína nos diferentes estágios de vida (em epimastigotas foram encontradas em reservossomos – atuam como reservatórios de macromoléculas -, e em amastigotas e tripomastigotas dispersas no citoplasma) sugerem que essa proteína possui importante função regulatória. Diferenças em sua expressão também foram verificadas durante a metaciclogênese, não tendo sido detectada em tripomastigotas metacíclicas (GUERRA-SLOMPO et al., 2012). TcRBP19, uma proteína de 17 kDa, também é diferencialmente expressa durante o ciclo de vida de *T. cruzi*, tendo sido detectada em altos níveis em amastigotas, quando comparado com epimastigotas (PÉREZ-DÍAZ et al., 2007, 2012, 2013).

Um total de 10 proteínas (TcPUF1-TcPUF6) com domínio PUF (Pumilio) foram identificadas em *T. cruzi* e separadas em três grupos de acordo com sua especificidade (CARO et al., 2006). Formado por oito repetições imperfeitas, sendo

que cada repetição se dobra em três hélices α , localizadas internamente em uma estrutura em forma de arco-íris, TcPUFs apresentam alta afinidade por seus alvos (CARO et al., 2006; WANG et al., 2002). Esse domínio está envolvido com a estabilidade e a eficácia de processos de tradução de vários mRNAs. PUF6, homóloga a PUF1 de *T. brucei*, foi identificada juntamente com nove mRNAs que são regulados pela mesma, através de ensaios de purificação e *microarray*. Ao se ligaram a complexos de degradação, PUF6 provoca uma diminuição dos níveis de expressão de seus alvos, o que foi observado nas formas epimastigotas, mas quando atuam sozinhos promovem regulação positiva dos transcritos alvos (DALLAGIOVANNA et al., 2008). Apresenta especificidade de ligação por sequências de 32 nucleotídeos similares aos *nanos response elements* (NREs), localizadas na região 3' UTR dos transcritos de *hunckback* materno (hb) (CARO et al., 2006).

O domínio dedo de zinco CCCH, encontrado em várias proteínas de tripanossomatídeos está envolvido em vários aspectos do controle celular, homeostase e diferenciação. TcZFP1 e TcZFP2, homólogas a TbZFP1 e TbZFP2 de *T. brucei*, foram descritas em *T. cruzi* por possuir um domínio do tipo C2H2 (HENDRIKS et al., 2001; MÖRKING et al., 2004). TcZFP1 possui afinidade por ribopolímeros ricos em citosina presentes na 3' UTR de transcritos alvos, diferentemente de TcZFP2 que mostrou preferência por sequências poli(A) (MÖRKING et al., 2012). TcZFP1 é diferencialmente expressa durante o ciclo de vida do parasita, tendo maior quantidade de mRNAs associados à ribossomos durante a metaciclogênese, indicando um papel fundamental na diferenciação do parasita em condições de estresse nutricional (MÖRKING et al., 2004). A mesma importância no papel da diferenciação é atribuída a TcZFP2, onde mRNAs associados a esta proteína têm regulação positiva em tripomastigotas metacíclicas, quando há baixa expressão da proteína (MÖRKING et al., 2012).

A proteína homóloga em *T. brucei*, TbZFP1 foi descrita por ser enriquecida durante a diferenciação de formas sanguíneas para formas procíclicas, permanecendo aumentada na forma procíclica, no inseto. Este foi o primeiro trabalho em *T. brucei* que por meio de nocaute caracterizou uma proteína envolvida no desenvolvimento do parasita. Através da interrupção dos dois alelos que expressam a proteína ZFP1 em formas sanguíneas obtiveram-se parasitas que não expressavam

mRNAs correspondentes, o que permitiu verificar que na ausência desta proteína, a diferenciação é comprometida (HENDRIKS; MATTHEWS, 2005).

Outra RBP contendo domínios dedo de zinco caracterizada é a TcZC3H39, a qual sequestra mRNAs altamente expressos e seus ribossomos associados em parasitos submetidos à metaciclogênese. Devido ao seu mecanismo de ação a TcZC3H39 é considerada como parte de complexos ribonucleoproteicos, que envolvem mRNAs, ribossomos, fatores de tradução e RBPs, o que tem como consequência o retardo da tradução para os mRNAs associados. Foi observado que esses complexos são formados em maior grau durante a metaciclogênese, sugerindo também um papel na regulação da diferenciação de epimastigotas de *T. cruzi*. Esse resultado é reforçado pela natureza dos transcritos regulados, que em sua maioria codificam para proteínas ribossômicas e proteínas do complexo citocromo c (COX), que são altamente expressos em tripomastigotas metacíclicos (ALVES et al., 2014).

Recentemente, pela primeira vez em *T. cruzi* uma RBP nomeada TcZC3H31 foi caracterizada por meio de nocaute do gene codificador da mesma. Resultados de *western blot* mostraram que esta é uma proteína dedo de zinco exclusivamente expressa em epimastigotas e tripomastigotas metacíclicas, sendo a única RBP deste grupo a possuir três domínios do tipo dedo de zinco. Epimastigotas nocautes para TcZC3H31 mostraram-se incapazes de se diferenciar *in vitro* e *in vivo* em tripomastigotas metacíclicas sob condições de estresse nutricional. Efeito contrário foi observado com a superexpressão da TcZC3H31 na qual constatou-se aumento da diferenciação de epimastigotas em formas infectivas. Apesar desse grande efeito sobre a metaciclogênese, quase nenhuma diferença foi observada no crescimento de parasitos nocautes comparadas com culturas WT (ALCANTARA et al., 2018).

Em contraste ao *T. brucei*, o efeito da deleção de genes de RBPs tem sido pouco estudado em *T. cruzi*. Em *T. brucei*, esses estudos tiveram enormes avanços principalmente devido ao uso da técnica de silenciamento de genes que codificam para estas proteínas, por meio de RNAi (RNA de interferência). Neste trabalho, será abordado o uso da técnica de nocaute para caracterizar uma RBP em *T. cruzi*.

1.5 Estudos do transcriptoma nas diferentes fases do ciclo de vida do *T. cruzi*

Outros estudos do perfil de expressão gênica realizados nas diferentes fases de vida do *T. cruzi* mostraram que vários genes, além daqueles que codificam para as RBPs, apresentam expressão diferencial. Tecnologias de *microarray* de cDNA, RNA-seq e ferramentas de bioinformática, associadas a análises estatísticas permitiram expandir o conhecimento acerca da mudança de expressão gênica no parasita.

Experimentos de *microarray* a partir da hibridização de cDNA de epimastigotas, tripomastogotas metacíclicas, tripomastigotas derivadas de cultura de tecidos, e amastigotas axênicas possibilitou a análise do transcriptoma, em condições de equilíbrio, para uma cepa de *T. cruzi* brasileira. Os resultados mostraram que mais de 50% (4.992) dos transcritos analisados são regulados em pelo menos uma das formas de vida, independente da inclusão de genes que codificam para famílias multigênicas, corroborando os resultados de análises proteômicas (ATWOOD, 2005; MINNING et al., 2009).

Com o avanço das técnicas de sequenciamento, que até meados de 2005 era exclusivamente realizado pelo método automatizado de Sanger, outras técnicas para o estudo do transcriptoma também foram desenvolvidas, como por exemplo, a técnica de RNA-seq (sequenciamento de RNA). Esta técnica permite quantificar, descobrir e caracterizar RNAs codificantes e não codificantes, com custo menor e maior sensibilidade na detecção da expressão de genes em comparação a técnica de *microarray*. O produto final gerado por essa técnica, denominado *reads*, é alinhado a um genoma de referência ou transcritos de referência, ou montadas *de novo* sem a sequência genômica para produzir um mapa de transcrição, que consiste tanto na estrutura transcricional quanto no nível de expressão de cada gene (WANG et al., 2009).

Através dessa técnica foi analisado o perfil de expressão gênica global de epimastigotas, células HFF infectadas com tripomastigotas e amastigotas intracelulares da cepa Y. Os resultados mostram que em situações de equilíbrio as diferenças de abundância de aproximadamente 2000 a 3500 transcritos é duas vezes maior, ou ainda, que a abundância de 15-30% dos transcritos codificadores de proteínas sofre alterações significativas para as formas de vida analisadas, divergindo

da estimativa encontrada por MINNING et al., (2009) (LI et al., 2016). De maneira similar, análise de RNA-seq revelou os processos biológicos de maior relevância em cada uma das três formas de vida da cepa Dm28c. Epimastigotes apresentaram regulação positiva para genes relacionados ao metabolismo energético, cadeia respiratória, fosforilação oxidativa, e para os genes relacionados à biossíntese de proteínas e esteróides. Genes que codificam para proteínas de superfície também aparecem por estarem aumentados em tripomastigotas (499 de 560 genes) (BERNÁ et al., 2017).

Epimastigotas da cepa Dm28c (do grupo TcI) tratadas com hidroxiuréia permitiram a análise do transcriptoma ao longo do ciclo celular (fases G1, S, G2 e mitose final). Os resultados mostraram mais uma vez a existência de genes diferencialmente expressos (GDE) nessa forma de vida, sendo 305 (9%) regulados exclusivamente pelo ciclo celular, com expressão aumentada 1,5 vezes, sobretudo, genes relacionados ao controle do ciclo celular (SANTIAGO et al., 2017).

Recentemente o nosso grupo descreveu uma análise do transcriptoma em diferentes fases do ciclo de vida de dois clones de *T. cruzi*, o clone CL Brener e o clone CL-14. A análise de RNA-seq de tripomastigotas, amastigotas provenientes de infecção de células HFF 60 e 96 horas pós-infecção e de tripomastigotas de cultura de tecidos possibilitou a comparação dos perfis de expressão gênica nessas condições, nos dois clones *T. cruzi* que apresentam distintos graus de virulência (BELEW et al., 2017). Os dados mostraram que existem grandes mudanças de expressão quando comparadas principalmente tripomastigotas de ambos os clones e amastigotas 60 horas pós-infecção. Por outro lado, comparações entre amastigotas 96 horas e tripomastigotas não revelaram grande remodelação do transcriptoma em CL Brener. Diferentemente, para o clone CL 14, essa mesma comparação revelou mudanças consideráveis na expressão, sobretudo de genes relacionados a famílias multigênicas que codificam para proteínas de superfície (BELEW et al., 2017). Esses resultados foram de fundamental importância para a compreensão da diferença de infecciosidade entre os clones, reforçando a hipótese de que a expressão reduzida de genes que codificam proteínas como MASPs, trans-sialidases e mucinas são responsáveis, em parte, pelo fenótipo avirulento do clone CL-14 (BELEW et al., 2017).

As análises de BELEW et al. (2017) revelaram ainda que o número de genes potencialmente codificadores de RBPs no clone CL Brener era maior do que o previsto, com base em estudos anteriores (CARO et al., 2006; DE GAUDENZI et al, 2005; GUERRA-SLOMPO et al., 2012; KRAMER et al., 2010): a busca de sequências contendo domínios RRM, dedo de zinco, PUF e Alba resultou na identificação de 253 sequências que correspondem a 147 genes codificadores de RBPs, 13 genes a mais em relação aos genes já identificados (BELEW et al., 2017).

Posteriormente, foram incorporados nessas análises, dados do transcriptoma de epimastigotas do clone CL Brener. Análises visando identificar RBPs diferencialmente expressas em todos os estágios do ciclo de vida do *T. cruzi* (clone CL Brener) mostraram a existência de 11 RBPs cujos níveis de mRNA encontram-se significativamente aumentados em epimastigotas (5 RBPs), amastigotas (5 RBPs) e tripomastigotas (1 RBP) (BELEW et al., 2017; VALENTE et al., manuscrito em preparação). Com base nesses dados foi escolhido o gene TcCLB.506739.99 para caracterização, uma vez que este mostrou maior diferença de expressão em comparação com amastigotas 60 horas pós infecção de células HFF, e tripomastigotas liberadas em cultura (Figura 4).

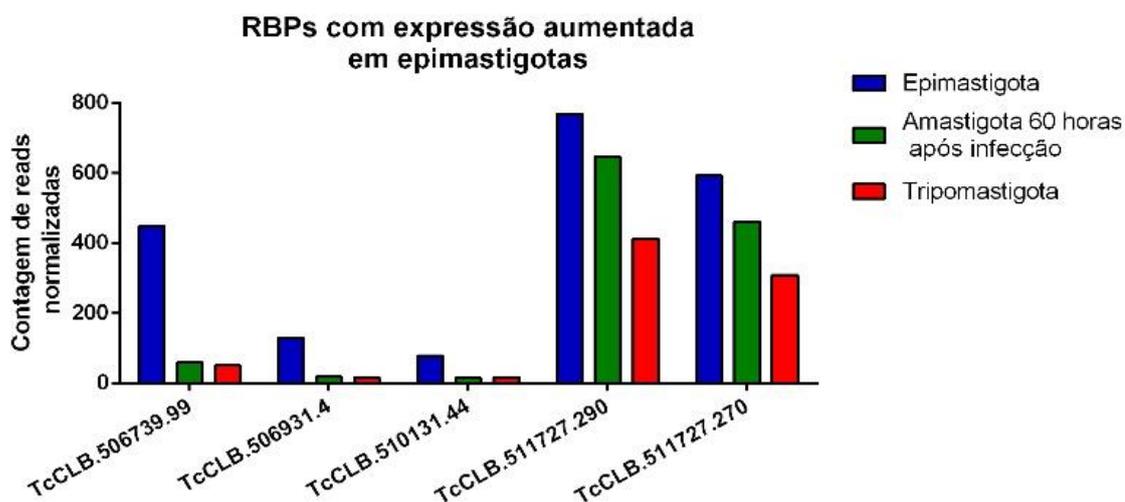


Figura 4. Proteínas de ligação à RNA (RBPs) com expressão aumentada em epimastigotas. Níveis de expressão em epimastigota comparadas com amastigotas coletadas 60 horas após a infecção e tripomastigotas liberadas em cultura. Contagens normalizadas das *reads* que alinham nas RBPs com expressão aumentada em formas epimastigotas fornecem o perfil de expressão nas diferentes formas. Em azul está representada as formas epimastigotas, em verde amastigotas coletadas 60 horas após a infecção e vermelho formas tripomastigotas. (VALENTE et al. – manuscrito em preparação)

Valente et al. (manuscrito em preparação), por meio de buscas por domínios de ligação a RNA constatou a presença do domínio dedo de zinco na estrutura da TcRBP39, semelhante ao encontrado na proteína já caracterizada e anotada como proteína *zinc finger 1* (ZFP1) (Gene ID: TcCLB.511511.6). Como descrito anteriormente, ZFP1 havia sido caracterizada na cepa Dm28c como uma RBP com expressão aumentada em epimastigotas submetidas a estresse nutricional, sugerindo papel associado à diferenciação das formas epimastigotas em tripomastigotas metacíclicas (MÖRKING et al., 2004). Outro estudo mostra que a proteína homóloga presente em *T. brucei*, TbZFP1 (Tb927.6.3490) foi identificada com níveis aumentados durante a diferenciação de formas sanguínea em formas procíclicas (HENDRIKS et al., 2001).

Com o objetivo de investigar o papel da proteína codificada pelo gene TcCLB.506739.99, nomeada TcRBP99, linhagens nocautes para o gene codificador dessa proteína foram geradas (VALENTE et al., manuscrito em preparação). A caracterização dessas linhagens comparadas com epimastigotas *wild-type* (WT) mostrou que nos parasitos nocautes o processo de metaciclologênese foi afetado, pois estes possuem maior capacidade de diferenciação em tripomastigotas metacíclicas após serem submetido a estresse nutricional. Em contrapartida, a superexpressão da proteína TcRBP99 levou a uma diminuição na proporção de epimastigotas se diferenciando em tripomastigotas metacíclicas nas mesmas condições (VALENTE et al., manuscrito em preparação). Ao analisar esses dados, a pergunta sobre quais genes cujas expressões estariam sendo afetadas pela perda da expressão do gene codificador da TcRBP99 foi levantada, visando ser respondida neste trabalho.

1.6 Análise de expressão diferencial e suas implicações

Dado o conceito do transcriptoma como sendo o conjunto de transcritos (RNAs mensageiros, RNAs ribossômicos, RNAs transportadores e os microRNAs) de um dado organismo, órgão, tecido ou linhagem celular, é possível desenvolver uma infinidade de análises que vão desde a catalogação de todas as espécies de transcritos até a identificação de novos genes, descoberta de rearranjos na transcrição, identificação de SNPs, padrões de *splicing* e outras modificações pós-

transcricionais (CLOONAN et al., 2008; MOROZOVA et al., 2009; WANG et al., 2009). Para além disso, a partir desses dados podemos traçar o perfil de expressão gênica de qualquer organismo, através da quantificação de transcritos presentes em determinado ponto do desenvolvimento e condição a ser analisada (WANG et al., 2009).

O maior desafio, contudo, é a análise dos dados produzidos por diferentes técnicas de sequenciamento, as quais vêm se desenvolvendo rapidamente nos últimos anos, permitindo a geração de até 20 bilhões de *reads* por corrida, como é o caso da plataforma Illumina *Novaseq 6000* (ILLUMINA, 2017; KOBOLDT et al., 2010). Além disso, com o desenvolvimento da técnica de RNA-seq (sequenciamento de RNA), a qual permite o sequenciamento de todo o transcriptoma de um organismo, de forma precisa e alta reprodutibilidade, o uso de ferramentas computacionais torna-se essencial para a interpretação dos dados gerados (HAN et al., 2015; NAGALAKSHMI et al., 2008).

Para a quantificação dos níveis de expressão em diferentes condições e inferência de diferenças para determinados genes é necessário a realização de um estudo de análise de expressão diferencial, a qual implica em processamentos computacionais de dados, *a priori*. Filtragem de sequências de baixa qualidade, por exemplo, contribui para a redução de falsos positivos, por eliminar sequências de adaptadores e contaminantes (etapa de pré-processamento) (ISAKOV; SHOMRO, 2011; THAKUR; VARSHNEY, 2010). Além disso, o mapeamento de *reads* contra um genoma de referência desempenha um papel chave na identificação de genes GDE. Por fim, a contagem de *reads* mapeadas para cada gene permite a identificação dos verdadeiros níveis de expressão, através de métodos estatísticos atribuídos à diferentes pacotes desenvolvidos com esta finalidade (FINOTELLO; DI CAMILLO, 2015; WANG et al., 2009).

Detectar genes que são diferencialmente expressos em determinadas condições contruibui para esclarecer o papel dos mesmos nos organismos de interesse ou suas implicações em outros genes. Isso é feito através da identificação de mudanças significativas na abundância de mRNAs entre as comparações realizadas (FINOTELLO; DI CAMILLO, 2015; OSHLACK et al., 2010).

Comparações entre organismos nocautes ou silenciados para determinados genes com organismos do tipo selvagem contribui para o entendimento do efeito destes genes em diversos processos, e ainda permite a identificação de possíveis expressões coordenadas, ou seja, se um determinado gene influencia nos níveis de expressão de outros genes. Dessa forma, pacotes de análise estatística focadas no estudo de expressão diferencial têm se tornado alvo de estudos, com o objetivo de decidir o mais adequado para determinada condição experimental (OSHLACK et al., 2010).

EdgeR, limma e Deseq2 são pacotes utilizados com a finalidade de detectar genes diferencialmente expressos em condições normais ou mesmo em condições onde o organismo foi submetido à alguma técnica de nocaute, silenciamento, tratado com alguma droga ou mesmo submetido a algum tipo de estímulo, seja ele físico, químico ou biológico (ANDERS, 2010; ROBINSON et al., 2010; SMYTH, 2005). Os resultados fornecidos pelos três pacotes são baseados em metodologias distintas de filtragem das contagens de *reads*, distribuição, normalização, cálculo de dispersão e cálculo dos valores de expressão inerentes a cada gene analisado. Sendo assim, alguns parâmetros são levados em consideração na escolha do pacote a ser utilizado, como por exemplo, número de replicatas e taxa de falsos positivos (FDR – *False Discovery rate*). Estudos de *benchmarking* levam em conta estes parâmetros para compará-los e destacar as vantagens e desvantagens para cada um deles (GUO et al., 2013; SCHURCH et al., 2016; SONESON; DELORENZI, 2013).

A normalização dos dados de contagens é uma etapa extremamente importante, uma vez que existe o viés de maior quantidade de *reads* mapeando para genes maiores, o que acaba “mascarando” os níveis de expressão em relação a genes menores, onde se espera que menor quantidade de *reads* estejam sendo mapeadas (BULLARD et al., 2010). Portanto, diminuir as diferenças entre e dentro das amostras é essencial para a obtenção de GDE com maior precisão (MORTAZAVI et al., 2008; OSHLACK et al., 2010; ROBINSON; OSHLACK, 2010). O método utilizado para cada pacote estatístico aqui citado será melhor abordado nas seções de Metodologia e Discussão.

A distribuição assumida na análise de GDE é outro fator determinante na qualidade dos resultados obtidos. As primeiras metodologias desenvolvidas

assumiam distribuição de Poisson para os dados de contagens de diferentes amostras, a qual assume média e variância igual entre elas (BULLARD et al., 2010; MARIONI et al., 2008). Contudo, a variabilidade biológica entre as amostras é evidente e assumir este tipo de distribuição acaba levando a uma taxa de falsos positivos elevada. Sendo assim, metodologias baseadas na distribuição binomial negativa, uma extensão da distribuição de Poisson, passaram a receber maior atenção, assim acontece para os pacotes edgeR (ROBINSON et al., 2010) e Deseq2 (LOVE; HUBER; ANDERS, 2014), onde a variância não é assumida como sendo igual a média, devendo, portanto, ser calculada a partir de um parâmetro adicional (BULLARD et al., 2010; ROBINSON; SMYTH, 2007a, 2007b). Nas seções Metodologia e Discussão o método estatístico de distribuição será também abordado com mais detalhes.

2. JUSTIFICATIVA

Diante dos avanços em metodologias de sequenciamento, ferramentas de bioinformática e métodos estatísticos frente a tantas perguntas sem respostas fica evidente a importância do estudo aqui proposto.

Os recentes estudos do transcriptoma de *T. cruzi* possibilitaram a expansão do conhecimento acerca dos mecanismos de transcrição e tradução do parasita, elucidando eventos de *splicing* e poliadenilação alternativos. Não menos importante, o genoma montado e anotado de *T. cruzi*, juntamente com ferramentas de bioinformática e biologia molecular de alto rendimento permitiram a identificação e caracterização de vários elementos atuando em *cis* ou *trans* para regular as taxas de expressão gênica, garantindo a eficiência dos processos de diferenciação e consequente adaptação do parasita às diferentes condições metabólicas ao qual é submetido.

Apesar disso, muitos estudos ainda são necessários para que se tenha uma visão completa dos elementos responsáveis pela regulação de diversos genes. Tendo em vista que apenas 13 RBPs foram caracterizadas funcionalmente em *T. cruzi* até o presente momento, a caracterização de outras RBPs, identificando as taxas de expressão nas diferentes formas de vida do parasita, o mecanismo pelo qual atuam sobre mRNAs e seus potenciais alvos, bem como a identificação de motivos de

sequência presentes na região UTR de mRNAs, contribuem para o entendimento de um dos principais mecanismos de regulação pós-transcricional existente em *T. cruzi*.

A identificação de elementos importantes para a sobrevivência do parasita, seja no hospedeiro invertebrado ou vertebrado, nós leva a propor um potencial alvo para o desenvolvimento de novos fármacos que podem atuar no controle das taxas de expressão de determinadas RBPs, no mecanismo de ação, podendo impedir o ligamento das mesmas à mRNAs alvos ou impedir a formação de complexos importantes, ou ainda atuar sobre mRNAs que tenham sido identificados como fundamentais para o processo de diferenciação celular, crescimento, motilidade ou sobrevivência das formas de vida de *T. cruzi*. Dessa forma, o planejamento para o desenvolvimento de novos fármacos poderá ser mais eficiente, ao passo que estratégias que visem um alvo bem conhecido poderão ser traçadas, resultando em fármacos com alta eficácia e de baixa toxicidade, contrapondo os fármacos atualmente utilizados para o tratamento da doença de Chagas: benzonidazol e Nifurtimox.

A busca e identificação de motivos de sequência envolvidos no processamento pós-transcricional exige a compreensão de regiões UTRs, como por exemplo, o tamanho das mesmas e a disposição de elementos presentes, que desempenham diferentes papéis, podendo atuar como ponto de reconhecimento de RBPs ou ainda como sinalizadoras para processos de *splicing* e poliadenilação alternativa. Sendo assim, este estudo fornece informações adicionais para o entendimento da regulação da expressão exercida pela TcRBP99.

3. OBJETIVOS

3.1 Objetivo Geral

- Analisar o efeito da deleção de uma proteína de ligação ao RNA no controle da expressão gênica de *T. cruzi*, comparando a expressão gênica global de epimastigotas selvagens e linhagens nocautes para TcRBP99.

3.2 Objetivos específicos

- Analisar e remover regiões de baixa qualidade de *reads* brutas, obtidas a partir do sequenciamento de bibliotecas de cDNA de epimastigotas selvagens e nocautes para a TcRBP99, para obtenção de *reads* de boa qualidade (*phred* \geq 20);
- Mapear *reads* no genoma de referência completo (alelos *Esmeraldo-like*, *Non-Esmeraldo* e *contigs* não inseridos na versão final do genoma) do clone CL Brener de *T. cruzi*;
- Identificar genes diferencialmente expressos em epimastigotas nocautes comparadas com epimastigotas selvagens;
- Determinar os limites das regiões 3' UTRs dos mRNAs identificados como potenciais alvos da proteína TcRBP99;
- Identificar elementos conservados (motivos) nas regiões 3' UTRs dos mRNAs que apresentaram mudança de expressão na comparação entre epimastigotas selvagens e epimastigotas nocautes;

4. METODOLOGIA

4.1 Isolamento de RNA total e construção de bibliotecas

Bibliotecas de cDNA foram obtidas a partir de dois clones nocautes para o gene codificador da proteína TcRBP99 de epimastigotas, em duplicata. Foram também geradas bibliotecas de cDNA em triplicatas, a partir de epimastigotas *wild-type* (WT). A extração do RNA foi conduzida pela aluna de doutorado Bruna Mattioly Valente e a geração de bibliotecas de cDNA foi realizada no Laboratório de Genômica e Biologia Computacional, localizado no Instituto René Rachou – Fiocruz Minas.

Para a construção das bibliotecas de cDNA foi utilizado o *kit v2 Illumina® TruSeq® RNA Sample Preparation*. Como mostrado na figura 5, RNA total, ligado previamente à sequências *barcodes* (em cinza), é filtrado por meio de duas eluições com *beads* magnéticas contendo oligo-dT (etapa 1), as quais capturam apenas mRNAs que possuem cauda poli(A). Qualquer RNA ribossômico (rRNA) ou outras espécies de RNAs contaminantes foram eliminadas por meio da adição de *Bead Washing Buffer* (BWB) à mistura de RNA e *beads* magnéticas. Durante a segunda eluição do RNA contendo cauda poli(A), o mesmo é fragmentado (etapa 2) e misturado a hexâmeros aleatórios que atuam como *primers* para posterior síntese da primeira cadeia de cDNA. A síntese desta cadeia foi então realizada pela atividade da Transcriptase Reversa (TR) *SuperScript II* (etapa 3). No passo seguinte foi realizada a síntese da segunda cadeia de cDNA pela TR, e remoção do *template* de RNA. Por fim, a ligação de múltiplos adaptadores de indexação (em preto) às extremidades das cadeias duplas de cDNA foi realizada, preparando-os para hibridização à adaptadores complementares presentes em uma *flow cell*.

A partir do procedimento descrito anteriormente, fragmentos de 300pb, em média, foram gerados e posteriormente sequenciados (etapa 4). Diversas *reads* foram obtidas por meio do sequenciamento das bibliotecas de cDNA (técnica RNA-seq), o qual foi realizado na plataforma *Illumina MiSeq®* no Instituto René Rachou – Fiocruz Minas.

Para o sequenciamento, fragmentos de cDNA que continham adaptadores em ambas as extremidades foram seletivamente enriquecidos por meio da “PCR em

“ponte” (formação de *clusters*), utilizando um *cocktail* de *primers*, os quais anelam nas extremidades dos adaptadores ligados a *flowcell*. Procedimentos de validação foram conduzidos, como a quantificação das bibliotecas de cDNA, por meio de qPCR e análise do controle de qualidade das mesmas no *Agilent Technologies 2100 Bioanalyzer*. O rendimento obtido foi de aproximadamente 8,75 Mb de seqüências para cada biblioteca sequenciada.

As amostras foram nomeadas de acordo com a natureza das epimastigotas (nocautes ou WT): “nocaute 01” e “nocaute 02”, duplicatas do primeiro clone nocaute selecionado; “nocaute 03” e “nocaute 04”, duplicatas do segundo clone nocaute selecionado; “WT 01”, “WT 02” e “WT 03”, triplicatas de epimastigotas WT.

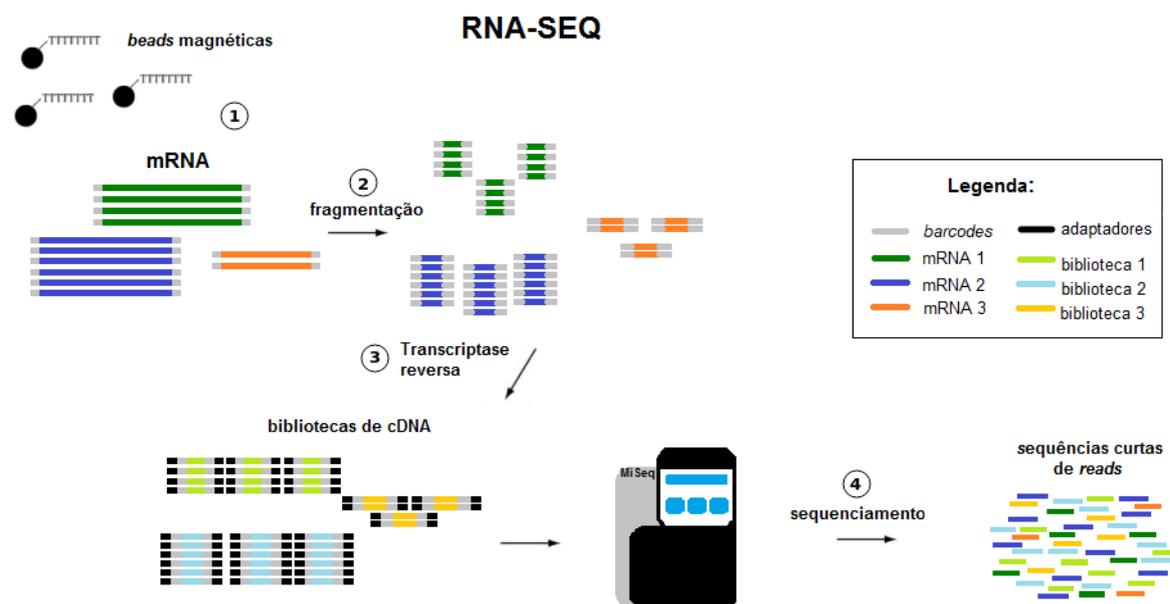


Figura 5. Esquema representativo da técnica de RNA-Seq. RNA total previamente ligado a seqüências *barcodes* (em cinza), é extraído da célula e mRNA é capturado através de *beads* magnéticas com oligodT (1). Realiza-se então a fragmentação do mRNA em seqüências curtas de tamanho escolhido (300pb) (2). Dessa forma, cDNA é produzido através de uma reação de amplificação envolvendo a enzima Transcriptase Reversa (gera DNA a partir de uma molécula de RNA) (3). Posteriormente, adaptadores são ligados nas extremidades de cada fragmento (em preto), de forma a permitir a ligação de *primers*. Feito isso, o cDNA é sequenciado, dando origem a bibliotecas de *reads* (4), que irão passar por controle de qualidade e serem usadas para futuras análises.

4.2 Análise da qualidade geral das *reads*

As sete bibliotecas de *reads* correspondentes ao transcriptoma de epimastigotas nocautes e epimastigotas WT de *T. cruzi* foram analisadas quanto a qualidade das bases sequenciadas. Para isso foi utilizada a ferramenta FastQC (Versão 0.10.1) (ANDREWS, 2010).

As bibliotecas foram analisadas com base nos seguintes parâmetros fornecidos pela ferramenta: número total de *reads* por biblioteca (estatísticas básicas), qualidade das sequências por base, *score* de qualidade por sequência, conteúdo de sequência por base (proporção de cada uma das quatro bases (A, T, C e G), em cada posição sequenciada), conteúdo GC por sequência, conteúdo de N por base e distribuição do tamanho de sequências.

Foi estabelecido um *cut-off* de qualidade de bases, sendo este igual a 20 (*phred score* \geq 20) para todas as bibliotecas analisadas, ou seja, admitiu-se a possibilidade de 1 erro a cada 100 bases sequenciadas. Sendo assim, na etapa posterior, foi realizada a filtragem de *reads* de baixa qualidade.

4.3 Trimagem e filtragem de *reads* com bases de baixa qualidade

As bibliotecas sequenciadas (*paired-end*, predominância de *reads* de 250 pb) foram submetidas ao processo de eliminação de nucleotídeos de baixa qualidade, bem como a remoção de adaptadores contaminantes, usando a ferramenta Trimmomatic (Versão 0.32) (BOLGER et al., 2014). Para isso foram realizados alguns testes consecutivos visando avaliar o melhor conjunto de parâmetros com capacidade de gerar dados de alta qualidade e sem uma perda expressiva do total de *reads* disponíveis em cada biblioteca. O conjunto estabelecido segue a configuração baseada na linha de comando abaixo:

```
trimmomatic-0.32.jar      PE      -phred33      input_read_foward.fastq.gz  
input_read_reverse.fastq.gz      output_foward_paired.fastq  
output_foward_unpaired.fastq      output_reverse_paired.fastq
```

```
output_reverse_unpaired.fastq ILLUMINACLIP:TruSeq2-PE.fa:2:30:10 LEADING:5  
TRAILING:5 SLIDINGWINDOW:4:20 MINLEN:50
```

Descrição dos parâmetros:

ILLUMINACLIP: remove sequências de adaptadores presentes nos fragmentos;

LEADING: remove nucleotídeos com qualidade abaixo de 5 no início da sequência;

TRAILING: remove nucleotídeos com qualidade abaixo de 5 no final da sequência;

SLIDINGWINDOW: janela deslizante de tamanho 4, que executa a remoção de porções das *reads* onde a média de qualidade cai abaixo de 20;

MINLEN: remove *reads* menores que 50 nucleotídeos de comprimento.

4.4 Mapeamento das *reads*

As *reads* de alta qualidade foram mapeadas contra o genoma de referência de *T. cruzi*, cepa CL Brener, utilizando o programa Tophat2 (KIM et al., 2013). A referência foi formada a partir das sequências FASTA dos haplótipos *Esmeraldo-like* e *Non Esmeraldo-like*, bem como de *contigs* de CL Brener que não foram atribuídos a nenhum destes haplótipos (*reads unassigned*). Os arquivos FASTA foram obtidos no banco de dados TritrypDB (<http://tritrypdb.org/>) (ASLETT et al., 2010) a partir da versão mais recente disponibilizada no momento da execução desta etapa (versão 31, “TriTrypDB-31_TcruziCLBrenerEsmeraldo-like_Genome.fasta”, disponibilizada em 03 de junho de 2017).

Antes de iniciar o mapeamento das *reads*, foi necessário gerar o índice do genoma. Para isso foi utilizado o programa bowtie2-build, que faz parte do pacote de programas instalados junto com o programa Bowtie2 (LANGMEAD; SALZBERG, 2012). A indexação foi executada segundo a linha de comando abaixo:

```
bowtie2-build TriTrypDB-31_TcruziCLBrener_Genome.completo.fasta  
tophat_complete.gff
```

Para a execução do Tophat2 o seguinte comando foi executado:

```
tophat2 -o tophat_complete_sample-1-1 -g 20 -p 6 -N 2 --no-discordant --no-mixed  
--keep-fasta-order prefix_index input_trimmed_read_foward.fastq  
input_trimmed_read_reverse.fastq
```

Descrição dos parâmetros:

-N: número máximo de *mismatches* permitidos: 2.

-o/--output-dir: caminho para o *output*.

-p/--num-threads: número de *threads* a serem usados: 6.

-g/--max-multihits: número máximos de alinhamentos permitidos: 20.

--no-discordant: não relata alinhamentos discordantes.

--no-mixed: não mostra no *output reads* que mapeiam apenas umas das extremidades.

--keep-fasta-order: classifica os alinhamentos na mesma ordem no arquivo genoma *fasta*.

Os alinhamentos (arquivos *.bam*) gerados foram ordenados e indexados usando o programa *samtools* (<http://samtools.sourceforge.net>) a fim de considerar apenas os pares de *reads* apropriados (*foward e reverse*). Os arquivos BAM (versão binária com as coordenadas das *reads* mapeadas) foram então convertidos para formato SAM (versão texto, contendo as mesmas informações do arquivo BAM) utilizando também o pacote de programas *samtools* (LI et al., 2009).

4.5 Contagem das *reads* por gene

A contagem da quantidade de *reads* (*forward e reverse*) que mapearam em cada gene do genoma de *T. cruzi* foi realizada por meio do programa *Htseq* (versão 0.6.0) (ANDERS et al., 2015) segundo a linha de comando abaixo:

```
htseq-count -s no -q output_tophat.sam TriTrypDB-31_TcruziCLBrenner_Genome.completo.fasta tophat_complete.gff > output-htseq
```

Descrição dos parâmetros:

-s: a *read* é considerada sobreposta com uma *feature*, independentemente de ser mapeada para o mesmo sentido ou para o sentido oposto da fita de DNA, representada pela *feature*.

-q: impede o aparecimento do relatório de progresso e avisos.

O total de *reads* mapeadas para cada gene foi tabulado, gerando assim uma tabela de contagens para cada uma das bibliotecas de RNA-seq analisadas. Esta tabela foi construída de forma a permitir a seguinte comparação na etapa de análise de expressão diferencial:

$$\text{Comparação} = \frac{\text{epi. } \textit{knockout}}{\text{epi. } \textit{wild-type}}$$

4.6 Análise das contagens e expressão diferencial

As amostras nocautes e WT foram avaliadas quanto à reprodutibilidade. Foi investigada a ocorrência de possíveis *outliers* através de agrupamento hierárquico (baseados em distância euclidiana) e análise de componente principal (PCA). Além disso, os genes cuja soma das contagens para todas as bibliotecas foi inferior a 50, não foram considerados nas análises posteriores.

A tabela resultante da filtragem de baixas contagens foi inicialmente transformada para a unidade de expressão FPKM (*Fragments Per Kilobase Million*), a qual reflete a “normalização” das contagens brutas levando em consideração o tamanho da biblioteca e o tamanho de cada gene, diminuindo o viés de maior quantidade de *reads* mapeadas para genes maiores. Esta tabela foi utilizada para avaliação de expressão a partir das contagens.

Contudo, a tabela de contagens brutas, que passaram apenas pelo filtro de qualidade e baixas contagens, foi utilizada para análise de expressão diferencial, a qual foi conduzida no ambiente R de programação (versão 3.2.3). Para tanto, três pacotes estatísticos foram abordados nesta etapa: DESeq2 (LOVE et al., 2014), limma (SMYTH, 2005) e EdgeR (ROBINSON et al., 2010), através da execução de *scripts* construídos na linguagem R. Os resultados fornecidos por todos eles foram comparados por meio do agrupamento de genes diferencialmente expressos comuns aos três pacotes. O desempenho dos mesmos também foi avaliado extensamente por meio de estudos de *benchmarking* encontrados na literatura (GUO et al., 2013; SCHURCH et al., 2016; SONESON; DELORENZI, 2013). Diante dos resultados obtidos (descritos na próxima seção) foi decidido considerar apenas os genes diferencialmente expressos (GDE) identificados pelo DESeq2.

As análises de expressão diferencial foram conduzidas de modo que, para ser considerado gene diferencialmente expresso, o mesmo deveria obedecer dois parâmetros pré-estabelecidos: p ajustado (p_{adj}) $< 0,05$ e $-1,1 > \text{Log}_2\text{FoldChange} > 1,1$ (mudança de expressão maior ou menor que 2,14 vezes). Os métodos de normalização e distribuição assumidos são inerentes a cada um dos pacotes. Para edgeR e limma, o método de normalização utilizado é o TMM (*Trimmed mean of M-values*), que estima fatores de escala entre as amostras baseado na "produção relativa de RNA de 2 amostras", sob o pressuposto de que a maioria dos genes não são diferencialmente expressos (LAW et al., 2014; ROBINSON; OSHLACK, 2010). Deseq2 considera um método baseado em "fatores de tamanho" específico de cada amostra que tornam comparáveis as contagens de diferentes amostras (ANDERS; HUBER, 2010). Deseq2 e edgeR assumem distribuição binomial negativa, ao contrário de limma, que assume distribuição log normal (ANDERS; HUBER, 2010; LAW et al., 2014; ROBINSON et al., 2010).

Considerando o pacote DESeq2 para análise de GDE, primeiramente foi estabelecida a comparação a ser avaliada: *epi. knockouts vs epi. wild-type*, a qual foi armazenada no objeto denominado **colData**. As quatro bibliotecas nocautes foram consideradas como um único grupo, assim como as bibliotecas WT. A função **DESeqDataSetFromMatrix** foi utilizada para a criação de outro objeto, denominado **ddsData**. Para isso, foram aplicadas as contagens filtradas para baixas contagens e

o objeto **colData**, direcionados pela condição estabelecida para o *design* experimental. Por fim, a função **DESeq** foi utilizada para estimar os fatores de normalização dos dados (função **estimateSizeFactor**), cálculo da dispersão entre as contagens observadas (função **estimateDispersions**) e aplicação do modelo GLM binomial negativo, o qual testa a significância dos coeficientes para o Log2FoldChange (Log2FC) calculado para cada gene, em cada amostra (função **nbinomWaldTest**). A tabela de resultados foi gerada a partir da função **results**, a qual extrai valores Log2FC, padj e *p-values* resultantes da análise de GDE, para cada gene (Apêndice 1).

Para a visualização e posterior análise dos resultados foi construído um gráfico de Log2FC (MA-plot) de todos os genes analisados, através da função **plotMA** (biblioteca DESeq2), **vennDiagram** e **plot_ly** (limma) foram utilizados para produzir o diagrama de Venn da comparação entre os três pacotes de análise de GDE e MA plot em barras, respectivamente (Apêndice 1).

4.7 Mapeamento e alinhamento das regiões 3' UTR de genes diferencialmente expressos

Para determinar o comprimento das regiões 3' UTR foram utilizadas *reads* geradas a partir de mRNA das formas epimastigotas e amastigotas obtidas 96 horas após a infecção de células HFF (esta última foi utilizada apenas para comparação com a biblioteca de epimastigotas), em colaboração com o Prof. Nagib El-Sayed da Universidade de Maryland (BELEW et al., 2017). O mapeamento destas *reads* no genoma do clone CL Brener (versão 31, "TriTrypDB-31_TcruziCLBrenerEsmeraldo-like_Genome.fasta") foi visualizado utilizando a ferramenta IGV (*Integrative Genomics Viewer*). Dessa forma, as *reads* previamente mapeadas para todo o genoma puderam ser exploradas por meio de interface gráfica, ou seja, por meio da visualização interativa de toda a sequência genômica e posição exata do mapeamento das *reads* para cada *feature* (CDS).

Um *pipeline* específico, que considera a identificação de sítios de poliadenilação e *trans-splicing* foi utilizado para a correta identificação das regiões UTRs. Primeiramente, *reads* que a princípio não mapearam no genoma (armazenadas

no arquivo *unmapped.bam* gerado após o mapeamento total de *reads*) passaram por uma etapa de trimagem, de forma que todas as sequências poli(A) ou poli(T) *downstream* (máximo de 20 “A” ou “T”) de no mínimo quatro nucleotídeos de adenina ou timina foram excluídas, ou ainda, toda a sequência *upstream* (nucleotídeos do 1 ao 31) de *reads* contendo pelo menos as quatro últimas bases da sequência 5'-AACGCTATTATTGATACAGTTTCTGTAATATTG-3' ou seu complemento reverso. Essa etapa foi realizada por meio da ferramenta Cutadapt (versão 1.13; MARTIN, 2015). Para isso, as seguintes linhas de comando foram utilizadas:

```
cutadapt -g "A{20}" -m 20 -O 4 --discard-untrimmed -o poli(A).fastq input_file.fastq  
– para identificar regiões poli(A)
```

```
cutadapt -g "T{20}" -m 20 -O 4 --discard-untrimmed -o poli(T).fastq input_file.fastq  
– para identificar regiões poli(T)
```

```
cutadapt -g AACGCTATTATTGATACAGTTTCTGTAATATTG -g  
CAATATAGTACAGAACTGTATCAATAATAGCGTT -m 20 -O 4 --discard-untrimmed  
-o spliced-leader.fastq input_file.fastq – para identificar regiões de spliced leader
```

Descrição dos parâmetros:

-a: sequência do adaptador a ser cortado na extremidade 3'.

-g: sequência do adaptador a ser cortado na extremidade 5'.

-m (*minimum-length*): discarta *reads* menores que 20 nt.

-O (*overlap*): overlap mínimo do adaptador com a sequência buscada nas *reads*.

--discard-untrimmed: leva para o *output* apenas *reads* nas quais os adaptadores foram encontrados.

O algoritmo da ferramenta Cutadapt opera de modo a realizar um alinhamento semi-global, onde as sequências só são penalizadas na região sobreposta entre elas (MARTIN, 2017). Sendo assim, são consideradas todas as sobreposições possíveis entre duas sequências, mas para cada alinhamento é calculado um *score* baseado

em *matches* e *mismatches*. Alinhamentos de sequências sobrepostas que excedam a taxa de erro permitida (estabelecida na linha de comando ou 0,1 (10%) por *default*) são descartados e o alinhamento que apresente a maior região sobreposta e com o maior número de *matches* é o escolhido (MARTIN, 2017).

Esse filtro teve como objetivo retirar sequências que não estão presentes no genoma, mas sim, em mRNAs processados e conseqüentemente em bibliotecas de cDNA. Posteriormente, as *reads* trimadas foram remapeadas no genoma de CL Brener utilizando a ferramenta TopHat2. Para isso, as seguintes linhas de comando foram executadas:

```
tophat2 -o tophat_complete_sample-1-1 -g 1 -p 6 -N 2 --no-discordant --no-mixed --keep-fasta-order prefix_index input_trimmed_poly(A)read_foward.fastq

tophat2 -o tophat_complete_sample-1-1 -g 1 -p 6 -N 2 --no-discordant --no-mixed --keep-fasta-order prefix_index input_trimmed_poly(T)read_foward.fastq

tophat2 -o tophat_complete_sample-1-1 -g 1 -p 6 -N 2 --no-discordant --no-mixed --keep-fasta-order prefix_index input_trimmed_sliced-leader.read_foward.fastq input_trimmed_sliced-leader.read_reverse.fastq
```

Os mapeamentos resultantes das *reads* trimadas para regiões poli(A) e *spliced leader* de epimastigotas e amastigotas 96 horas foram carregados no visualizador IGV juntamente com os mapeamentos totais. As regiões *downstream* e *upstream* de cada *read* trimada foram verificadas para a determinação de sítios de poliadenilação e adição de *spliced-leader*. Dessa forma, definiu-se regiões 3' UTR como sendo regiões que vão desde a primeira base após o códon de parada do gene em questão, até onde havia *reads* trimadas que foram mapeadas, ou seja, até os sítios de poliadenilação identificados.

A partir das coordenadas obtidas para cada 3' UTR, as sequências correspondentes foram extraídas da base de dados TritrypDB (<http://tritrypdb.org/>) (ASLETT et al., 2010) utilizando a ferramenta *Genome browser*. Então, de forma manual, foram realizadas buscas por tratos de polipirimidina, definidos como sendo a

maior sequência rica em pirimidina, interrompida por no máximo uma purina após a 3' UTR.

Um resumo deste *pipeline* pode ser visto na Figura 6.

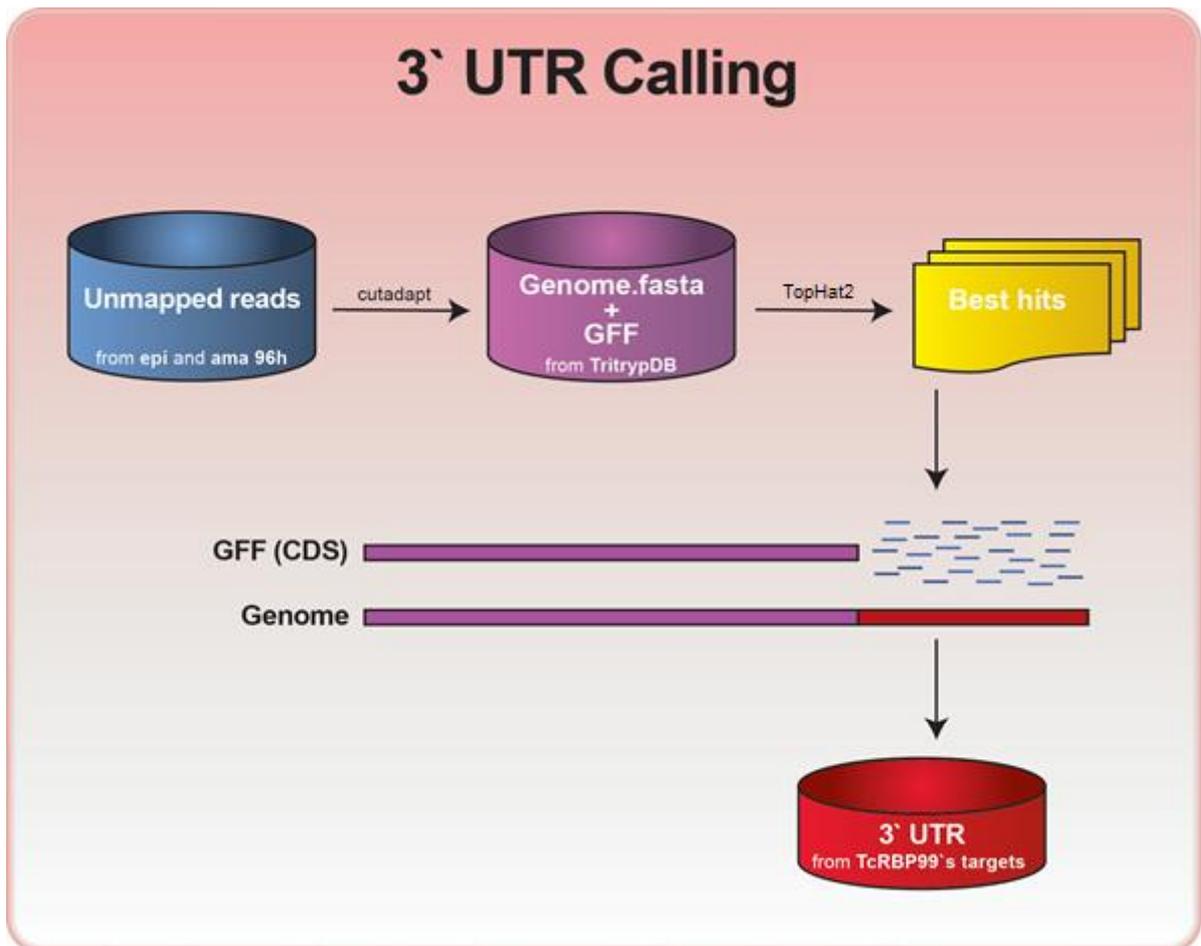


Figura 6. Pipeline para análise de 3' UTR de *Trypanosoma cruzi*. Para determinar as coordenadas de regiões 3' UTR de RNAs mensageiros (mRNA) alvos da proteína TcRBP99 foram utilizadas *reads unmapped* obtidas de mapeamento prévio de epimastigotas e amastigotas 96 horas pós infecção de células HFF. Estas passaram por etapa de trimagem para a remoção de sequências poly(A) ou Poly(T) na extremidade 3' e foram então remapeadas ao genoma de CL Brener, seguindo coordenadas contidas no arquivo GFF. *Best hits* foram utilizados para definir as coordenadas das regiões 3' UTRs.

4.8 Alinhamento múltiplo local de regiões 3' UTRs de genes alvos da TcRBP99 e comparação com as CDS correspondentes

As sequências previamente identificadas como 3' UTR dos mRNAs que interagem com a proteína TcRBP99 foram alinhados através da ferramenta *MUSCLE*, versão

3.8.31 (EMBL-EBI). A partir do resultado foi construída uma árvore filogenética baseada no algoritmo *Neighbour-joining* (EDGAR, 2004).

Da mesma forma, as CDS dos potenciais genes alvos da TcRBP99 foram alinhadas também com a ferramenta *MUSCLE* e a correspondente árvore filogenética foi construída pelo método citado anteriormente. Posteriormente os resultados de ambos foram comparados.

4.9 Predição de motivos na região 3' UTRs

Dada a importância da presença de elementos regulatórios em regiões 3' UTRs em *T. cruzi*, os mesmos foram identificados nos genes que potencialmente podem ter sua expressão alterada pela presença da proteína TcRBP99.

De posse das sequências das UTRs previamente retiradas do banco de dados TritypDB (<http://tritypdb.org/>) (ASLETT et al., 2010) foi utilizado uma versão *online* do *software* MEME suite (*Multiple EM for Motif Elicitation*), que é capaz de identificar motivos em um grupo de sequências relacionadas de DNA, RNA ou proteínas (BAILEY et al., 2015). Os resultados fornecidos pelo algoritmo são baseados em um *e-value* que deve ser menor que 0,05. Este parâmetro estatístico é baseado no logaritmo da razão de verossimilhança, largura e contagem dos sítios, *background* e o tamanho do conjunto de treinamento, ou seja, representa a chance de o motivo ter sido encontrado de forma aleatória, baseado no *background*. O resultado é dado em representações gráficas de acordo com a frequência das bases em cada posição. Além disso, diagramas em blocos representando os motivos descobertos são fornecidos juntamente com o arquivo de saída gerado para os gráficos de frequência. Opções de formato para o arquivo de saída são fornecidas: HTML, PNG, XML, de modo que se torna compatível com outras ferramentas da plataforma, como a ferramenta MAST (*Motif Alignment and Search Tool*), que permite comparar os motivos preditos entre as sequências dadas como entrada. Essa ferramenta considera exatamente uma ocorrência de cada motivo por sequência e o grau de similaridade entre as sequências. Por sua vez, os resultados são tidos como positivos a partir do *p-value* atribuído a cada sequência, o qual é baseado nas ocorrências de motivos individuais.

Para executar a predição de motivos no MEME, o seguinte comando foi considerado:

```
meme 3UTRs.txt -rna -oc . -nostatus -time 18000 -maxsize 60000 -mod anr -nmotifs 10 -minw 6 -maxw 50 -evt 1e-5
```

Descrição dos parâmetros:

-rna: especifica o tipo de sequência dada no *input*.
-oc: cria um diretório chamado dir.
-nostatus: não imprime mensagem de status na tela.
-time: tempo máximo, em segundos, que o algoritmo buscará por motivos.
-maxsize: tamanho máximo permitido do conjunto de dados.
-mod anr: qualquer número de repetições.
-nmotifs: número máximo de motivos buscados.
-minw: tamanho mínimo dos motivos.
-maxw: tamanho máximo dos motivos.
evt: *e-value* máximo permitido.

Por último, os motivos identificados no MEME foram submetidos à ferramenta MAST. Foram considerados motivos com *e-value* > 1e-5 para avaliação mais abrangente da distribuição dos motivos ao longo das sequências.

```
mast meme.xml 3UTRs.txt -oc . -nostatus
```

5. RESULTADOS

5.1 Identificação dos níveis de expressão e análises *in silico* da TcRBP99

Buscas no transcriptoma de *T. cruzi* e em dados adicionais, gerado a partir do mRNA de formas epimastigotas, amastigotas coletadas após 60 e 96 horas de infecção e formas tripomastigotas liberadas em cultura, mostraram que o gene que codifica a proteína TcRBP99 (TcCLB.506739.99) apresenta expressão 8 vezes aumentada em epimastigotas em comparação com as demais condições analisadas (Figura 7) (BELEW et al., 2017; VALENTE et al., manuscrito em preparação).

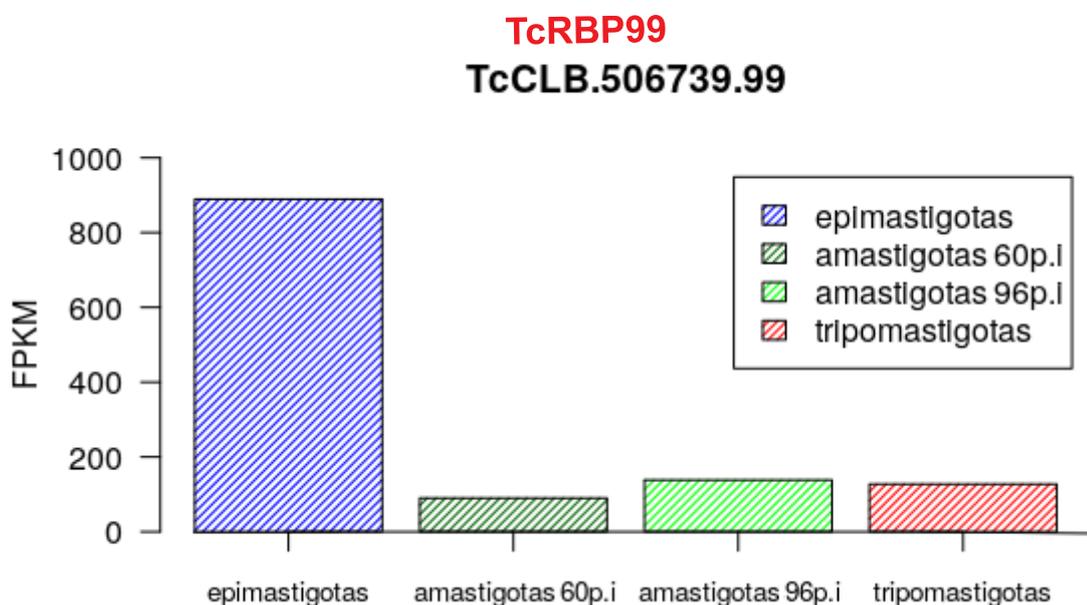


Figura 7. Níveis de expressão do alelo TcCLB.506739 do clone CL Brener nas diferentes formas de vida de *T. cruzi*. (Valores de expressão obtidos a partir do transcriptoma gerado pelo Prof. Nagib El-Sayed e colaboradores, BELEW et al., 2017)

Predições *in silico*, na base de dados TritypDB, da TcRBP99 mostraram que o alelo TcCLB.506739.99 está localizado no cromossomo 3 do haplótipo *Esmeraldo-like* do clone CL Brener de *T. cruzi*. Além disso, as buscas indicaram a presença de um alelo codificador para a mesma proteína no segundo haplótipo de CL Brener, denominado *Non-Esmeraldo-like*. Este alelo é identificado na base de dados pelo ID TcCLB.510819.119, o qual compartilha várias características com TcCLB.506739.99 (Tabela 1). Outra informação obtida é a de que a TcRBP99 se trata de uma proteína

de 19,54 kDa. Sua região codificadora é composta de 519 pb e a proteína de 173 aminoácidos. O ponto isoelétrico (pI) predito, igual a 8,49, o que indica o caráter básico da proteína, reforçando a função de ligação à ácidos nucleicos, que por sua vez possui cargas negativas.

Tabela 1. Principais características dos alelos que codificam a TcRBP99.

	TcCLB.506739.99	TcCLB.510819.119
Haplótipos	<i>Esmeraldo-like</i>	<i>Non Esmeraldo-like</i>
Localização	Posição 46.294 a 46.815 do cromossomo 3	Posição 46.651 a 47.172 do cromossomo 3
Tamanho do gene	522 bases	522 bases
Tamanho da proteína	173 aminoácidos	173 aminoácidos
Localização do motivo dedo de zinco	Entre as bases 15 e 41	Entre as bases 15 e 41
Ponto isoelétrico	8,49	8,29

A pequena diferença observada no pI pode ser devido a divergência de dois aminoácidos entre as sequências dos alelos TcCLB.506739.99 e TcCLB.510819.119 (Figura 8), sobretudo em relação aos aminoácidos tirosina, presente no primeiro alelo e histidina, presente no segundo alelo. Esses aminoácidos possuem características químicas diferentes, uma vez que tirosina possui um grupo aromático em sua cadeia lateral (R) e a histidina possui R carregado positivamente.

TcCLB.506739.99	1	MPKANKHAEVKPSKYRTALCEFYLRQEECPFGTRCAFAHGEHELQTEERNVELLKATGLQ	60	TcRBP99
TcCLB.510819.119	1	MPKANKHAEVKPSKYRTALCEFYLRQEECPFGTRCAFAHGEHELQTEERNVELLKATGLQ	60	alelo <i>Non-Esmo</i>
TcCLB.506739.99	61	RLDAAI SPTTRCSAMKAESSLNAVFPVTSLSRRHPVIVSLPCWTESRVVGFERSEPTTP	120	TcRBP99
TcCLB.510819.119	61	RLDAAI PTPTRCSAM AESSLNAVFPVTSLSRRHPVIVSLPCWTESRVVGFERSEPTTP	120	alelo <i>Non-Esmo</i>
TcCLB.506739.99	121	GKRLDEREAAPSFSPAREVHRARCSSRFSA TCSTEAPVMYRHNPYA LSTHYE	173	TcRBP99
TcCLB.510819.119	121	GKRLDEREAAPSFSPAREVHRARCSSRFSA TCSTEAPVMYRHNPYA LSTHYE	173	alelo <i>Non-Esmo</i>

Figura 8. Alinhamento local das sequências correspondentes aos alelos TcCLB.506739.99 e TcCLB.510819.119 de *T. cruzi*. A primeira sequência alinhada corresponde ao alelo que codifica a TcRBP99, localizado no haplótipo *Esmeraldo-like*. A segunda sequência corresponde ao alelo localizado no haplótipo *Non-Esmeraldo-like (Non-Esmo)*. Os retângulos azul e vermelho correspondem as diferenças de aminoácidos entre os alelos. A primeira diferença é entre lisina e isoleucina, ambos aminoácidos com cadeia lateral alifática, não polar; a segunda diferença é entre os aminoácidos tirosina e histidina, com cadeia lateral aromática e carregada positivamente, respectivamente.

Outra informação que pôde ser constatada por meio de alinhamento múltiplo de todas as sequências sintênicas ao gene TcCLB.506739.99, depositadas no TritypDB, é a presença e alta conservação do domínio dedo de zinco, semelhante ao encontrado na proteína já caracterizada e anotada como proteína zinc finger 1 (ZFP1) (Figura 9). Este domínio está localizado na região N-terminal da proteína que mostrou ser também altamente conservada dentre as sequências alinhadas.



Figura 9. Alinhamento múltiplo das sequências dos genes sintênicos a TcRBP99 depositados no banco de dados TritypDB. A proteína TcRBP99 está destacada na figura pela cor azul. O domínio dedo de zinco C-x8-C-x5-C-x3-H está indicado pela cor vermelha em todas as sequências alinhadas. Os resíduos centrais do domínio (CCH) estão indicados pela cor amarela.

5.2 Análise de reads obtidas no sequenciamento de bibliotecas de cDNA de epimastigotas WT e nocautes para a TcRBP99

Bibliotecas de cDNA foram geradas a partir de amostras de RNA obtidas em duplicata de dois clones de epimastigotas nocautes para a TcRBP99 e amostras de RNA obtidas em triplicata de culturas de epimastigotas WT. Posteriormente estas bibliotecas foram sequenciadas e analisadas quanto à qualidade. Adaptadores e sequências de baixa qualidade (*phred score* < 20) foram então removidos. Por fim, resultados para os dados brutos e filtrados com base na qualidade foram visualizados por meio de arquivos HTML fastqc. Um resumo com os principais resultados obtidos por essas etapas está na Tabela 2.

Tabela 2. Resultados do sequenciamento das bibliotecas de cDNA de duas linhagens de epimastigotas nocautes e epimastigotas WT.

Biblioteca	Total de reads	Total de reads pós trimagem
Nocaute 01	3.513.887	2.741.084
Nocaute 02	3.895.258	3.088.360
Nocaute 03	4.143.756	3.274.477
Nocaute 04	3.521.679	2.820.681
WT 01	3.294.361	2.626.697
WT 02	3.609.438	2.862.761
WT 03	3.469.573	2.709.211
Total	25.447.952	20.123.271

*Nocautes 01 e 02 foram obtidos de um primeiro clone e os nocautes 03 e 04 foram obtidos de um segundo clone selecionado, mas todos foram considerados como pertencentes ao grupo “nocaute” para análises posteriores.

O grau de reprodutibilidade entre as bibliotecas, quando dados filtrados pela qualidade foram analisados por PCA (Análise de Componente Principal) (Figura 10) e *HeatMap* (Figura 11), mostrou-se adequado, ou seja, nenhuma amostra foi identificada como *outlier*. Dois grupos distintos foram identificados no plano vertical do gráfico do PCA, individualizando amostras nocautes e amostras WT. O mesmo resultado pode ser observado no *HeatMap*, no qual amostras nocautes e WT formam dois grupos distintos, confirmando a existência de réplicas técnicas.

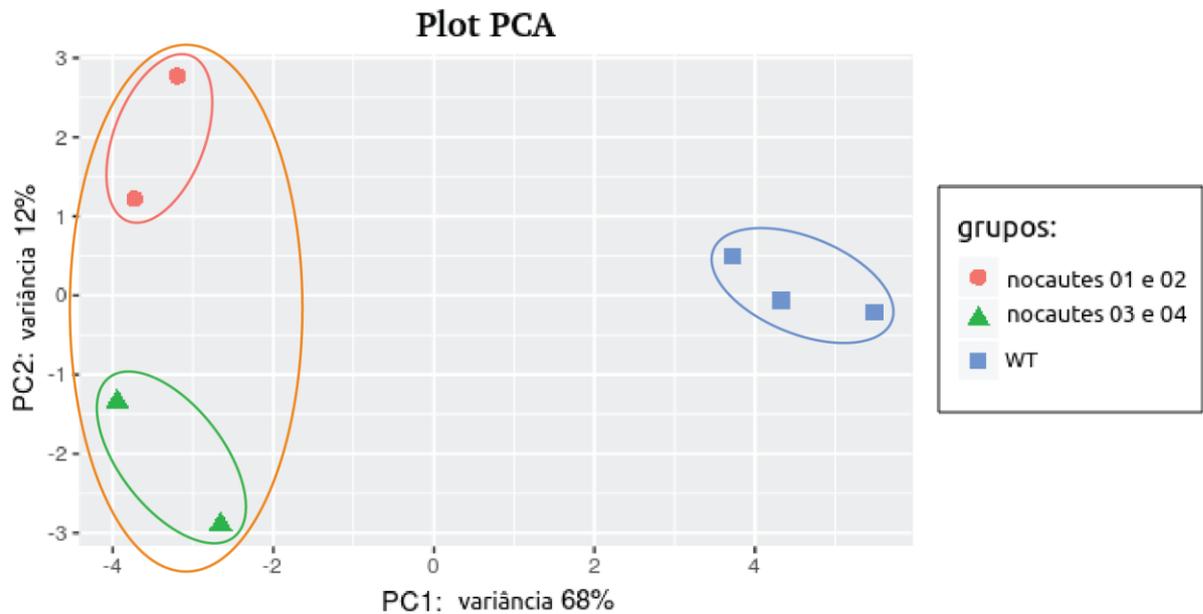


Figura 10. Análise de Componente Principal (PCA) das bibliotecas de cDNA. As elipses laranja e azul marcam os dois principais grupos da análise: nocautes para a TcRBP99 e WTs, respectivamente. As elipses rosa e verde marcam as duplicatas dos dois clones estudados: nocautes 01 e 02 para o primeiro clone e nocautes 02 e 03 para o segundo clone.

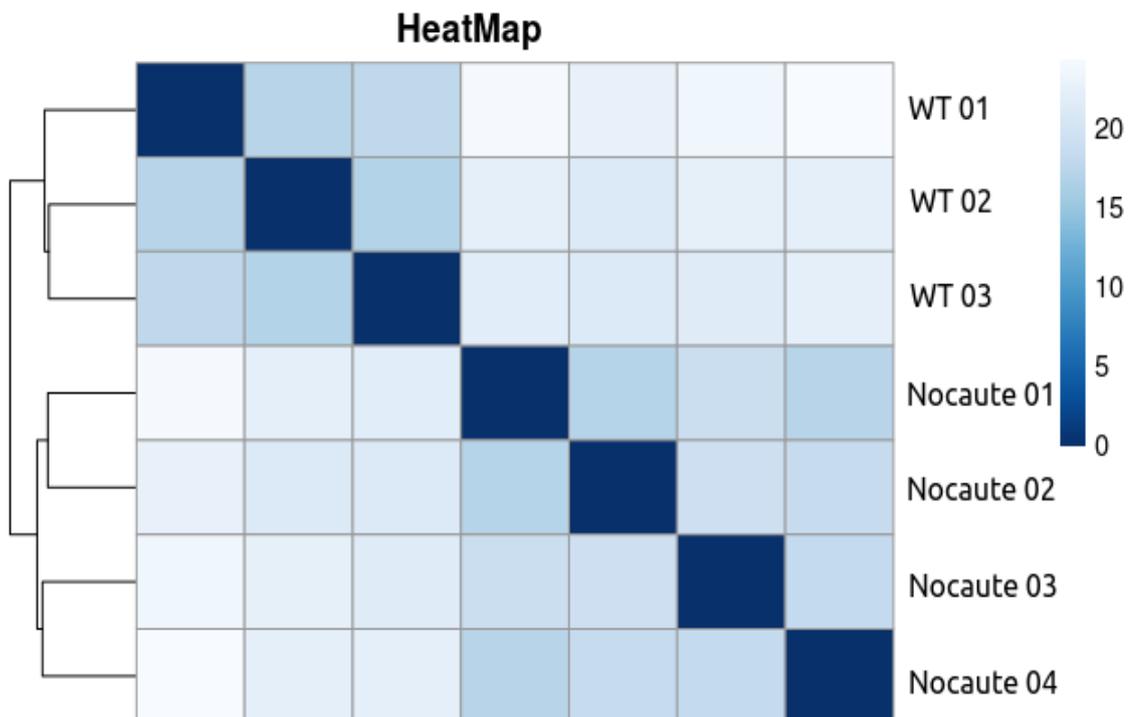


Figura 11. HeatMap das bibliotecas de cDNA. A proximidade estatística das amostras WT e nocautes é dada pela distância euclidiana existente entre as mesmas, o que é representado pela escala de azul. Método de clusterização hierárquico: ligação completa (*complete-linkage clustering*).

No total, foram obtidas sete bibliotecas com aproximadamente 3,6 milhões de *reads* cada, totalizando mais de 25 milhões. A qualidade das *reads* foi analisada antes e após a remoção de bases de baixa qualidade utilizando a ferramenta FastQC e respeitando o “*cut-off*” de qualidade média (*phred* \geq 20) para cada uma das bases. Sendo assim, um padrão de boa qualidade foi obtido para todas as bibliotecas brutas (Figura 12 e 13) e para as bibliotecas cujas sequências de baixa qualidade foram removidas (Figura 14 e 15).

A maioria das *reads* que passaram pelo filtro de qualidade, possuíam 250 pb de comprimento. Contudo, *reads* que possuem um comprimento entre 51 e 249 pb também foram conduzidas as análises subsequentes.

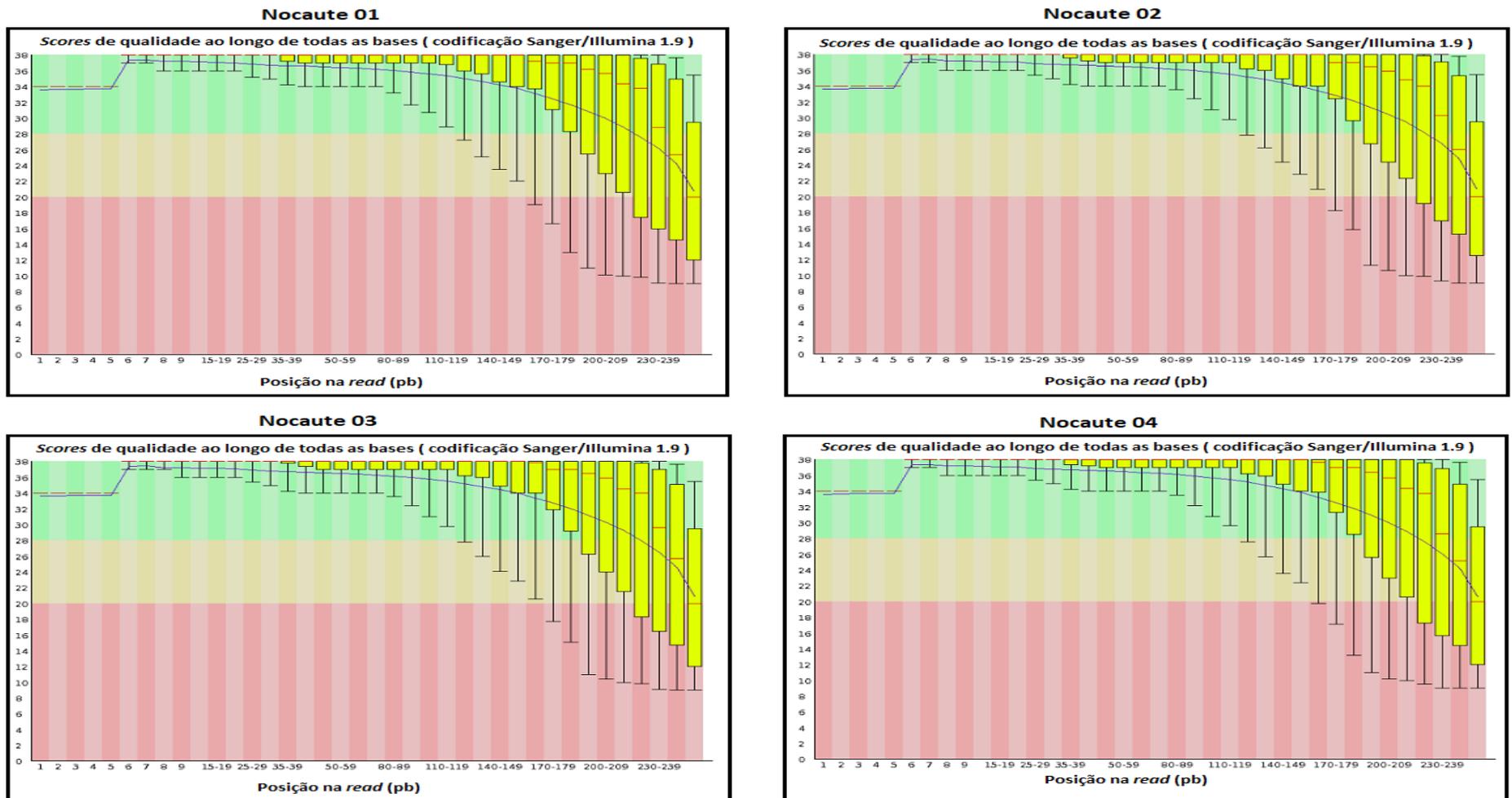


Figura 12. Gráficos de qualidade por base de reads brutas provenientes de bibliotecas de cDNA de epimastigotas nocautes para TcRBP99. Qualidade de reads (*forward*) por base baseada na pontuação de *phred score* (eixo y) para cada uma das quatro bibliotecas nocautes gerados. Ao final de todas as reads ocorre a queda da qualidade para níveis abaixo de 20.

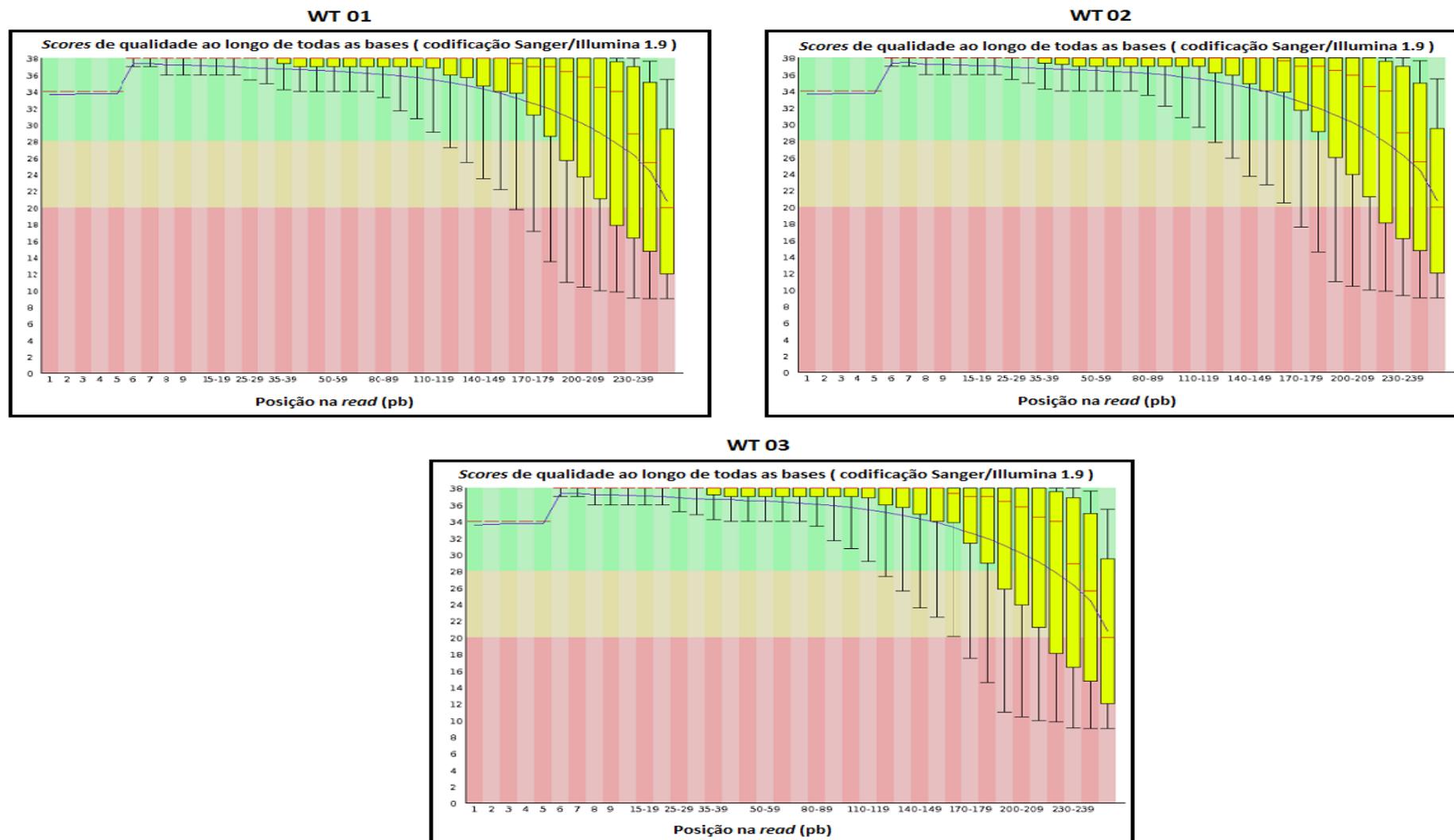


Figura 13. Gráficos de qualidade por base de reads brutas provenientes de bibliotecas de cDNA de epimastigotas WT. Qualidade de reads (*foward*) por base baseada na pontuação de *phred score* (eixo y) para cada uma das três bibliotecas WTs geradas. Ao final de todas as reads ocorre a queda da qualidade para níveis abaixo de 20.

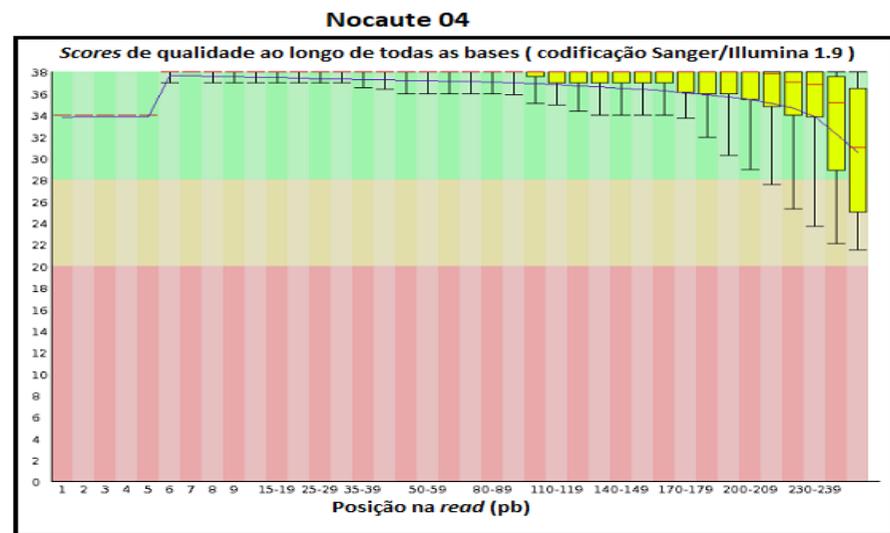
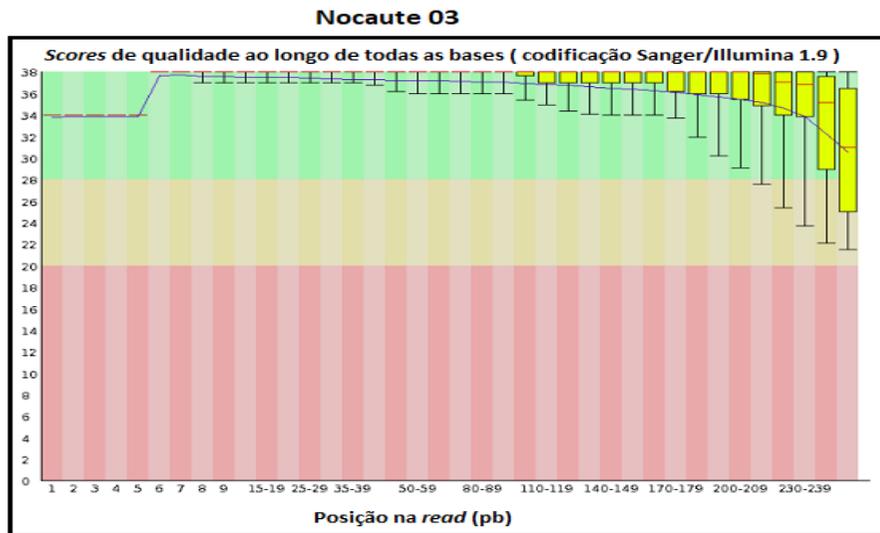
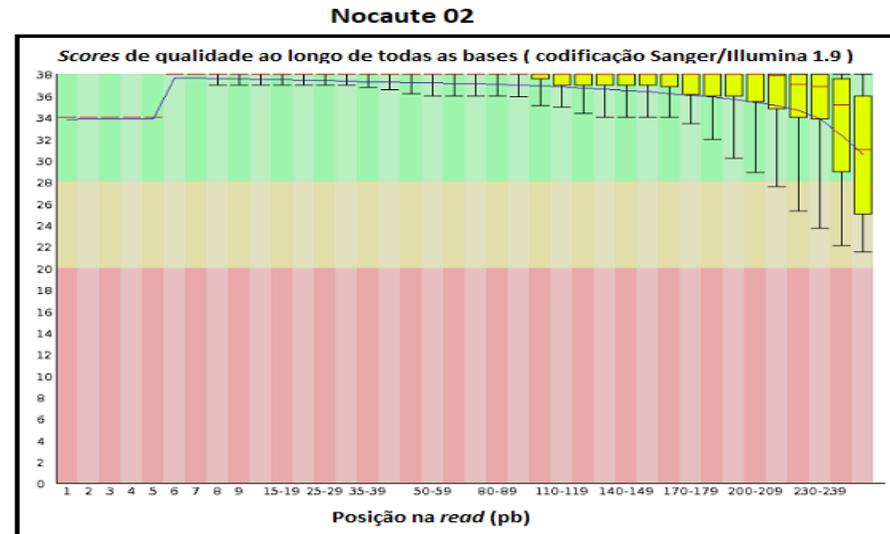
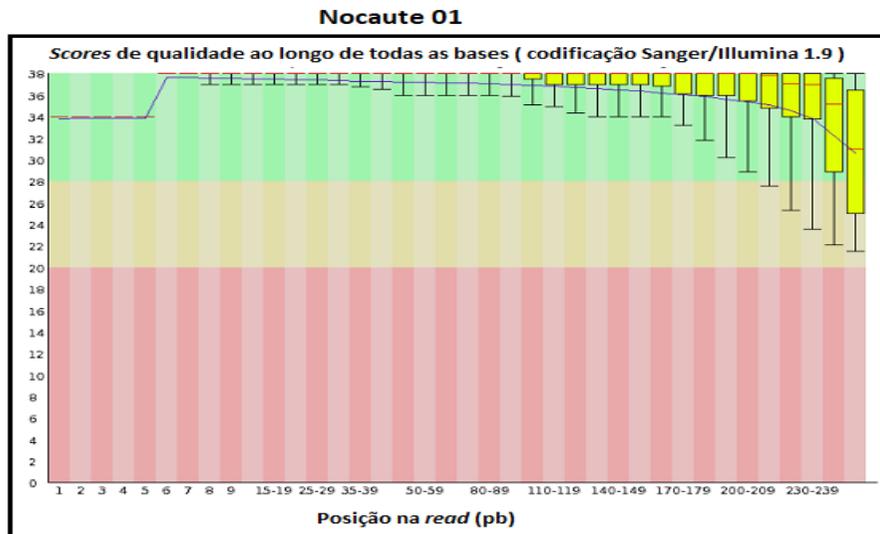


Figura 14. Gráficos de qualidade por base de reads trimadas provenientes de bibliotecas de cDNA de epimastigotas nocautes para TcRBP99. Qualidade de reads (*forward*) por base baseada na pontuação de *phred score* (eixo y) para cada uma das quatro bibliotecas nocautes gerados. As bases de baixa qualidade no final de cada read foram removidas.

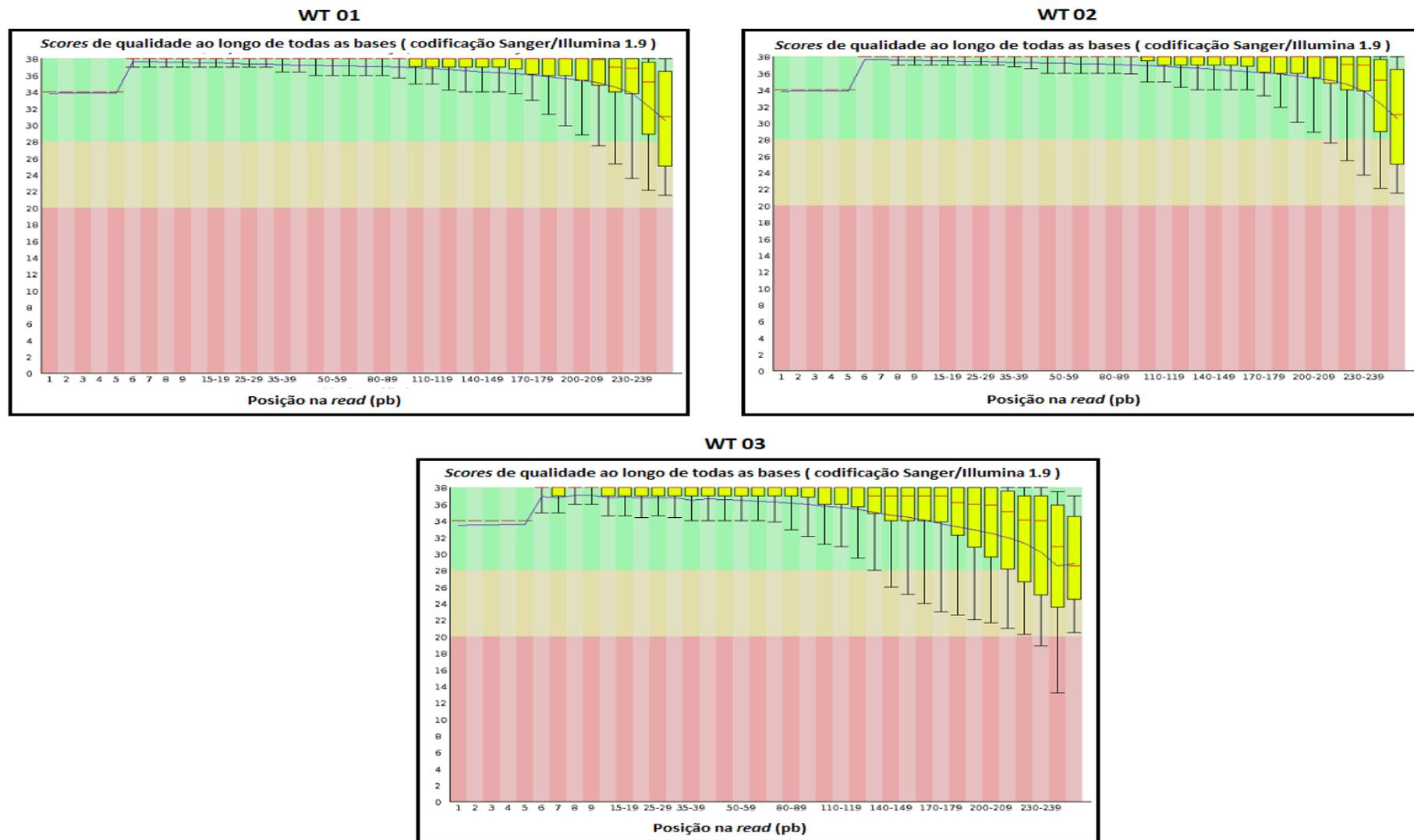


Figura 15. Gráficos de qualidade por base de reads trimadas provenientes de bibliotecas de cDNA de epimastigotas WT. Qualidade de reads (*foward*) por base baseada na pontuação de *phred score* (eixo y) para cada uma das três bibliotecas WTs geradas. As bases de baixa qualidade no final de cada *read* foram removidas.

5.3 Mapeamento e contagem de *reads* alinhadas ao genoma do clone CL Brener de *T. cruzi*

Ao mapear as *reads* de todas as bibliotecas no genoma anotado de *T. cruzi*, para os alelos *Esmeraldo-like* e *Non Esmeraldo-like* e *reads unassigned* (*reads* que não foram atribuídas a nenhum dos haplótipos) do clone CL Brener (www.tritrypdb), foi possível obter uma média de 73,28% de mapeamento, considerando apenas *reads paired-end* que passaram pelo filtro de qualidade (Tabela 3).

Tabela 3. Número total de *reads* mapeadas contra o genoma *Esmeraldo-like* e *Non Esmeraldo-like* de *T. cruzi*.

Biblioteca	Total de <i>reads</i>	Total de <i>reads</i> pós trimagem	<i>Reads</i> mapeadas	% <i>reads</i> mapeadas
Nocaute 01	3.513.887	2.741.084	1.927.421	70,2%
Nocaute 02	3.895.258	3.088.360	2.254.514	72,9%
Nocaute 03	4.143.756	3.274.477	2.394.984	73,1%
Nocaute 04	3.521.679	2.820.681	2.099.668	74,4%
WT 01	3.294.361	2.626.697	1.970.460	75,0%
WT 02	3.609.438	2.862.761	2.143.399	74,9%
WT 03	3.469.573	2.709.211	1.965.395	72,5%
Total	25.447.952	20.123.271	14.755.814	73,2%

A tabela acima apresenta também um resumo do número total de *reads* obtidas com o RNA-seq de epimastigotas nocautes para a TcRBP99 e epimastigotas WT e *reads* que passaram pelo filtro de qualidade (phred score > 20) (Total de *reads* pós trimagem).

A opção de mapeamento múltiplo permitiu que as *reads* pudessem ser mapeadas em mais de um local, porém, apenas o local em que foi obtido maior *score* foi contabilizado. Dessa forma, não houve duplicidade no mapeamento para nenhuma *read*.

Para todas as bibliotecas foi possível observar *reads* mapeando em regiões intercodificadoras e *reads* que não mapearam em nenhum dos haplótipos ou *reads*

unassigned (essas foram armazenadas no arquivo *unmapped.bam*), portanto, essas foram descartadas na etapa de contagem de *reads*.

Após o mapeamento foram contabilizadas as quantidades de *reads* mapeadas para cada um dos genes anotados no genoma, utilizando a ferramenta Htseq. Os resultados mostraram que cerca de 23,53% das *reads* mapearam em regiões não anotadas (UTRs e regiões intercodificadoras), ou seja, não mapeavam dentro de regiões codificadoras (CDS) relacionadas no arquivo GFF. Estas, portanto, não foram contabilizadas na contagem de nenhum dos genes, o que diminuiu a quantidade de *reads* consideradas na etapa de análise de expressão diferencial (Tabela 4).

Tabela 4. Contagem de *reads* mapeadas para cada gene do genoma *Esmeraldo-like* e *Non Esmeraldo-like* de *T. cruzi* de acordo com o arquivo GFF.

Biblioteca	Contagem de <i>reads</i> em CDS	% contagem de <i>reads</i> em CDS
Nocaute 01	1.375.561	50,18%
Nocaute 02	1.671.020	54,10%
Nocaute 03	1.769.604	53,78%
Nocaute 04	1.555.930	55,16%
WT 01	1.443.438	54,95%
WT 02	1.587.946	55,46%
WT 03	1.437.038	53,04%
Total	10.840.537	53,81%

Foi construída uma tabela de contagens brutas contendo o número de *reads* mapeadas para cada gene (25.100 genes para os dois haplótipos), para as sete bibliotecas (Tabela S1, disponível em: <https://www.dropbox.com/s/f5tumqz1vwulu4z/Tabela%20S1%20-%20Contagens_%20nocaute%20x%20WT.xlsm?dl=0>). Essa tabela foi então utilizada para análise de expressão diferencial entre epimastigotas nocautes e WT, após passar por uma filtragem para baixas contagens.

As contagens brutas foram transformadas em unidades de expressão FPKM, o que possibilitou a construção de uma tabela contendo tais valores para cada gene,

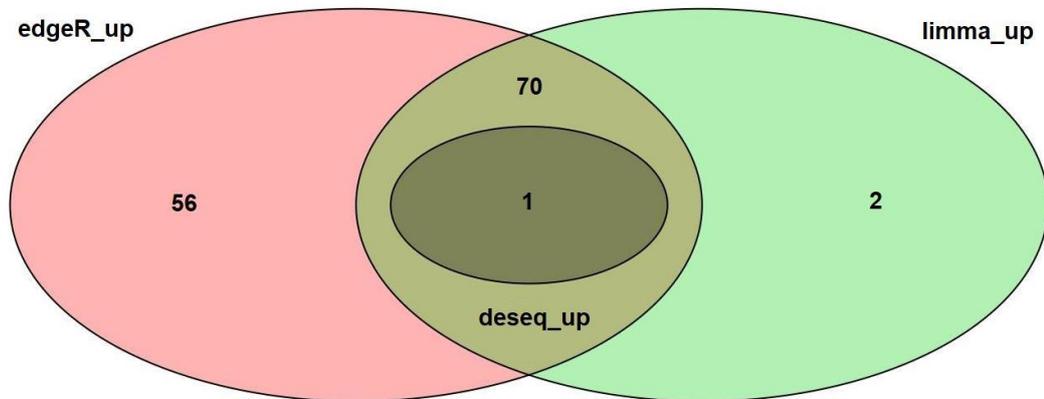
para as sete bibliotecas (Tabela S2, disponível em: <https://www.dropbox.com/s/ueywt3aormea3z0/Tabela%20S2%20-%20Fpkm_nocautes%20x%20WT.xlsm?dl=0>). Para melhor visualização das diferenças de expressão estes valores foram transformados em log-FPKM (Tabela S3, disponível em: https://www.dropbox.com/s/frs0shis1j0inc5/Tabela%20S3%20-%20logFPKM_nocautes%20x%20WT.xlsm?dl=0). A partir desses dados foi possível verificar o nível de expressão de todos os genes de epimastigotas WT e epimastigotas nocautes para a TcRBP99 de *T. cruzi*. Análises de expressão gênica baseadas em valores FPKM permitiram a comparação de diferentes genes de uma mesma amostra, ao contrário de contagens brutas que não levam em consideração o tamanho da biblioteca e o comprimento dos diferentes genes.

5.4 Análise de expressão diferencial, identificação de possíveis mRNA alvos da TcRBP99 e comparação de pacotes utilizados

A princípio a análise de expressão diferencial foi conduzida utilizando três pacotes de análise estatística: EdgeR, limma e Deseq2. Cada pacote identificou uma quantidade diferente de genes tanto para genes com expressão aumentada, quanto para genes com expressão diminuída. Essa comparação possibilitou a construção de um diagrama de Venn (Figura 16), mostrando que todos os genes identificados pelo Deseq2 foram também identificados nos outros pacotes. Esse resultado juntamente com o fato do Deseq2 ser mais conservador em relação aos demais, nos levou a considerar somente os 10 genes identificados pelo mesmo nas análises subsequentes.

A

Genes com expressão aumentada

**B**

Genes com expressão diminuída

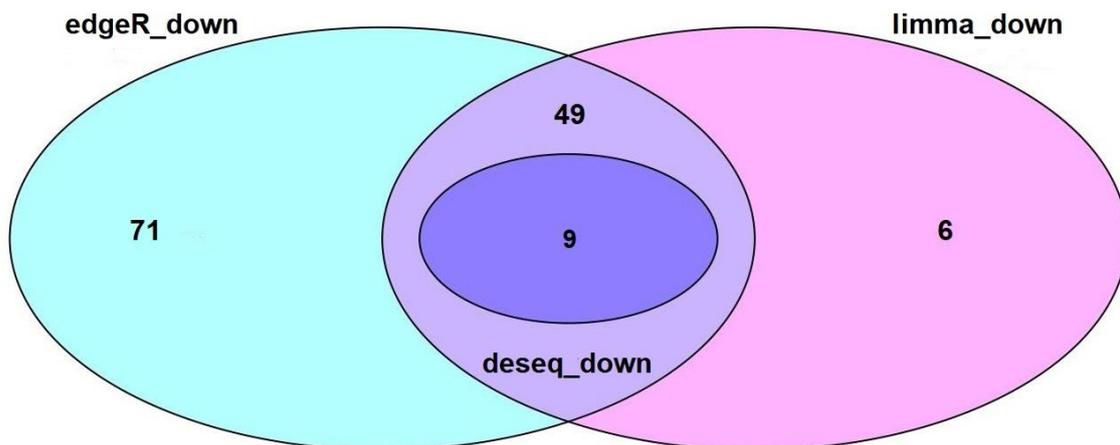


Figura 16. Diagrama de Venn - Quantidade de genes diferencialmente expressos identificados com os pacotes EdgeR, limma e Deseq2. A - Genes com expressão aumentada em epimastigotas nocautes comparadas a epimastigotas WT. Um gene identificado para o pacote Deseq, 56 exclusivamente identificados pelo edgeR e 2 pelo limma. Outros 70 genes foram compartilhados pelos pacotes edgeR e limma. B - Genes com expressão diminuída em epimastigotas nocaute comparadas a epimastigotas WT. Nove genes identificados para o pacote Deseq, 71 exclusivamente identificados pelo edgeR e 6 pelo limma. Outros 49 genes foram compartilhados pelos pacotes edgeR e limma.

A análise de expressão diferencial utilizando o pacote DESeq2, que assume distribuição binomial negativa para a contagem de *reads* e normalização dos “fatores de tamanhos”, leva em consideração a profundidade de sequenciamento de todas as amostras. Dessa forma e segundo parâmetros estabelecidos ($p_{adj} < 0,05$ e $-1,1 < \text{Log}_2\text{FoldChange} > 1,1$) 10 genes foram identificados como tendo expressão diferencial (Figura 17 e 18), sendo apenas 1 com expressão aumentada em

epimastigotas nocautes em comparação a epimastigotas WT (Tabela 5) e 9 com expressão diminuída considerando a mesma comparação (Tabela 6).

MA Plot – Genes com expressão aumentada/diminuída Nocaute X WT

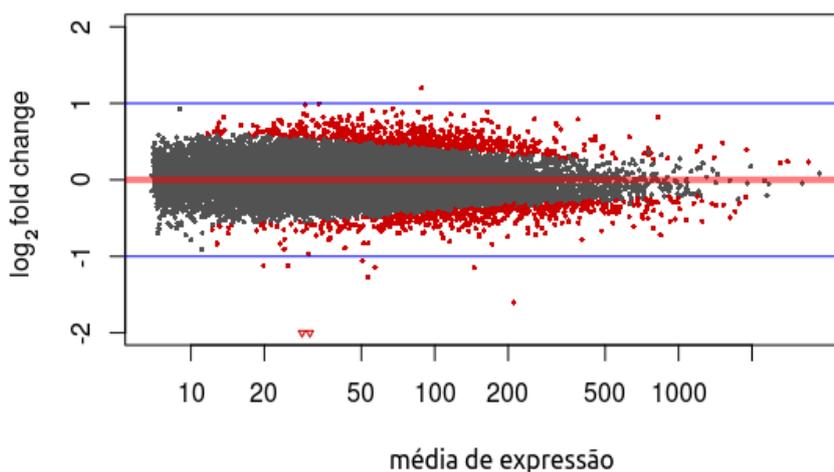


Figura 17. Níveis de expressão gênica em *T. cruzi*. Resultado da análise de expressão diferencial entre epimastigotas nocautes para TcRBP99 e epimastigotas WT, considerando $-1,1 > \log_2\text{FoldChange} > 1,1$ e $\text{padj} < 0,05$. Os pontos vermelhos, situados entre os limites das linhas azuis ($-1,1 < \log_2\text{FoldChange} < 1,1$) representam todos os genes com expressão diferencial ($\text{padj} < 0,05$), mas com mudança de expressão menor que 2 vezes. O ponto vermelho acima (1) e abaixo (9) dos limites das linhas azuis representam genes com $\text{padj} < 0,05$ e $-1,1 < \log_2\text{FoldChange} > 1,1$ respectivamente, ou seja, mudança de expressão maior que 2 vezes em epimastigotas nocautes comparadas com epimastigotas WT. Os triângulos vermelhos representam genes com os maiores valores de mudança de expressão e se encaixam no grupo de genes que apresentam expressão diminuída em mais de 2 vezes.

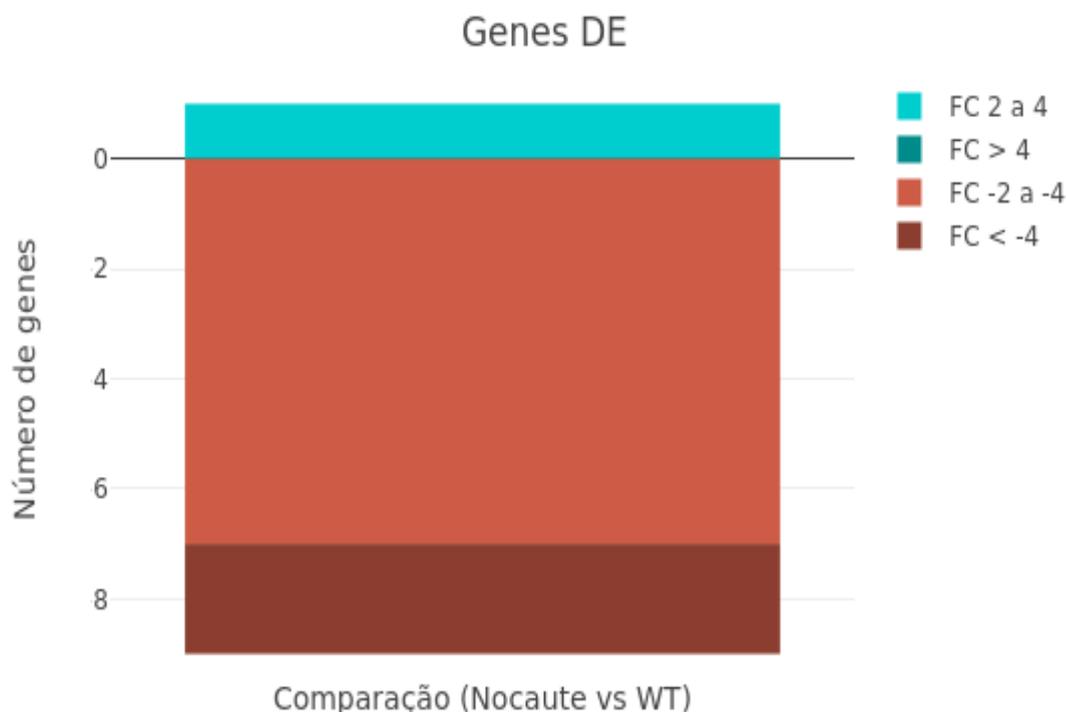


Figura 18. Genes diferencialmente expressos em epimastigotas nocautes comparadas com epimastigotas WT, segundo parâmetros de padj e LogFoldChange (FC) pré-estabelecidos. A cor verde-água representa o gene identificado com regulação positiva na ausência da TcRPB39, com logFoldChange entre 2 e 4. As cores marrom claro e marrom escuro representam a quantidade de genes com regulação negativa também na ausência da TcRPB39, com LogFC entre -2 e -4 e menor que -4, respectivamente.

Tabela 5. Genes com expressão aumentada em epimastigotas nocaute para TcRBP99 comparado com epimastigotas *wild-type*.

ID	Anotação	log2FC	pvalue	padj
TcCLB.510131.40	Hidrolase haloácido dealogenase, putativa	1,19	2,95e-13	5,59e-10

Como somente um gene foi identificado como tendo expressão aumentada na ausência da TcRBP99 e essa diferença de expressão foi muito próxima do “*cut-off*” estabelecido (Log2FC = 1.19), decidimos não investigar esse gene no momento. Futuramente este gene será validado experimentalmente por meio de qRT-PCR para confirmação se ocorreu o aumento da sua expressão nos nocautes.

Tabela 6. Genes com expressão diminuída em epimastigotas nocaute para TcRBP99 comparado com epimastigotas *wild-type*.

ID	Anotação	log2FC	pvalue	padj
TcCLB.509559.20	Proteína retrotransposon hot spot (RHS, pseudogene), putativa	-2,86	6,98e-40	1,05e-35
TcCLB.505943.110	Proteína retrotransposon hot spot (RHS, pseudogene), putativa	-2,74	3,00e-36	1,52e-32
TcCLB.503733.80	Proteína hipotética, conservada	-1,50	3,74e-23	1,41e-19
TcCLB.506153.10	Transportador de aminoácidos, putative	-1,26	2,20e-11	3,03e-08
TcCLB.506551.10	Proteína associada à diferenciação 4, putative	-1,14	4,00e-18	1,01e-14
TcCLB.504173.60	Proteína hipotética, conservada	-1,14	1,02e-10	1,04e-07
TcCLB.504173.40	Proteína hipotética, conservada	-1,12	1,08e-07	3,64e-05
TcCLB.506825.40	Proteína hipotética	-1,12	2,19e-07	6,19e-05
TcCLB.507659.10	Proteína hipotética, conservada	-1,11	7,34e-09	4,12e-06

As linhas da tabela marcadas com retângulos vermelhos indicam membros da família multigênica das RHS (do inglês *retrotransposon hot spot*), que juntamente com às proteínas marcadas com retângulos azul (transportador de aminoácido, TA) e verde (proteína associada à diferenciação, PAD) se destacam por serem genes com anotações que indicam produtos de tradução conhecidos, diferentemente dos genes que codificam para proteínas hipotéticas.

A proteína TcRBP99 possivelmente regula outros genes em epimastigotas, ainda que em menor proporção (LogFoldChange (LogFC) entre 0 e 2 e entre 0 e -2). Na Figura 19 é possível comparar o número de genes que apresentaram diferenças de expressão estatisticamente significativas ($padj < 0,05$), independentemente do valor LogFC.

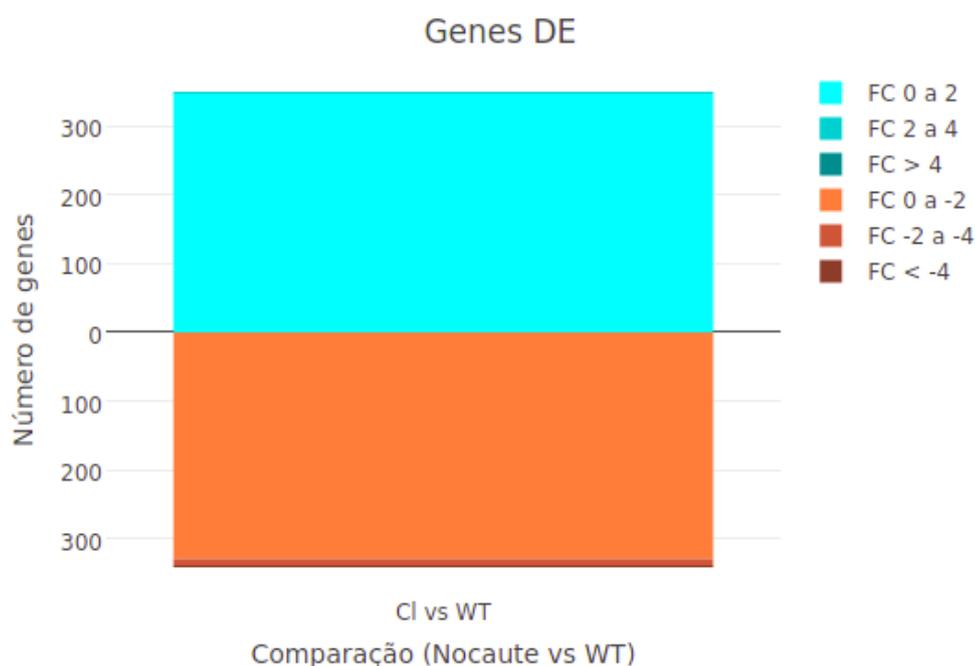


Figura 19. Genes diferencialmente expressos em epimastigotas nocautes comparadas com epimastigotas WT segundo parâmetros estatisticamente significativos ($p_{adj} < 0,05$). As cores azul-claro e verde-água representam a quantidade de genes identificados com regulação positiva na ausência da TcRBP39, com FC entre 0 e 2 (não passaram pelo filtro estabelecido neste trabalho, $FC > 2$) e FC entre 2 e 4, respectivamente. As cores laranja, marrom claro e marrom escuro representam a quantidade de genes com regulação negativa também na ausência da TcRBP39, com FoldChange entre 0 e -2, entre -2 e -4 e menor que -4, respectivamente.

Novas buscas realizadas no banco de dados do transcriptoma de *T. cruzi* (BELLEW *et al.*, 2017), mostraram que os potenciais alvos identificados da TcRBP99, codificadores da PAD, transportador de aminoácido e RHS, são transcritos com níveis aumentados em epimastigotas WT comparadas às demais formas de vida do parasita (Figura 20). O gene codificador de PAD apresenta expressão 12,77 vezes maior comparado com tripomastigotas; para o gene que codifica transportador de aminoácidos essa diferença é de 2,09 vezes. Em relação aos genes que codificam para as RHS, a diferença de expressão em relação a tripomastigotas é de 1,96 vezes e 2,16 vezes, respectivamente.

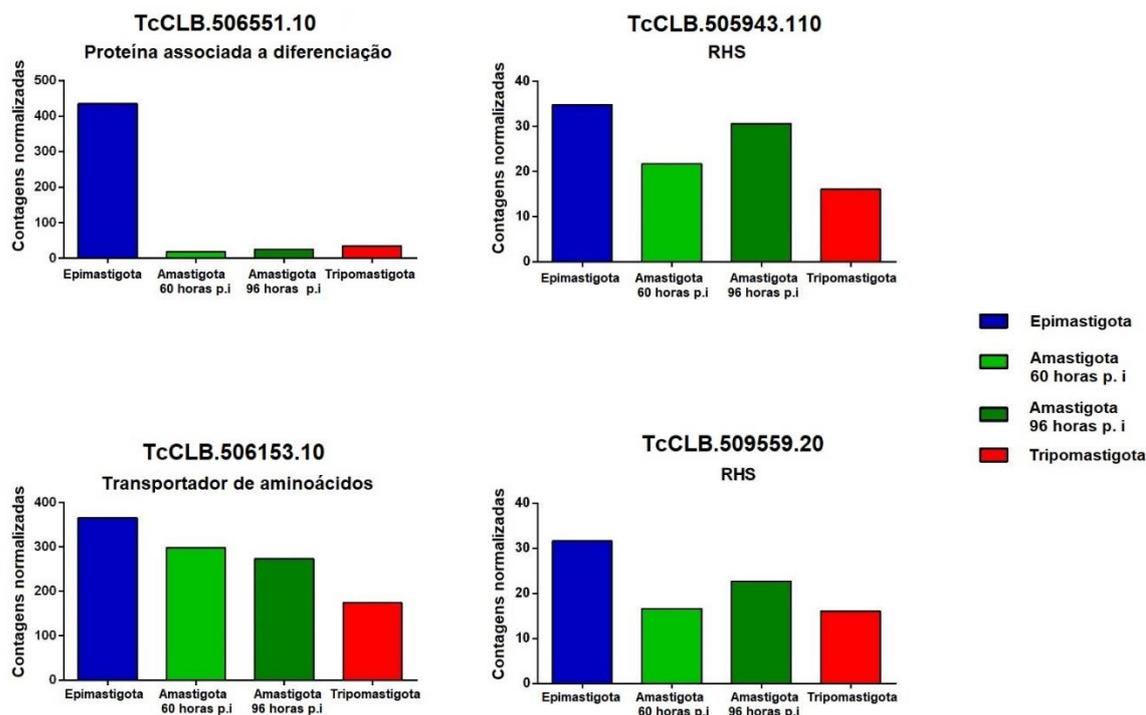


Figura 20. Níveis de expressão de possíveis genes alvos da TcRBP99 ao longo do ciclo de vida *T. cruzi*. A expressão dos genes nas formas epimastigota, amastigota 60 e 96 horas pós-infecção de fibroblastos humanos *in vitro* e tripomastigotas liberadas no sobrenadante dessas infecções é mostrada nos gráficos com base em dados de expressão previamente publicados (BELEW et al., 2017).

Buscas *in silico* mostraram que um dos genes identificados como tendo expressão diminuída na ausência da TcRBP99 é o gene que codifica um transportador de aminoácidos (TcCLB.506153.10). As buscas na base de dados mostraram que este gene pertence a uma família de 17 membros e pertence ao haplótipo *Non-Esmeraldo-like* (o alelo *Esmeraldo-like* tem o ID: TcCLB.511411.30). Seu tamanho consiste em 1344 nucleotídeos, os quais codificam uma proteína de 448 aminoácidos. Além disso, este gene possui ortólogos em outros tripanossomatídeos dos gêneros *Trypanosoma*, *Leishmania* e *Leptomonas*. Dados de RNA-seq de *T. cruzi*, cepa Dm28c, depositados na base de dados indicam que a maior expressão desse gene ocorre em epimastigotas (SMIRCICH et al., 2015). O mesmo é visto para a cepa Y em um experimento realizado a partir de células humanas infectadas, onde os níveis de mRNA para o gene TcCLB.506153.10 foram em média 18 vezes maior em epimastigotas, comparado a amastigotas e tripomastigotas (LI et al., 2016).

Heatmaps gerados a partir das contagens filtradas utilizadas neste trabalho (Figura 20) e a partir dos dados de RNA-seq de BELEW et al. (2017) (Figura 22)

mostraram que 11 genes da família dos transportadores de aminoácidos (IDs: TcCLB.506053.10, TcCLB.507659.30, TcCLB.508017.30, TcCLB.507659.20, TcCLB.506153.10, TcCLB.506153.10, TcCLB.503837.20, TcCLB.506227.10, TcCLB.506053.5, TcCLB.507007.108, TcCLB.511543.30) estão aumentados em epimastigotas de *T.cruzi* em comparação às outras formas de vida. Além disso, é possível notar a diminuição do gene TcCLB.506153.10 em epimastigotas nocautes na Figura 21.

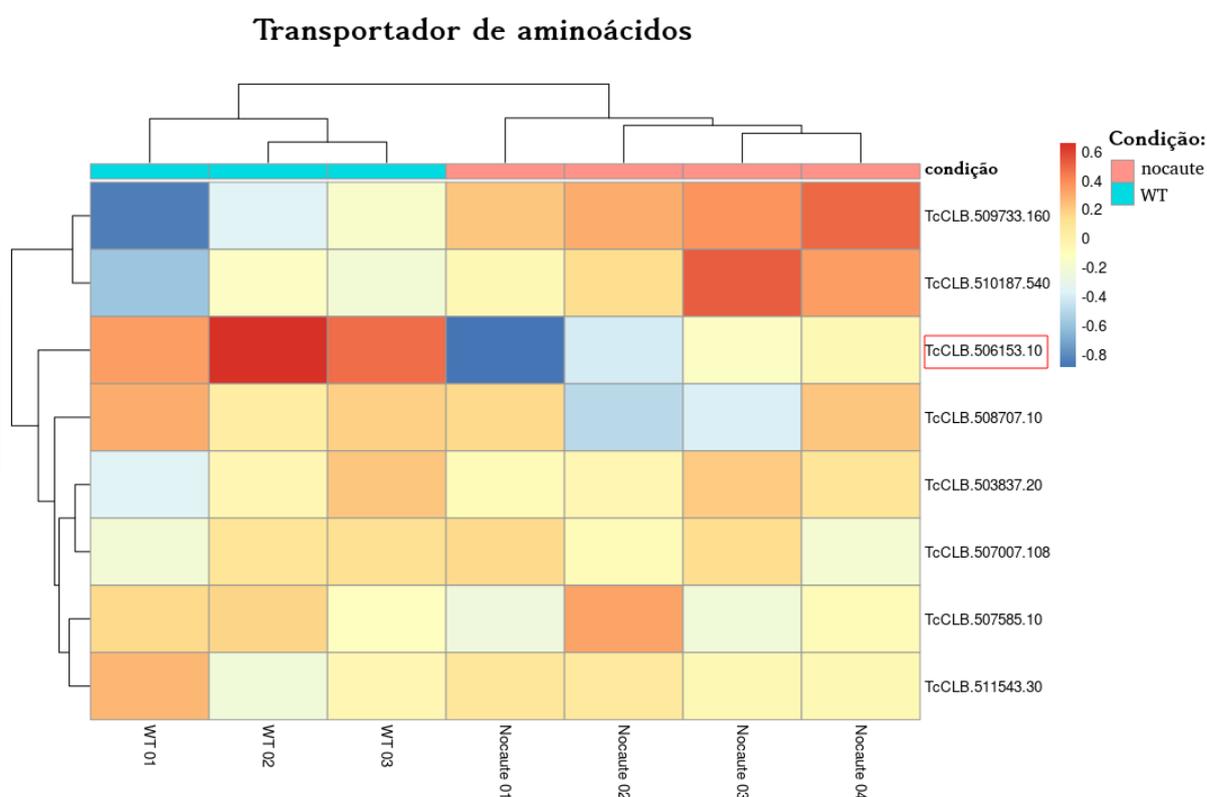


Figura 21. HeatMap de 8 membros da família multigênicas dos transportadores de aminoácidos gerado a partir de dados de expressão de epimastigotas WT e nocautes. Apenas os genes representados passaram pelo filtro de baixas contagens a partir dos dados de RNA-seq obtidos neste trabalho. O retângulo vermelho representa o gene diferencialmente expressos com $p_{adj} < 0.05$ e $-1.1 < \text{Log}_2\text{FC} > 1.1$. A escala de cores varia entre azul e vermelho, representando respectivamente menor e maior expressão dos genes. O método utilizado para o agrupamento dos genes e amostras foi baseado na distância euclidiana eles. Método de clusterização hierárquico: ligação completa (*complete-linkage clustering*).

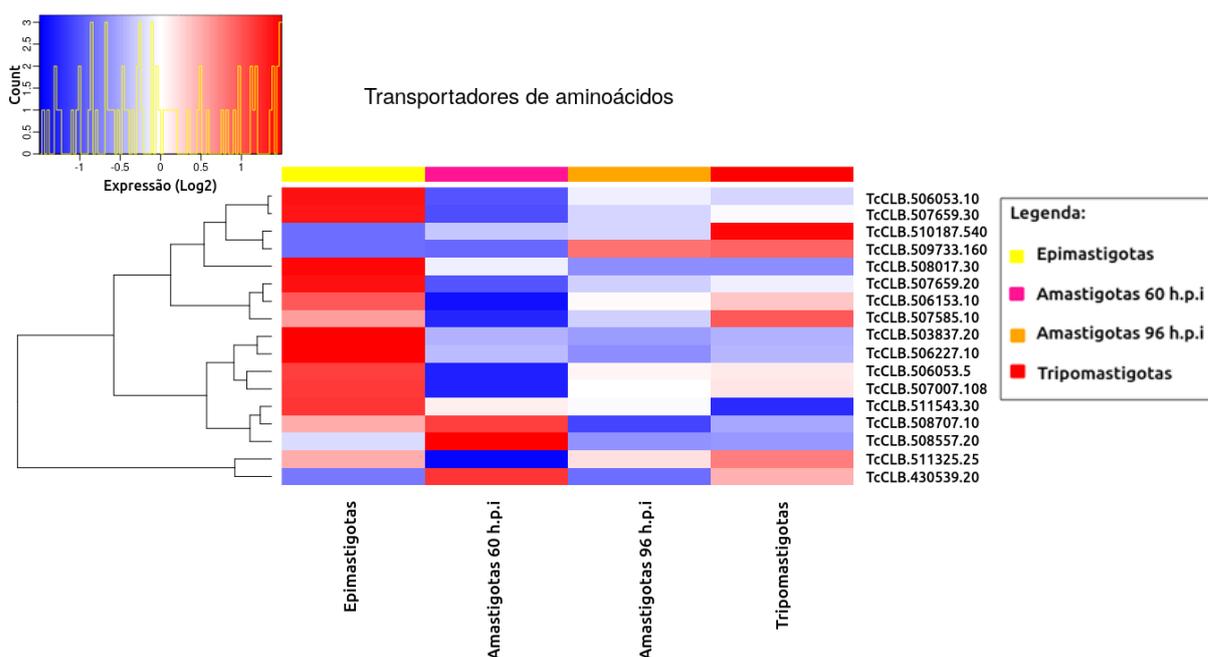


Figura 22. HeatMap dos 17 membros da família multigênica dos transportadores de aminoácidos gerado a partir de dados de expressão das formas epimastigotas, amastigotas coletadas 60 e 96 horas após a infecção e tripomastigotas (BELEW et al., 2017). A escala de cores varia entre azul e vermelho, representando respectivamente menor e maior expressão dos genes. O método utilizado para o agrupamento dos genes foi baseado na distância euclidiana. Método de clusterização hierárquico: ligação completa (*complete-linkage clustering*).

Outro gene, o qual está anotado como codificador da proteína associada à diferenciação (TcCLB.506551.10), também parece ser afetado pela presença ou ausência da TcRBP99, uma vez que na ausência da mesma a análise de expressão diferencial constatou diminuição nos níveis de seu mRNA. Pertence a uma família multigênica de 13 membros e está localizada no haplótipo *Esmeraldo-like*. Além disso, este gene possui 1851 nucleotídeos e codifica uma proteína de 617 aminoácidos. Genes ortólogos estão presentes nos demais tripanossomatídeos, como por exemplo *T. brucei* e espécies do gênero *Leishmania*. Dados de *microarray* e RNA-seq também depositados no TritypDB mostraram que os níveis de mRNA deste gene encontram-se aumentados em epimastigotas das cepas Dm28c e Y (LI et al., 2016; MINNING et al., 2009; SMIRCICH et al., 2015).

HeatMaps realizados para a família das PADs de *T. cruzi* mostraram que os 13 genes dessa família possuem expressão aumentada em epimastigotas comparadas as demais formas de vida do parasito (Figura 24). Além disso, na Figura 23 é observada mais uma vez a expressão diminuída do gene TcCLB.506551.10 em epimastigotas nocautes comparadas a epimastigotas WT.

Proteína associada à diferenciação

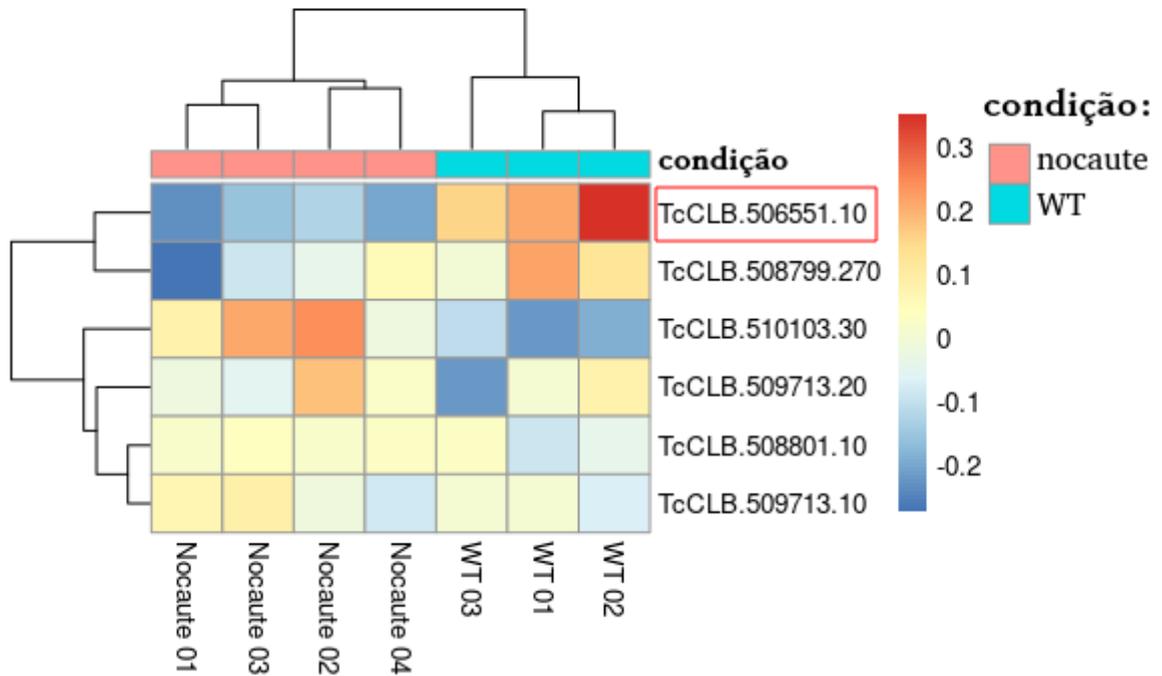


Figura 23. HeatMap de 6 membros da família multigênicas das proteínas associadas à diferenciação gerado a partir de dados de expressão epimastigotas WT e nocautes. Apenas os genes representados passaram pelo filtro de baixas contagens a partir dos dados de RNA-seq obtidos neste trabalho. O retângulo vermelho representa o gene diferencialmente expressos com $p_{adj} < 0.05$ e $-1.1 < \text{Log}_2\text{FC} > 1.1$. A escala de cores varia entre azul e vermelho, representando respectivamente, menor e maior expressão dos genes. O método utilizado para o agrupamento dos genes e amostras foi baseado na distância euclidiana. Método de clusterização hierárquico: ligação completa (*complete-linkage clustering*).

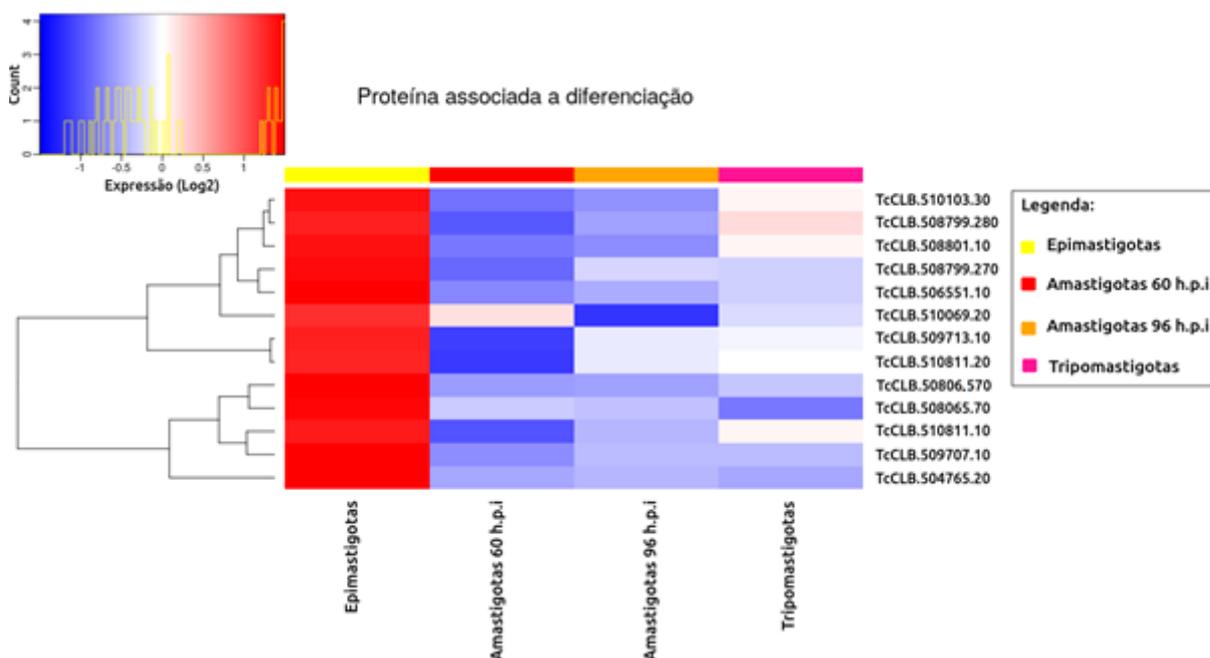


Figura 24. HeatMap dos 13 membros da família multigênica das proteínas associadas à diferenciação gerado a partir de dados de expressão das formas epimastigotas, amastigotas coletadas 60 e 96 horas após a infecção e tripomastigotas (BELEW et al., 2017). A escala de cores varia entre azul e vermelho, representando respectivamente menor e maior expressão dos genes. O método utilizado para o agrupamento dos genes foi baseado na distância euclidiana. Método de clusterização hierárquico: ligação completa (complete-linkage clustering).

O gene TcCLB.509559.20, pertencente a família das RHS e identificado como tendo expressão diminuída na ausência da TcRPB99, é na verdade um pseudogene com 2832 pb, pois apresenta 15 *stop códons* dentro da CDS predita. A tradução de seu respectivo mRNA resulta em uma sequência de 929 aminoácidos e não de 944 aminoácidos como seria esperado. Contudo, o primeiro *stop códon* aparece somente nas posições que vão do nucleotídeo 2.137 ao nucleotídeo 2.139. Sendo assim, é possível que ocorra a síntese de uma proteína funcional com 712 aminoácidos (sequência que vai desde a metionina até o primeiro *stop códon*).

Os resultados da análise de expressão diferencial indicam que o alelo identificado no haplótipo *Esmeraldo-like* (TcCLB.505943.110) também apresentou expressão diminuída em parasitas nocautes para a TcRBP99. Este gene está também anotado na base de dados TritypDB como pseudogene, devido à presença de 5 *stop códons* presentes na CDS predita, a qual é composta de 2814 pb. Uma proteína de 938 aminoácidos poderia ser esperada, contudo, devido à presença do primeiro *stop códon*, nas posições que vão do nucleotídeo 2.137 ao nucleotídeo 2.139, é possível

que uma proteína funcional de no máximo 712 aminoácidos seja sintetizada, assim como ocorre para o alelo descrito anteriormente.

HeatMaps realizados a partir das contagens filtradas (Figura 25), comparando epimastigotas WT e epimastigotas nocautes, e a partir de dados de RNA-Seq das formas epimastigotas, amastigotas coletadas 60 e 96 horas após a infecção e tripomastigotas (BELEW et al., 2017) (Figura 25), mostraram que a maior parte dos membros da família das RHS (> 80%) têm expressão aumentada em epimastigotas. Na Figura 26 também é possível observar que a expressão dos genes TcCLB.505943.110 e TcCLB.509559.20 realmente é menor em condições nocautes para a TcRBP99 comparadas com epimastigotas WT, e que ao contrário dos demais genes da família das RHS, esses dois genes apresentam grande diferença de expressão nas condições analisadas, o que pode ser constatado pelo contraste de cores do *Heatmap* (azul – menor expressão e vermelho – maior expressão).

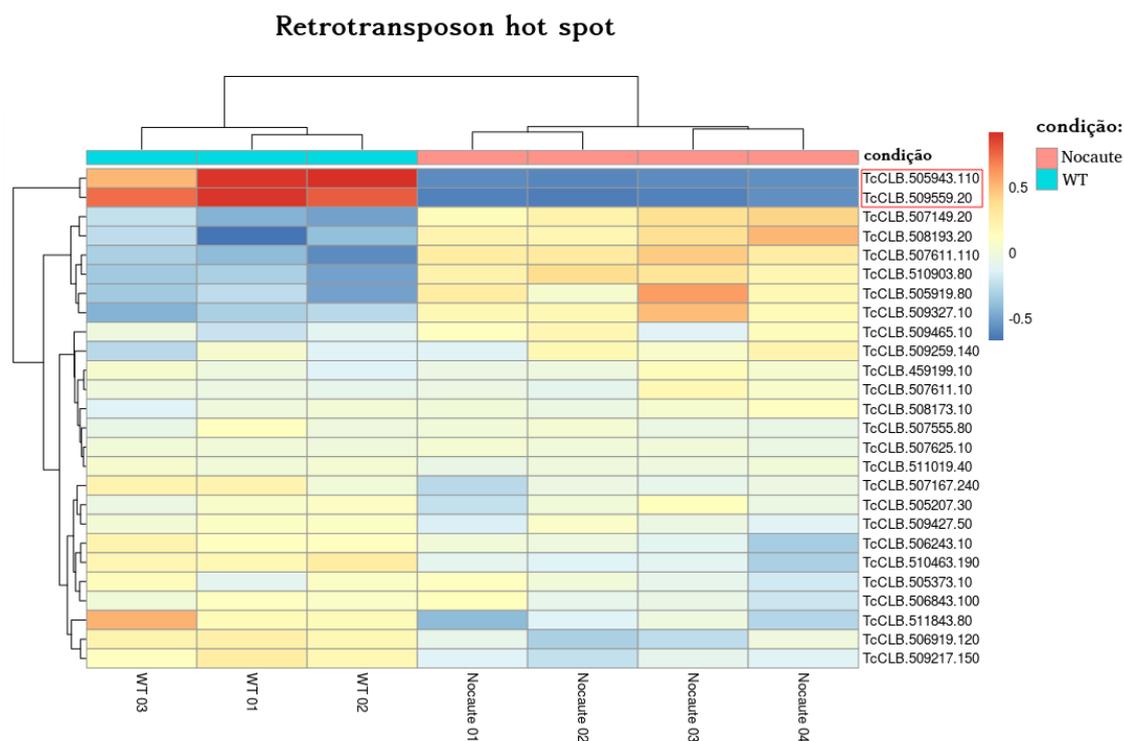


Figura 25. HeatMap de 26 membros da família multigênicas das RHS gerado a partir de dados de expressão de epimastigotas WT e nocautes. Apenas os genes representados passaram pelo filtro de baixas contagens a partir dos dados de RNA-seq obtidos neste trabalho. O retângulo vermelho representa os dois genes diferencialmente expressos com $p_{adj} < 0.05$ e $-1.1 < \text{Log}_2\text{FC} > 1.1$. A escala de cores varia entre azul e vermelho, representando respectivamente menor e maior expressão dos genes. O método utilizado para o agrupamento dos genes e amostras foi baseado na distância euclidiana. Método de clusterização hierárquica: ligação completa (*complete-linkage clustering*).

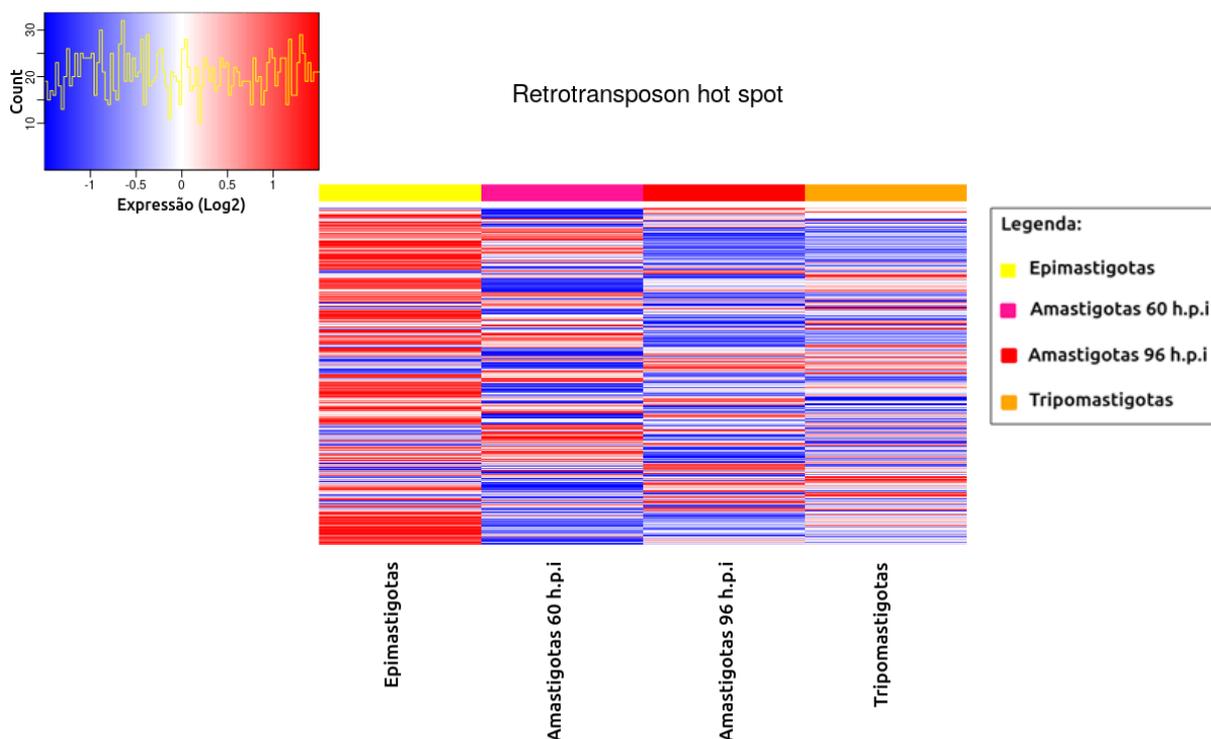


Figura 26. *HeatMap* dos 524 membros da família multigênica das RHS gerado a partir de dados de expressão das formas epimastigotas, amastigotas coletadas 60 e 96 horas após a infecção e tripomastigotas (BELEW et al., 2017). A escala de cores varia entre azul e vermelho, representando respectivamente, menor e maior expressão dos genes. O método utilizado para o agrupamento dos genes foi baseado na distância euclidiana. Método de clusterização hierárquico: ligação completa (*complete-linkage clustering*).

Os demais genes identificados como tendo expressão possivelmente afetada pela TcRBP99 estão anotados como proteínas hipotéticas (TcCLB.503733.80, TcCBL.404173.60, TcCLB.504173.40, TcCLB.506825.40 e TcCLB.506825.40) e não foram foco deste trabalho, porém, análises de validação dos níveis de expressão dos mesmos serão futuramente realizadas por meio de RT-PCR, assim como para os mRNAs dos genes PAD, transportador de aminoácidos e RHS.

Mas, de forma geral, por meio de análises de expressão e buscas *in silico* realizadas na base de dados TritypDB foi possível concluir que 10 genes possivelmente são regulados pela ligação da TcRBP99 aos seus respectivos mRNAs, nas formas epimastigotas. Destes, apenas 1 gene apareceu como tendo expressão aumentada em parasitas nocautes para essa RBP, enquanto 9 genes foram identificados como tendo expressão diminuída, indicando que de forma direta ou indireta a TcRBP99 influencia na estabilização desses mRNAs em epimastigotas WT.

Porém, novas análises são necessárias para a validação dos resultados obtidos e para o melhor entendimento do papel da TcRBP99 na regulação desses mRNAs, bem como do papel desses alvos nos diversos processos biológicos que envolvem a sobrevivência de epimastigotas e diferenciação em tripomastigotas metacíclicas.

5.5 Determinação das regiões 3' UTRs de mRNAs possíveis alvos da TcRBP99

Após a identificação de genes possivelmente regulados pela proteína TcRBP99 e a possível correlação dos mesmos com processos sabidamente importantes em formas epimastigotas, como a metaciclogênese e transporte de aminoácidos (BRINGAUD et al. 2006; SILBER et al., 2005; VALENTE et al., manuscrito em preparação), buscamos identificar os limites das regiões 3' UTRs dos mesmos. Além de essa ser uma informação que não consta no banco de dados de tripanossomatídeos, a posterior identificação de elementos regulatórios presentes nessas regiões, responsáveis pela ligação da RBP e pelos efeitos de mudança dos níveis equilíbrio é crucial para o entendimento dos mecanismos de regulação pós-transcricional (ARAÚJO et al., 2011; BARTHOLOMEU et al., 2002).

Sequências das bibliotecas do transcriptoma de epimastigotas e amastigotas, obtidas em colaboração com o Prof. Nagib El-Sayed da Universidade de Maryland, e mapeadas contra o genoma do clone CL Brener com níveis elevados de cobertura (em média, 103 vezes) foram utilizadas para conduzir esta análise de determinação do tamanho das 3' UTRs. Além disso, o remapeamento de *reads unmapped* que passaram por uma etapa de remoção de sequências poli(A) ou poli(T), forneceu um perfil de mapeamento em regiões 3' UTR e contribuiu para a definição dos limites das mesmas.

No total 1.068.789 *reads*, provenientes do arquivo *unmapped*, foram mapeadas para a biblioteca de epimastigotas cujas sequências de adeninas, timina ou *spliced leader* foram removidas e 2.020.191 para a biblioteca de amastigotas 96 horas. Diante disso, pode-se observar que o mapeamento dessas *reads* coincidia com o mapeamento total das bibliotecas de epimastigotas e amastigotas mapeadas, dando maior confiança aos limites das regiões 3' UTRs estabelecidas. Além disso, foram observadas *reads* mapeando em regiões CDS, porém estas foram desconsideradas nesta análise.

Foram detectados sítios de poliadenilação e *trans-splicing* com base nos tratos de polipirimidina e nas *reads* mapeadas. Sendo assim, verificou-se que para os genes TcCLB.506153.10, TcCLB.506551.10 e TcCLB.509559.20 há mais de um sítio de *trans-splicing* e que existe uma distância média de 140 nt entre sítios de poliadenilação (PAS) e sítios de *trans-splicing* (SAS) para o gene *downstream*, o qual é caracterizado pelo dinucleotídeo AG próximo a uma sequência rica em polipirimidina (HUMMEL et al., 2000; PARSONS et al., 1984; SIEGEL et al., 2005; TEIXEIRA; DAROCHA, 2003; ZWIERZYNSKI et al., 1989).

Região 3' UTR maior que a região codificadora foi encontrada para o gene TcCLB.506153.10, cujo mRNA codifica para transportador de aminoácidos, o que sugere que possa haver sequências codificadoras (ORFs) não anotadas dentro dessa região ou ainda que essas UTR contêm vários elementos regulatórios importantes na regulação da expressão desse gene.

Foi possível determinar os limites de regiões 3' UTR para os 9 genes identificados como tendo expressão diminuída pela TcRBP99. Estes resultados estão sendo abordados na Tabela 7. Os genes anotados como transportador de aminoácidos, proteína associada à diferenciação e RHS (TcCLB.506153.10; TcCLB.506551.10; TcCLB.505943.110 e TcCLB.509559.20) receberam maior atenção por permitir conclusões mais precisas a respeito do papel dessa RBP na regulação da expressão gênica e por isso somente o mapeamento desses genes está sendo mostrado nas figuras 27 e 28.

Tabela 7. Relação entre regiões codificadoras e regiões 3' UTRs dos possíveis genes alvos de regulação exercida pela TcRBP99.

Gene ID	CL Brener	Localização	nº nt (CDS)	nº nt (3'UTR)
TcCLB.508153.10	CL Brener Non-Esmeraldo-like	TcChr12- P:451,352..452,698(-)	1344	1632
TcCLB.508551.10	CL Brener Esmeraldo-like	TcChr27- S:202,017..203,870(-)	1851	880
TcCLB.511843.80	CL Brener Non-Esmeraldo-like	TcChr20- P:580,851..583,158(+)	2505	620
TcCLB.505943.110	CL Brener Esmeraldo-like	TcChr12- S:235,028..237,844(+)	2814	621
TcCLB.509559.20	CL Brener Non-Esmeraldo-like	TcChr12- P:233,641..236,472(+)	2829	380
TcCLB.503733.80	CL Brener Non-Esmeraldo-like	TcChr20- P:580,851..583,158(+)	6552	1907
TcCLB.504173.60	CL Brener Non-Esmeraldo-like	TcChr31- P:432,562..435,225(+)	2661	732
TcCLB.504173.40	CL Brener Non-Esmeraldo-like	TcChr31- P:428,037..428,888(+)	849	1405
TcCLB.508825.40	CL Brener Esmeraldo-like	TcChr27- S:674,103..675,650(+)	1545	1985

Os genes TcCLB.506153.10 e TcCLB.506551.10 possuem alelos situados no outro haplótipo (TcCLB.511411.30 em CL Brener *Esmeraldo-like* e TcCLB.509713.10 em CL Brener *Non-Esmeraldo-like*, respectivamente), cuja expressão não foi detectada como tendo sido estatisticamente diferencialmente expresso. Esse fato pode ser explicado pela baixa cobertura (8 vezes) obtida com o sequenciamento ou mesmo devido a erros de sequenciamento ou mapeamento.

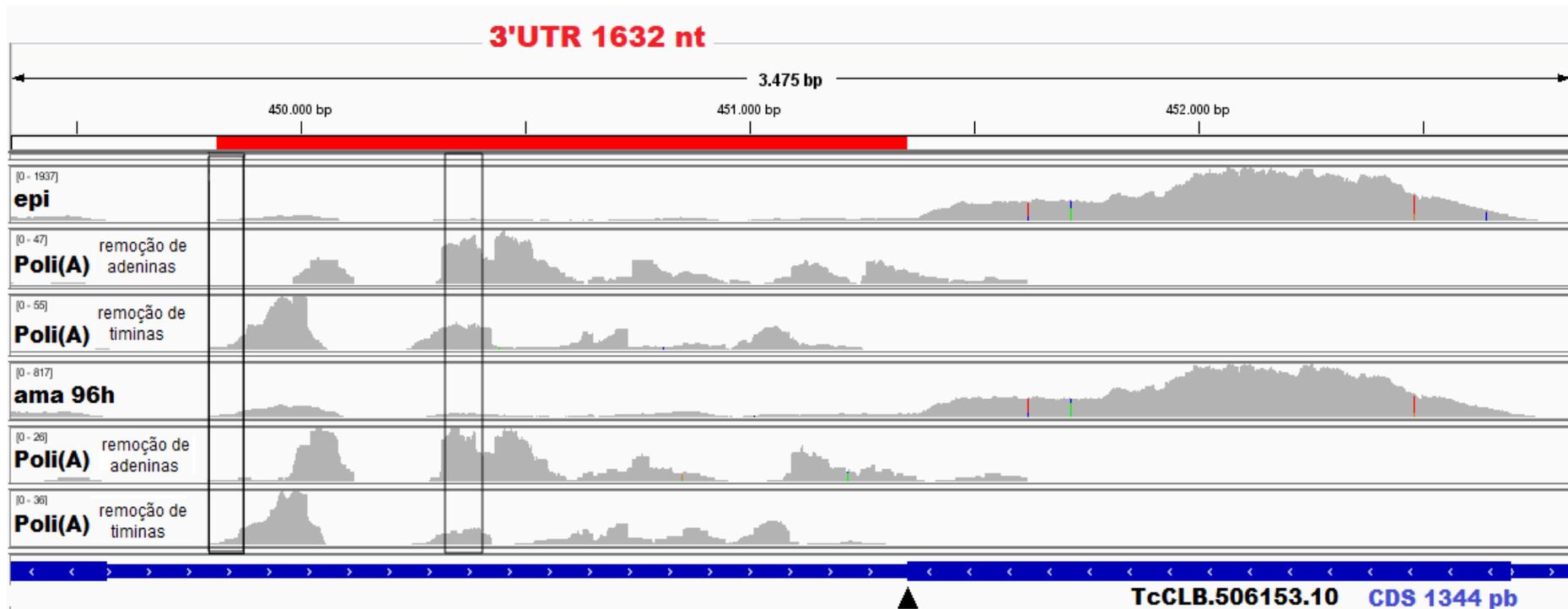


Figura 27. Visualização do mapeamento de reads de epimastigotas, amastigotas 96 horas pós-infecção de células HFF e reads cuja extremidades foram aparadas após a presença de sequências de mais de 4 adeninas ou 4 timinas, correspondentes as duas primeiras bibliotecas citadas. Todas as bibliotecas foram mapeadas contra o genoma de CL Brener (representado pela barra azul inferior). A região codificadora do gene relacionado a transportador de aminoácidos (TcCLB.506153.10) está indicado pelo retângulo azul na barra inferior, tendo seu final representado pelo triângulo preto. Os limites da região 3' UTR está indicado pela linha vermelha. Os retângulos pretos demarcam regiões onde houve mapeamento de reads cujas sequências de adenina ou timina foram removidas.

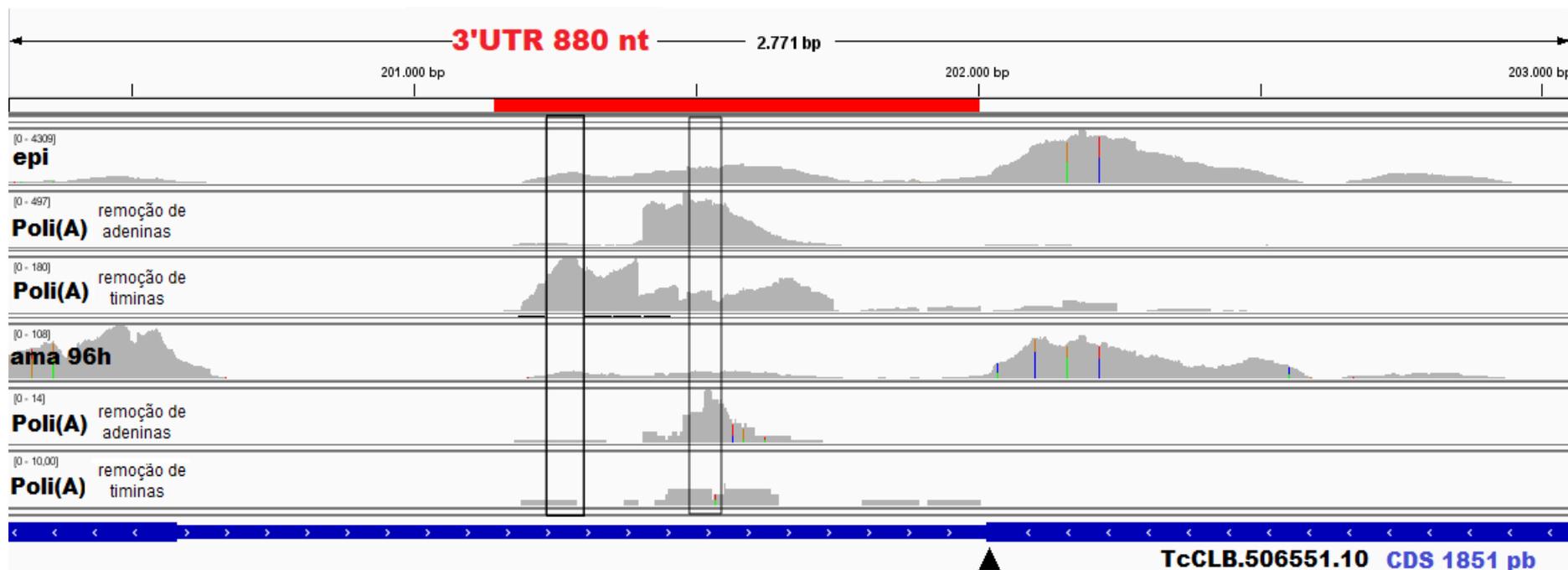


Figura 28. Visualização do mapeamento de *reads* de epimastigotas, amastigotas 96 horas pós-infecção de células HFF e *reads* cuja extremidades foram aparadas após a presença de sequências de mais de 4 adeninas ou 4 timinas, correspondentes as duas primeiras bibliotecas citadas. Todas as bibliotecas foram mapeadas contra o genoma de CL Brener (representado pela barra azul inferior). A região codificadora do gene relacionado a proteína associada à diferenciação (TcCLB.506551.10) está indicado pelo retângulo azul na barra inferior, tendo seu final representado pelo triângulo preto. Os limites da região 3' UTR está indicado pela linha vermelha. Os retângulos pretos demarcam regiões onde houve mapeamento de *reads* cujas sequências de adenina ou timina foram removidas.

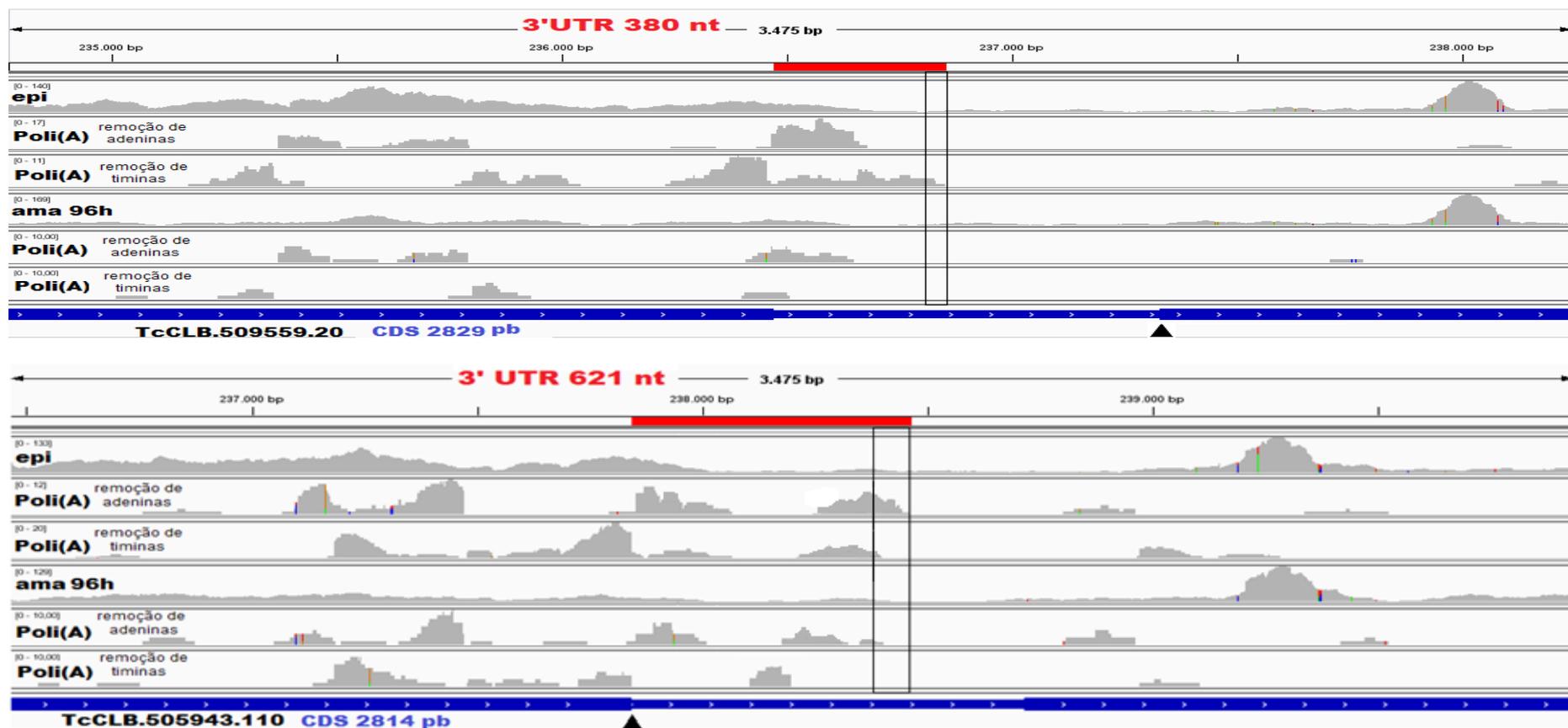


Figura 29. Visualização do mapeamento de *reads* de epimastigotas, amastigotas 96 horas pós-infecção de células HFF e *reads* cuja extremidades foram aparadas após a presença de sequências de mais de 4 adeninas ou 4 timinas, correspondentes as duas primeiras bibliotecas citadas. Todas as bibliotecas foram mapeadas contra o genoma de CL Brener (representado pela barra azul inferior). As regiões codificadoras dos genes pertencentes a família multigênica das RHS (TcCLB.505943.110 e TcCLB.509559.20) estão indicados pelos retângulos azuis na barra inferior de ambas as figuras, tendo seus finais representados pelo triângulo preto. Os limites das regiões 3' UTR estão indicados pelas linhas vermelhas. Os retângulos pretos demarcam regiões onde houve mapeamento de *reads* cujas sequências de adenina ou timina foram removidas.

5.6 Alinhamento múltiplo local das regiões 3' UTRs

Para avaliar o grau de similaridade entre as regiões 3' UTRs estabelecidas para os genes: TcCLB.506551.10, TcCLB.506153.10, TcCLB.505943.110 e TcCLB.509559.20, foi realizado um alinhamento múltiplo entre elas, utilizando a ferramenta MUSCLE, resultando na construção de uma árvore filogenética, traçando o perfil de similaridade (Figura 30).

O mapeamento mostrou clara divergência entre as sequências alinhadas, com no máximo três bases conservadas consecutivamente, em duas posições diferentes (Figura 31).

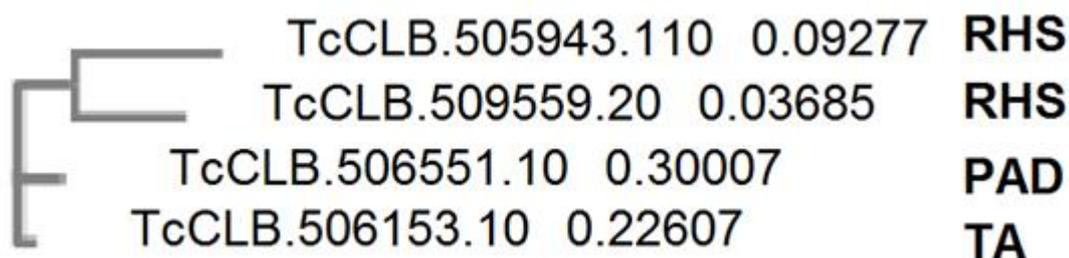


Figura 30. Árvore filogenética de sequências nucleotídicas das regiões 3' UTRs de mRNAs cuja expressão foi possivelmente diminuída pela ação da TcRBP99. Os genes relacionados aos membros da família RHS (TcCLB.505943.110 e TcCLB.509559.20) estão separados por pequenas distâncias (0.09277 e 0.03685) e partem de uma mesma raiz. Os genes relacionados a proteína associada à diferenciação (PAD) e ao transportador de aminoácidos (TA) (TcCLB.506551.10 e TcCLB.506153.10) estão separados por distâncias de 0.30007 e 0.22607 e partem de raízes distintas.

A árvore filogenética resultante do alinhamento entre as sequências de regiões 3' UTRs revelou maior similaridade entre os dois genes relacionados às proteínas da família das RHS (TcCLB.505943.110 e TcCLB.509559.20), como o esperado, mas também revelou maior proximidade entre as sequências das 3' UTRs de genes relacionados ao transportador de aminoácidos e PAD (TcCLB.506153.10 e TcCLB.506551.10).

```

TcCLB.505943.110 -----UUUGUGUG-
TcCLB.509559.20 -----UUUGUGUG-
TcCLB.506153.10 AGUGCUGCGUGCGCUGCGGAGUGGAAGGGGGGGGGGUUUAUUUUUGUUUUUUUGUGUGA
TcCLB.506551.10 -----UCUUUUUUUUUUUUUUUUUGG
TcCLB.511843.80 -----UUUUUGUG-
                                     *** * **

TcCLB.505943.110 -----CAUACG-----
TcCLB.509559.20 -----CAUACG-----
TcCLB.506153.10 GUUGGCGCACGUGAAACCGCCUGUGACCUCCCGCCGUCACCUCCCCACAUACGCCCCUGC
TcCLB.506551.10 -----CGGGCGUCACU-U
TcCLB.511843.80 -----UGUG-----
                                     *

TcCLB.505943.110 -----GUGC-----GGGUGCAC-----CAGCACCAGGUGAC--AGUGU-----U
TcCLB.509559.20 -----GUGC-----GGGUGCAC-----CAGCACCAGGUGAC--AUUAU-----U
TcCLB.506153.10 UGGAAUGUGCCCAGCGGGGCGCAUAAAUGGAAUAACACGAGAU-AC--AAUUUAAGAAU
TcCLB.506551.10 UUUUCCGCGCGU----GAGUCUCCG-----GGGAC--GGUGACGGAGCAU-----U
TcCLB.511843.80 -----UGUGCCUAC--GAGGUCUU-----GACAGUGAAUGCC--UUUCUU-----U
                                     * ** * * * * *

TcCLB.505943.110 ACAAAAAGAUGUCAAAAUGUCCAACGCUAUCG-----AACAUACAGUGUCAGAAGGAA
TcCLB.509559.20 ACAAAAAGGUCUCAAAAUGACCAACGCUAUCG-----AACAUACAGUGAGAGAAGGAA
TcCLB.506153.10 AGAAAAAGGAAUAAAUGAAAUGAGGUGUGG-----AGAAAUAUUUAACAAGGAA
TcCLB.506551.10 GCGGGGCGGUGUGCACGCGGACCGA---GCGA---AGGAUGUGGUGAUAAUAAGGUA
TcCLB.511843.80 UCAGAGAACC GUUU----UGUUCGGCAGUGCCACACUCACCACACCUCUUUUUCAUUUUUA
                                     * * * *

TcCLB.505943.110 -UUAG--GGAGG-----GAGUGGGGUUUUUG-GUUUCUGGU-----
TcCLB.509559.20 AUUAG--GGAGG-----GAGUGGGUUUUUUUUG-GUUUCUGGU-----
TcCLB.506153.10 ACCAACAGAGG-----UAGUGUUUUUUUUUUUGUUUUUCGUAUAGUAUACCCCC
TcCLB.506551.10 ACUUUUUUAUUCUGUUUUUGAGUUUUUUUUUACUGUUUUUCCUUUACUUUUUCUCCUC
TcCLB.511843.80 UUUUUAUUUAUCCACCAGAGGGCGCAUCGCCACUGGAGGCGUUGC-----
                                     * * * * *

TcCLB.505943.110 -----GAAAA-----ACAUUGGGUGA-----
TcCLB.509559.20 -----GAAAA-----ACAUUGGGUGA-----
TcCLB.506153.10 AACCCCUGCCACCACCACCGCCAGAGAAUCUCCCAUAUGAACGGAGUCAUGGUGUUUU
TcCLB.506551.10 -----GGUCCU-----UUUUGGCUGAA-----UUUUUAC
TcCLB.511843.80 -----GCACU-----ACAGGGGCCGCA-----UAA
                                     * *

TcCLB.505943.110 UUUUUAUUAUAGUAUGU-----UAGUUGUGUGUGUG---CACACGAG-----
TcCLB.509559.20 UUUUUAUUAUAGUAUGU-----UUUUGUGUGUGUG---CAUACGAG-----
TcCLB.506153.10 UUUUUUUUAGCACGCGGACUUGUUGGUGUAUAUGAA---CUCAUGGAU-UGGUA
TcCLB.506551.10 UUUCAUUUUUAUUUUU--UCCUCCUUUCUGGGCGUACAUGUUUUUAUUUGAGUGCGAUU
TcCLB.511843.80 UUUUCCCUU-GUAUUUA-----UGGGAGUUUGUGA-----GAA-----
                                     *** * * * * *

```

Figura 31. Alinhamento das seqüências correspondentes às regiões 3' UTRs de mRNAs cuja expressão foi possivelmente diminuída pela ação da TcRBP99.

Da mesma forma, o alinhamento entre as CDS dos quatro genes, utilizando a ferramenta *MUSCLE*, mostrou grande divergência entre as seqüências. Existem no máximo seis nucleotídeos consecutivos conservados (Figura 32). Mas, diferentemente do alinhamento das 3' UTRs, o alinhamento entre as CDS (Figura 33) resultou na construção de uma árvore que agrupou apenas as seqüências relacionadas as RHS, o que era esperado, devido ao fato de serem alelos do mesmo gene.

```

TcCLB.505943.110 GGAGGAGGATGAATCCACTCCTGACTGCTATGTGAACTGTGAGGTGGATCGAGTGTGGCA
TcCLB.509559.20 GGAGGAGGATGAATCACTCGTGACTGCTACGTGAACTGTGAGGTGGATCGAGTGTGGCA
TcCLB.506551.10 CGATCAGCAGGGGCCCTGTCGTGGCCATTTTG-AAGTCGTACATCGGACTGGGCTCCGCC
TcCLB.506153.10 TGTCGACGACGTTTAACTTTCGAGCGCCACG- - - CTTGGTGGCGGGGTGA- - TCAGCC
* * * * * * * * * * * * * * * * * * * * * *

TcCLB.505943.110 GATTGTCAGGAATGATCTAACC CGTGGTT CAGCCCTCACCTGGGGC-CTACTT-TACG
TcCLB.509559.20 GATTGTCAGGAATGACCTAACC GAATTTGTT CAGCCCTCACCTGGGGC-CTACTT-TACG
TcCLB.506551.10 -ATTGT-GGGAAGCATCCAGCTGGCC TTTTTCAGGGAGGCGGACCCTACTTCTATT
TcCLB.506153.10 - - -TGGCAGCGGGCTTTCAGATGAGTGGTGTGCGCCTTCTCCATCATCC- - TTCTTGT-CC
* * * * * * * * * * * * * * * * * * * * * *

TcCLB.505943.110 CCTAGGCGGCGTGTGTTGATTGGGACGCCCGGGATTGGGAAATCTTTGGCTGCCGGCTCG
TcCLB.509559.20 CCTAGACGACGTGTGTTGATTGGGACGCCCGGGATCGGGAAGTCGATGAATGCCGGCTCG
TcCLB.506551.10 TTCTGATGGTGC-TGTTTTTTGTGA- - - CTGGAGCTGCGGGTTTCTCCTTGTG- - - CCA
TcCLB.506153.10 TGGTGACGGTGTGACTGTCTATTCCG- - TCGGCCTGATGATGCAGGCTGTGGAGATGAC
* * * * * * * * * * * * * * * * * * * * * *

TcCLB.505943.110 TACCTCCTCTACCAGCTGCTGCAATGCGATGTCAAGAAACTCCAAGTG-GTTGTCTACAG
TcCLB.509559.20 TACCTCCTCTACCAGCTGCTGCACTGCGATGCCAAGAAACTCCATTTG-GTTATCTACAG
TcCLB.506551.10 CTCCGTCGTACCACCTGACTGGCTATGAGG- - -AGAAACACCTTGGC-ATCGAGGAAAA
TcCLB.506153.10 CGGCTATAAT-TCATACGCGGAGCTCTCAA- - -AGAAACTGTTTGGCCGGGGGTGGGAC
* * * * * * * * * * * * * * * * * * * * * *

TcCLB.505943.110 TTTTGGTGGA- -AGCACGACGTACGTGTTTGACAAGACCATCAA- -GACGGTGACAAGAT
TcCLB.509559.20 TTTTGGTGGA- -AATACGACGTACGTGTTTGACAAGACCATCAA- -GGCGGTGACAAGAT
TcCLB.506551.10 GGAGAGACGACTGGCACG-CAAATCCGTTTACCTCCGCCAGCAACCACCCACAATTCGCT
TcCLB.506153.10 CACTTACGGTGGTCTCACGTGGCTGTTACCTTTGGCACCTGT-GTGGGCTATGTCAT
* * * * * * * * * * * * * * * * * * * * * *

TcCLB.505943.110 ACGTGGGTAGAGAAGCATCCAAGGAA-TGTTTGTGTGATTTGCGGGATTTGAAGATGAAA
TcCLB.509559.20 ACGTGGGTGGAGGAGCATCCAAGGAA-TGTTTGCCTGATTTTGGCGTTTGAAGATGAAG
TcCLB.506551.10 TCGCGATCGGCATTGCGTTTGTGCTGCTGTTATATACTTGCCACTGCAGAGCGCAC
TcCLB.506153.10 A-GCCATTGGCCGTCTCCTTGAGCCCGTGT-GTCCGATCCGTCCCTGCCGGAGTTTTGG
* * * * * * * * * * * * * * * * * * * * * *

TcCLB.505943.110 GGGTATATTATCTACGATGTGACAAGGCAAGGGACACCGCCAGAGGAATATTTTTTGCCT
TcCLB.509559.20 GGGTATGTTATCTACGACGTGACAAGGCAAGGGACACCTCCAGAGGAATATTTTTTGCCT
TcCLB.506551.10 TGTTTGCATCTGGGGTGGGGGAGGACGCAGCGCATCAT-ATTTGCGTCCATCTTGATT
TcCLB.506153.10 CGCTCCGAGGCTGGCAATCGCGCGGTGACAAGTGCAATCTGGTTTGTGGGCATGTT- -CT
* * * * * * * * * * * * * * * * * * * * * *

```

Figura 32. Alinhamento das seqüências correspondentes às regiões CDS de mRNAs cuja expressão foi possivelmente diminuída pela ação da TcRBP99. O retângulo vermelho representa a maior seqüência de nucleotídeos consecutiva conservada. Os retângulos azuis representam seqüências de três nucleotídeos consecutivos conservados.

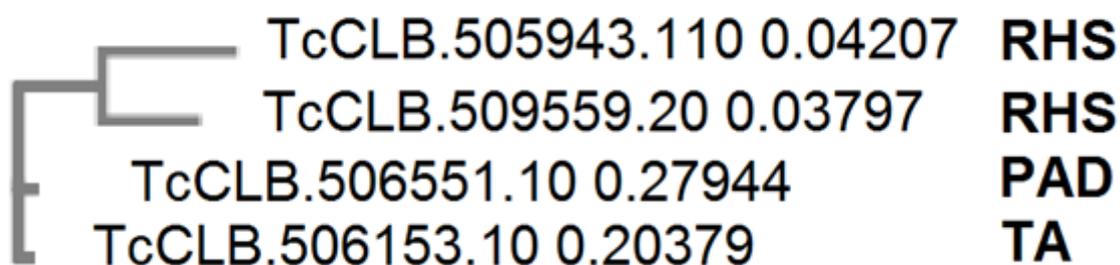


Figura 33. Árvore filogenética das regiões CDS de mRNAs cuja expressão foi possivelmente diminuída pela ação da TcRBP99. Os genes que codificam para membros da família RHS (TcCLB.505943.110 e TcCLB.509559.20) correspondem a distâncias de 0.04207 e 0.03797 e partem da mesma raiz. Os genes que codificam para proteína associada à diferenciação e transportador de aminoácidos (TcCLB.506551.10 e TcCLB.506153.10) estão separados por distâncias de 0.27944 e 0.20379 e partem de raízes distintas.

5.7 Sequências ricas em uracila e citosina presentes nas regiões 3' UTRs

Como já ressaltado anteriormente, a identificação de motivos regulatórios que estão presentes em regiões UTRs é de suma importância para a compreensão dos mecanismos de regulação da expressão gênica, afinal, estes são responsáveis por permitir a interação dos mRNAs com elementos que atuam em *trans* ou ainda por desempenharem seu papel com atuação em *cis*. Esta etapa foi executada para os mRNAs que codificam PAD, transportador de aminoácidos e RHS, cujas coordenadas de regiões 3' UTR foram determinadas anteriormente. A ferramenta *web MEME suite* foi utilizada para a identificação dos motivos.

O conjunto de sequências 3' UTRs previamente alinhadas (referentes aos genes: TcCLB.506551.10, TcCLB.506153.10, TcCLB.505943.110 e TcCLB.509559.20) foi inserido na plataforma, permitindo a identificação de três motivos (Figura 34) com significância estatística (*e-value* < 1e-5), baseada na probabilidade-dependente da posição. Outro parâmetro estabelecido foi considerar qualquer quantidade de ocorrências em cada sequência analisada. Assim, os resultados não limitaram o fato de poder haver a ocorrência de repetições para cada um dos motivos. O motivo 3 (barra verde maior) é o que apareceu mais vezes comparado aos outros 2, chegando a aparecer 4 vezes em uma mesma sequência, enquanto que os motivos 1 e 2 (retângulo vermelho e azul claro, respectivamente) aparecem no máximo três vezes (Figura 35). Além disso, o motivo 1 é caracterizado por ser rico em uracilas, assim como o motivo 3, rico em citosinas.

Apesar da identificação de três motivos estatisticamente significativos, nenhum dos motivos identificados é encontrado nas quatro sequências analisadas. Quando as sequências 3' UTRs dos mRNAs de PAD e transportador de aminoácidos são analisadas separadamente, observa-se que o motivo 3 é comum a ambas. Diferentemente, notamos que para as sequências de RHS, os motivos 1 e 2 encontram-se em ambas as sequências 3' UTRs.

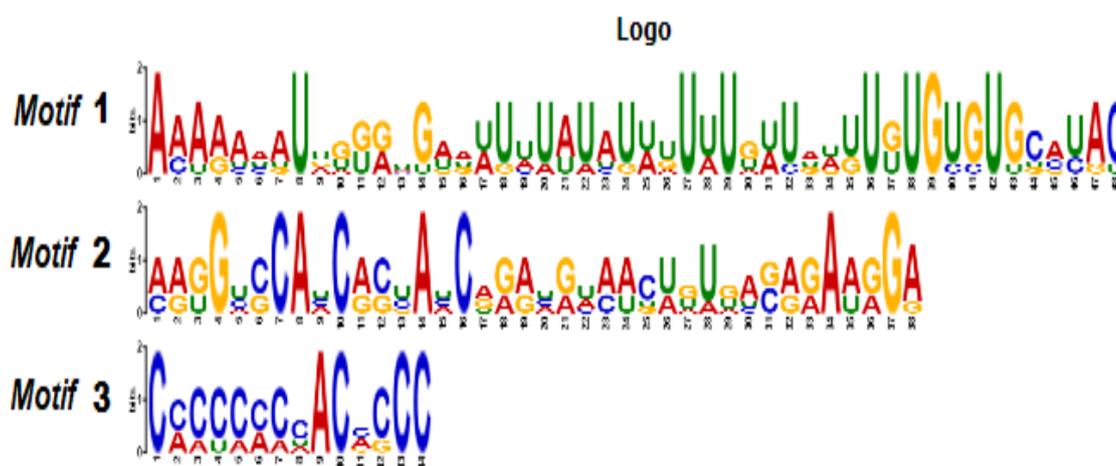


Figura 34. Logo dos motivos descobertos entre as sequências de regiões 3' UTR dos possíveis mRNAs alvos da TcBP99. Só estão representados motivos que apresentaram significância estatística ($e\text{-value} < 1e\text{-}5$).

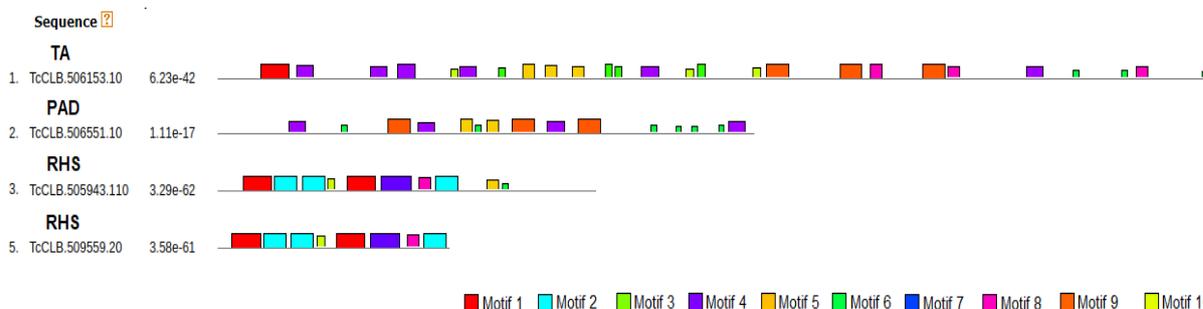


Figura 35. Visualização de ocorrências dos motivos descobertos pela ferramenta *MEME suit*. Os motivos de 1 ao 3, apresentaram significância estatística ($e\text{-value} < 1e\text{-}5$). Os motivos de 4 ao 10 foram identificados com $e\text{-value} > 1e\text{-}5$.

Para analisar o padrão de ocorrência dentro das regiões 3' UTRs analisadas, ou seja, verificar a existência de motivos localizados em posições similares foi realizado o alinhamento destes utilizando o arquivo gerado pelo MEME. Esta etapa foi concluída através de outra ferramenta da mesma plataforma, denominado MAST.

Mas, com o objetivo de realizar uma análise mais abrangente, foram considerados motivos com *e-value* maior que 1e-5. No total 10 motivos foram analisados quanto seus padrões de ocorrência (Figura 36 e 37).



Figura 36. Logo dos motivos considerados na análise de alinhamento realizados através da ferramenta MAST.

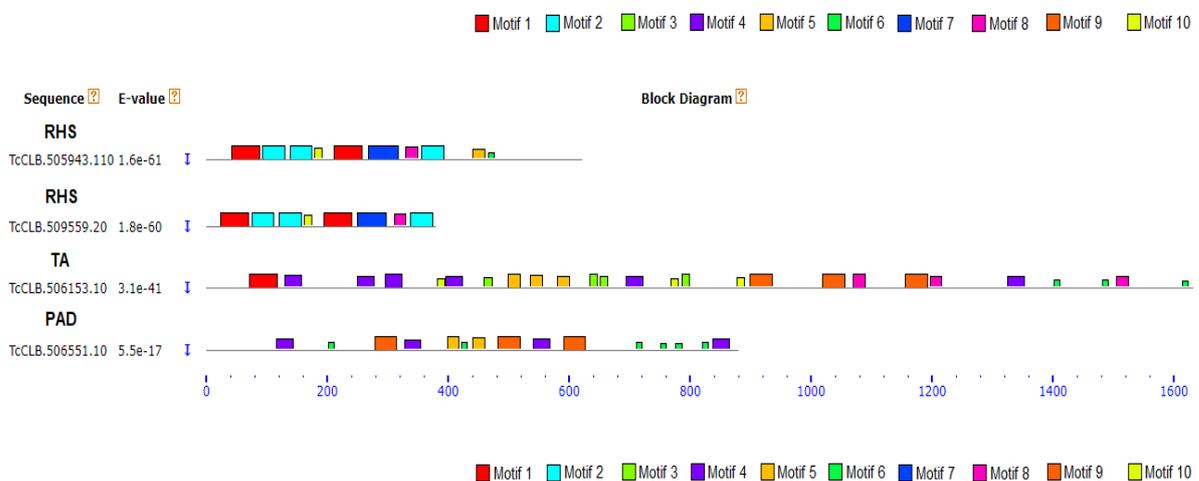


Figura 37. Diagrama dos motivos preditos presentes nas regiões 3' UTRs dos possíveis mRNAs de alvos da TcRBP99.

Os resultados mostram que apenas as regiões 3' UTRs dos genes TcCLB.505943.110 e TcCLB.509559.20 possui um padrão de ocorrência semelhante dentre os motivos identificados, tendo as posições dos motivos compartilhadas entre ambos.

Diante dos resultados obtidos pela busca e identificação de motivos presentes nas regiões 3' UTRs de possíveis alvos da TcRBP99 é possível concluir que existem sequências conservadas que podem ser importantes para a ligação da RBP nos mRNAs de RHS, PAD e transportador de aminoácidos, separadamente. Mas, não existe nenhum motivo comum a todas as sequências 3' UTRs. Sendo assim, podemos especular que existem outros fatores que estejam influenciando na estabilização desses mRNAs, juntamente com a TcRBP99. A hipótese é de que a TcRBP99 interage com outras proteínas, as quais se ligam a outros elementos presentes nos possíveis alvos dessa RBP ou ainda que a TcRBP99 regule mRNAs primários, sendo estes responsáveis pelo efeito estabilizador de alguns dos mRNAs que foram identificados na análise de expressão diferencial.

6. DISCUSSÃO

O estudo do transcriptoma desempenha papel chave no entendimento dos mecanismos de regulação gênica pós-transcricionais que ocorrem no *T. cruzi*. Os avanços das técnicas de sequenciamento de nova geração (NGS – *Next generation sequencing*) juntamente com a técnica de RNA-seq e o desenvolvimento de ferramentas computacionais têm facilitado o armazenamento e a análise de grande volume de dados (MUTZ et al., 2013; PARK; KIM, 2016). Dessa forma, a complexidade do parasita em questão aos poucos vem sendo elucidada, assim como os fatores associados as drásticas mudanças que ocorrem ao nível do transcriptoma, proteômica ou metabolômica em decorrência das diferentes condições celulares que o parasita enfrenta.

Um dos principais elementos envolvidos na regulação pós-transcricional são as RBPs, que como descrito anteriormente, atuam por meio da ligação a motivos presentes nas regiões 3' UTRs dos mRNAs alvos. Como consequência ocorre o aumento ou diminuição dos níveis de mRNA na presença ou ausência das mesmas (ALVES; GOLDENBERG, 2016; ARAÚJO; TEIXEIRA, 2011; RUIZ-ECHEVARRÍA; PELTZ, 2000). Outros efeitos decorrentes da ligação de RBPs a mRNAs, como a mudança na localização destes na célula, modulação de eventos de *splicing* ou tradução são relatados, mas não foram avaliados no presente trabalho (DE GAUDENZI et al., 2016a; EDWARDS, 2015; KISHORE et al., 2010).

A proteína TcRBP99, identificada como sendo uma RBP diferencialmente expressa em formas epimastigotas, mostrou estar possivelmente envolvida em processos de diferenciação e proliferação celular durante condições de estresse nutricional. A comparação entre epimastigotas WT e epimastigotas nocautes mostrou que a taxa de diferenciação é maior em condições nocautes ao mesmo tempo em que apresentam crescimento reduzido na fase final de crescimento em meio LIT (VALENTE et al, manuscrito em preparação). Sendo assim, neste trabalho, *reads* geradas a partir de epimastigotas nocautes e WT foram submetidas à análise de expressão diferencial a fim de permitir a identificação de genes possivelmente regulados pela TcRBP99 e a relação destes com o processo de diferenciação e crescimento do parasito. De forma complementar foram determinadas as regiões 3'

UTRs dos mRNA alvos da TcRBP99 para posterior identificação de motivos nessas regiões, os quais são essenciais para a ligação da RBP.

O sequenciamento MiSeq de bibliotecas de cDNA, proveniente de epimastigotas nocautes e WT permitiu a obtenção de uma quantidade limitada (8,75 Mb de *reads* por biblioteca) de *reads* em comparação com a plataforma HiSeq, mas suficiente para a detecção de mRNAs que interagem com a TcRBP99, devido ao pequeno tamanho de aproximadamente 105 Mb do genoma de *T. cruzi* (EL-SAYED et al., 2005; SOUZA et al., 2011). Sendo assim, por meio de análise de expressão diferencial, um total de 10 genes apresentou diferença de expressão em epimastigotas nocautes comparadas a epimastigotas WT, tendo tido relevância estatística para esta análise ($p_{adj} < 0,05$ e $-1.1 > \text{Log}_2\text{FC} > 1.1$). Dentre eles, quatro dos mRNAs que aparecem por ter expressão reduzida estão anotados no banco de dados TrypDB da seguinte forma: dois correspondem a genes da família multigênica das RHS, um ao gene que codifica para transportador de aminoácido e por último, um gene multicópia relacionado a proteína de diferenciação.

Pacotes estatísticos focados na análise de expressão diferencial gênica tem ganhado destaque no estudo do transcriptoma, devido a sua alta eficácia na detecção de genes positivamente ou negativamente regulados em diferentes estágios do ciclo de vida de *T. cruzi* ou mesmo nas diferentes fases do ciclo celular (G1, S, G2) (BELEW et al., 2017; LI et al., 2016; PASTRO et al., 2017; SANTIAGO et al., 2017; SMIRCICH et al., 2015). Estudos comparativos entre esses pacotes têm revelado características relevantes para o DESeq2, que mostrou ser mais conservador em relação a outros pacotes utilizados, como o edgeR e limma (SCHURCH et al., 2016; ZHANG et al., 2014; ZHU et al., 2015; GUO et al., 2013). Deseq2 assume distribuição binomial negativa para as *reads* mapeadas, assim como o pacote edgeR, porém admite menor quantidade de falsos positivos (maior controle das taxas de FDR - *False Discovery Rate*) fornecendo resultados mais eficazes, ideal para a situação de baixa cobertura (8X) existente neste trabalho (RAPAPORT et al., 2013; SCHURCH et al., 2016; SONESON; DELORENZI, 2013). Outro fator determinante na escolha do Deseq2 foi a menor variação sofrida em decorrência do número de replicatas biológicas, no qual este pacote mostrou ser menos tendencioso comparado ao pacote limma, por exemplo (SCHURCH et al., 2016; SONESON; DELORENZI, 2013).

Um dos genes que teve sua expressão diminuída em epimastigotas nocautes de TcRBP99, o gene TcCLB.506551.10, é um membro da família das PADs. Como mostra o resultado do *HeatMap* de membros dessa família, construído a partir de dados de transcriptoma de epimastigotas, amastigotas coletadas 60 e 96 horas após a infecção e tripomastigotas, esse gene apresenta níveis de transcritos aumentados em epimastigotas (BELEW et al., 2017). Sendo assim, podemos especular que esse gene tem sua expressão aumentada em epimastigotas pelo fato de que essas formas de vida necessitam permanecer tempo suficiente no intestino médio do vetor, a fim de se multiplicarem de forma a aumentar a eficiência da infecção (VAZ, 2017). Portanto, em situação contrária, onde esse gene apresenta expressão diminuída, como ocorre nas epimastigotas nocautes para a TcRBP99, os parasitos são induzidos a se diferenciarem. Com isso, os altos níveis de expressão desse gene permitem que epimastigotas possam se multiplicar adequadamente antes de iniciar o processo de metaciclogênese.

Expressão aumentada de genes PADs também foi observada em epimastigotas da cepa Y e Dm28c, quando valores de expressão FPKM foram analisados (LI et al., 2016). Em *T. brucei* duas PADs foram caracterizadas, recebendo os nomes de PAD1 e PAD2, as quais estão aumentadas em formas sanguíneas e procíclicas respectivamente. PAD2, que apresenta 56% de identidade com a PAD identificada como sendo possivelmente regulada negativamente pela TcRBP99, mostrou ter sua expressão fortemente induzida durante a diferenciação de formas sanguíneas para as formas procíclicas, presentes no vetor (DEAN et al, 2005). Esse resultado, juntamente com dados de genes ortólogos em *T. brucei* sugerem que as PADs atuam como transdutoras de sinal que disparam a diferenciação. Baseado nisso, especula-se que em *T. cruzi*, a expressão do gene PAD é induzida durante o processo de diferenciação de tripomastigotas extracelulares em epimastigotas, durante o hematofagismo, e que os níveis da mesma continuam elevados em epimastigotas, configurando-se como proteínas importantes para a manutenção dessas formas de vida no inseto vetor. Diante disso, a diminuição nos níveis de expressão desse gene levaria a uma redução da capacidade de proliferação, o que pode funcionar como um “gatilho” para disparar a sinalização para a diferenciação em tripomastigotas metacíclicas.

Para analisar o envolvimento da TcRBP99 e da proteína associada à diferenciação em vias de transdução de sinal mais estudos são necessários. Para isso, existem bancos de dados como KEGG que funciona como uma enciclopédia de genes e genomas permitindo busca e análise de redes de interações moleculares específicas (KANEHISA; GOTO, 2000). Estudos de nocaute para o gene que codifica a proteína associada à diferenciação e posterior análise da taxa de diferenciação de tripomastigotas em epimastigotas podem ajudar a elucidar se existe uma relação direta ou indireta da RBP em processos de diferenciação e proliferação de epimastigotas. Espectrometria de massa e eletroforese em gel bidimensional (SDS-PAGE) poderão contribuir para a identificação de outras moléculas que se ligam a esse mRNA e a relação dos mesmos com a diferenciação.

Recentemente, a interação da TcRBP99 com o mRNA que codifica PAD (TcCLB.506551.10) foi confirmada por meio de co-imunoprecipitação, mostrando que TcRBP99 se liga de forma específica a esse mRNA (VALENTE et al., manuscrito em preparação). Além disso, resultados que mostram aumento da taxa de diferenciação de epimastigotas nocautes para tripomastigotas metacíclicas e diminuição do crescimento de epimastigotas nocautes observados por VALENTE et al. (manuscrito em preparação) podem estar relacionados ao fato de que a TcRBP99 poderia estabilizar o mRNA relacionado à PAD, resultando nos fenótipos observados.

Um segundo possível alvo da TcRBP99 é o mRNA que codifica transportador de aminoácidos. Vários trabalhos publicados mostram um importante papel de genes relacionados a transportadores de aminoácidos em epimastigotas de *T. cruzi*, uma vez que o metabolismo energético de epimastigotas, na ausência de glicose, é baseado dentre outras fontes, no consumo de aminoácidos (principalmente L-prolina) obtidos por proteólise e sistemas de transporte (CANEPA et al., 2004; LAWRIE, 1935; PEREIRA et al., 1999; SILBER et al., 2005; URBINA; AZAVACHE, 1984). LI et al., (2016) por análises de expressão diferencial mostraram que em epimastigotas da cepa Y, vários genes que codificam proteínas envolvidas na aquisição de nutrientes (por exemplo, transportadores de aminoácidos, açúcares de hexose e nucleósídeos) estão pelo menos duas vezes mais aumentados, de forma seletiva, inclusive o gene identificado neste trabalho (TcCLB.506551.10). O mesmo resultado pôde também ser observado por meio do *HeatMap* da expressão de genes que codificam para

transportadores de aminoácidos em epimastigotas, uma vez que 11 dos 17 genes dessa família apareceram por estar aumentados em epimastigotas de *T. cruzi*.

Pressupõe-se que os aminoácidos obtidos pelo parasita são provenientes do sangue ingerido pelo inseto vetor, o qual tem a hemoglobina como principal proteína constituinte. Esses aminoácidos ficam então disponíveis na porção final do tubo digestivo do inseto vetor (KELLER et al., 2017). Dessa forma, transportadores de aminoácidos são indispensáveis no processo de obtenção desses nutrientes pelo parasita. Sendo assim o mRNA do gene TcCLB.506153.10, que codifica para transportador de aminoácidos, em epimastigotas WT, poderia estar sendo estabilizado pela TcRBP99, garantindo sua expressão aumentada nessas formas de vida. O crescimento reduzido de epimastigotas nocautes para a TcRBP99 observado por VALENTE et al. (manuscrito em preparação) corroboram o resultado de expressão reduzida desse gene, representando a consequência da modulação na expressão desse gene.

Uma vez disponíveis no organismo de epimastigotas, os aminoácidos são oxidados pelo ciclo dos ácidos tricarbóxicos (do inglês *tricarboxylic acid cycle* – TAC), levando a produção de ATP (revisado por BRINGAUD et al., 2006). Portanto, o papel da TcRBP99.39 na regulação de processos energéticos também deve receber atenção. Estudos similares aos propostos para a avaliação em processos de diferenciação podem ser realizados para o gene relacionado a transportador de aminoácido, bem como ensaios de inibição para a proteína relacionada e posterior avaliação das taxas de crescimento e diferenciação.

Membros da família RHS constituem o terceiro grupo de possíveis alvos da TcRBP99. As RHS constituem a quarta família multigênica com mais membros no genoma de *T. cruzi*, porém seu papel ainda é pouco conhecido (KIM et al., 2005). A maioria dos genes está localizada em regiões subteloméricas, entre os telômeros e a junção telomérica, a qual parece ser mais propensa a sofrer danos e consequentemente, reparos e recombinações no DNA (HORN; BARRY, 2005; KIM et al., 2005; LIRA et al., 2007). L1Tc é um retrotransposon de repetição terminal não longa (LTR, do inglês *non-long terminal repeat*), caracterizado por possuir sequências Seq3Tc flanqueando-o (OLIVARES et al., 2000). Essa característica estrutural do gene indica a inserção de L1Tc em um (pseudo) gene, evento que provavelmente descreve os genes da família das RHS.

Análises de sequências de aminoácidos provenientes de regiões subtelméricas mostraram que existe uma identidade de 28-30% com as proteínas RHS de *T. brucei*, incluindo um local de inserção putativo para retroelementos, (BRINGAUD et al., 2002), o que reforça a propensão dessas regiões a sofrerem constantes recombinações. Além disso, PARODI-TALICE et al. (2004), por análise proteômica, mostrou que as RHS são altamente expressas em epimastigotas, corroborando o resultado obtido pela análise de expressão diferencial que indicou diminuição da expressão na ausência da TcRBP99 e conseqüentemente maior expressão de membros das RHS em condições WT.

HeatMaps mostraram que em geral os genes dessa família estão aumentados em epimastigotas (> 80% dos genes de RHS). O mesmo resultado foi obtido por BELEW et al. (2017) quando mRNAs de epimastigotas, amastigotas coletadas 60 e 96 horas após a infecção de células HFF e tripomastigotas foram analisados por RNA-seq. Mas, em 2015, NAGULESWARAN et al. mostraram que uma proteína nuclear de ligação a RNA, denominada TbRRM1, interage com mRNAs que codificam para membros da família RHS, podendo inibir a tradução dos mesmos. Portanto, apesar de apresentar níveis aumentados de expressão, algumas proteínas RHS podem não ser sintetizadas em epimastigotas. O motivo pelo qual estes genes estariam aumentados nessas formas de vida ainda é desconhecido.

Mas, apesar dos indícios encontrados neste trabalho que sugerem que a TcRBP99 desempenha um importante papel na estabilização de 9 genes, dentre eles, genes que codificam para PAD, transportador de aminoácidos e proteínas RHS, outros experimentos são necessários para confirmar a expressão dos mesmos. Para isso PCR em tempo real (qRT-PCR) será realizada a partir dos mRNAs correspondentes a esses genes e os níveis de expressão serão avaliados comparando-se os níveis em epimastigotas WT e nocautes para TcRBP99. Outro experimento previsto é o *add-back* do gene da proteína que codifica a TcRBP99, no qual este gene será reintroduzido em epimastigotas nocautes e os níveis de expressão dos genes alvos serão avaliados novamente por qRT-PCR a fim de verificar se a expressão dos mesmos voltará a ser igual a condições normais. Experimentos de metaciclogênese *in vitro* e *in vivo* (por meio da infecção do vetor com as formas epimastigotas WT e nocautes) também serão conduzidos com o objetivo de avaliar o retorno da taxa de

diferenciação em condições normais, por meio da contagem de tripomastigotas metacíclicas após submeter epimastigotas a estresse nutricional.

Análises similares foram conduzidas para a RBP de *T. cruzi*, nomeada TcZC3H31. Esta é uma proteína exclusiva de epimastigotas e tripomastigotas metacíclicas, a qual se diferencia da TcRBP99, por possuir três domínios dedos de zinco. Outras características que a diferenciam é o fato de que, sua expressão em epimastigotas parece induzir a diferenciação em tripomastigotas metacíclicas, o que de forma ainda superficial, pode indicar um papel estabilizador dessa RBP em mRNAs que estariam induzindo esse processo. Assim, a TcZC3H31 parece ter um comportamento análogo a TcRBP99, mas complementar, uma vez que a superexpressão da TcZC3H31 indica maior diferenciação de epimastigotas e a superexpressão da TcRBP99 indica menor diferenciação desses parasitos (ALCANTARA, 2018).

Para a identificação de possíveis elementos regulatórios presentes em regiões não traduzidas dos genes identificados, foi realizada a delimitação das regiões 3' UTR. Para isso não foram utilizadas as *reads* mapeadas neste trabalho devido a limitada cobertura do sequenciamento (8 vezes), o que dificultaria identificar com precisão os sítios de poliadenilação e *trans-splicing*. Portanto foram utilizadas *reads* obtidas a partir do sequenciamento de epimastigotas e amastigotas 96 horas pós-infecção de células de fibroblastos (BELEW et al., 2017), que mostraram alta cobertura (aproximadamente 103 vezes), para a análise dessas regiões. Buscas por tratos de polipirimidina também foram essenciais nesta etapa, uma vez que influenciam na localização dos sítios de poliadenilação. Assim, os resultados mostraram que não há um padrão de comprimento para estes genes, não sendo este um fator limitante para a interação com a RBP. Porém, regiões 3' UTR maiores que as regiões CDS correspondentes foram identificadas, fato que leva à hipótese de que existem regiões codificantes dentro dessas regiões. Para verificar esse evento é possível realizar a busca por possíveis ORFs, utilizando ferramenta de bioinformática, como o *software ORF Finder* (do inglês *Open Reading Frame Finder*).

Apesar de ter sido possível determinar o comprimento de regiões 3' UTRs de possíveis mRNAs alvos da TcRBP99, os resultados encontrados são considerados preliminares, pois existe um viés de mapeamento inerente a metodologia utilizada. É possível observar que existem *reads* filtradas para regiões poli(A), mas que mapearam

em regiões CDS, por exemplo. Portanto, os tamanhos das regiões 3' UTRs devem ser confirmados por meio de amplificação e sequenciamento das mesmas.

Concomitantemente a determinação das coordenadas de regiões 3' UTR foi realizado o mapeamento de sítios aceptores de *spliced leader*, que mostrou haver mais de um sítio para alguns dos genes alvos, indicando que os mesmos podem ser regulados por *splicing* alternativo. Esse evento foi descrito para famílias multigênicas de *T. brucei* (GUPTA et al., 2013; KOLEV et al., 2010; NILSSON et al., 2010) e Leishmania (DILLON et al., 2015), como observado nos mRNAs que codificam RHS, identificados neste trabalho.

Busca por elementos regulatórios é essencial para o entendimento do mecanismo de ação de RBPs bem como dos efeitos causados na regulação da expressão após eventos transcricionais. Muitos trabalhos já relataram a importância de motivos presentes em regiões UTRs de todos os tripanosomatídeos como elementos de regulação (ALY et al., 1994; BRANDÃO; JIANG, 2009; NOZAKI; CROSS, 1995; PÉREZ-DÍAZ et al., 2013; TEIXEIRA et al., 1999) e mostram seu papel como sítios de ligação de proteínas (RBPs), o que acaba por influenciar na meia vida, eventos de degradação ou ainda na eficiência da tradução dos mRNAs envolvidos (COUGHLIN et al., 2000; DE GAUDENZI et al., 2003; HAILE; PAPADOPOULOU, 2007; HENDRIKS; MATTHEWS, 2007; WU et al., 2006).

Tendo em vista o papel chave desempenhado por motivos de sequências em regiões UTRs, este trabalho buscou identificá-los nas regiões 3' UTR de potenciais mRNAs alvos da TcRBP99. Sendo assim, foram encontrados três motivos para o conjunto das quatro proteínas anotadas como transportador de aminoácidos, proteína relacionada à diferenciação e RHS, ricos em uracilas e citosinas (motivo 1 e 3, respectivamente), e sete motivos quando submetidos às dez proteínas cuja região 3' UTR foi delimitada. Este resultado é similar ao encontrado por D'ORSO e FRASCH (2001), que identificaram motivos ricos em uracilas, responsáveis pela ligação da proteína TcUBP1. Além disso, MÖRKING et al. (2012), mostrou que TcZFP1 possui afinidade por sequências ricas em citosina e TcZPF2 por sequências ricas em adeninas, reforçando a importância dos motivos identificados neste trabalho como sendo possíveis sítios de ligação da TcRBP99. Contudo, nenhum dos motivos identificados é comum a todas as 3' UTRs analisadas, o que nos leva a pensar na possibilidade de ocorrência de interações com outras proteínas ou mesmo no efeito

indireto dessa RBP ao estabilizar um primeiro mRNA, o qual, de forma direta, estabilizaria os mRNAs identificados na análise de expressão diferencial.

Ademais, quando os motivos foram alinhados, a fim de observar padrões de ocorrência, apenas dois mRNAs correspondentes a genes das RHS (TcCLB.505943.110 e TcCLB.509559.20) se alinharam, divergindo pouco em relação as posições dos motivos encontrados. Este resultado reforça a ideia da existência de outros elementos que estejam atuando nos mRNAs concomitantemente a TcRBP99, ou seja, outras proteínas podem estar formando um complexo com esta RBP. A hipótese levantada aqui é a de que existem interações proteína-proteína que podem estar influenciando na ligação aos mRNAs alvos da RBP em questão e consequentemente na regulação estágio-específico dos mesmos.

Para avaliar a atuação de forma mais abrangente nos genes identificados como potenciais alvos da TcRBP99 são necessários estudos de interação proteína-RNA e também proteína-proteína para esta RBP, sendo possível ter uma visão mais precisa do seu papel na regulação da expressão dos mesmos e seu papel em processos previamente identificados, como diferenciação e proliferação em formas epimastigotas. Para isso existem metodologias de alta precisão validadas para esses estudos, como é o caso das técnicas de CLIP-seq e ensaio de desvio de mobilidade eletroforética (do inglês *Electrophoretic Mobility Shift Assay* (EMSA)), as quais são usadas para analisar interações proteicas com o RNA. Dessa forma torna-se possível localizar sítios de ligação das proteínas ou ainda, indicar se há a presença de outras proteínas formando um complexo.

Espectrometria de massa e SDS-PAGE, técnicas já citadas e outras como *pull-down*, que consiste na imobilização de proteínas que interagem com a proteína conhecida por meio de interações proteína-proteína em uma purificação de afinidade, podem ser realizadas para avaliar interações proteína-proteína (BERNÁ et al., 2017; EINARSON, 2001; OUNA et al., 2012). Co-imunoprecipitação também é passível de identificar interações proteína-proteína, mas baseia-se na utilização de anticorpos para a identificação de interações (LIN; LAI, 2017).

Métodos computacionais são bastante empregados para a predição de interações proteína-proteína, sendo métodos alternativos para estas análises. Buscas em bancos de dados de interações como o STRING ajudam a construir relações verdadeiras entre proteínas (RAO et al., 2014). Mas, existem ainda, algoritmos

baseados nas sequências da proteína que buscam identificar proteínas homólogas com potencial grau de interação com a proteína de interesse, neste trabalho a TcRBP99. Outros métodos mais complexos envolvem mineração de texto (*text mining*), aprendizagem de máquina (*machine learning*) ou redes neurais artificiais para a realização de buscas por interações proteína-proteína (ZAHIRI et al, 2013).

Sendo assim, com base nos resultados encontrados neste trabalho é possível propor um papel da TcRBP99 na estabilização de mRNAs que estão aumentados nas formas epimastigotas e que por sua vez estão envolvidos em processos essenciais para a sobrevivência das mesmas. A hipótese de que a TcRBP99 estabiliza mRNAs que contribuem para o aumento da proliferação nas formas epimastigotas é levantada aqui. Outra hipótese levantada é a de que devido a maior ploriferação, as epimastigotas acabam tendo a diferenciação para tripomastigotas metacíclicas reduzida ou inibida. Essas hipóteses são sustentadas pelo fato de que epimastigotas necessitam permanecer tempo suficiente no intestino médio do vetor a fim de se multiplicarem de forma a aumentar a eficiência da infecção e prolonga-la no hospedeiro vertebrado (VAZ, 2017). Os genes possivelmente relacionados a esses fenótipos são os identificados nesse trabalho como tendo expressão reduzida pela TcRBP99. O gene PAD em formas epimastigotas WT parece ser essencial para evitar a rápida diferenciação dessas formas de vida (Figura 38A). O gene relacionado a transportadores de aminoácidos indica ter um papel também importante para a ploriferação de epimastigotas, que tem metabolismo baseado nesse nutriente (CANEPA et al., 2004, URBINA; AZAVACHE, 1984) (Figura 38B). Com a maior expressão de transportadores de aminoácidos os parasitas têm maior eficiência na proliferação, para que somente depois de algum tempo se diferenciem em tripomastigotas metacíclicas.

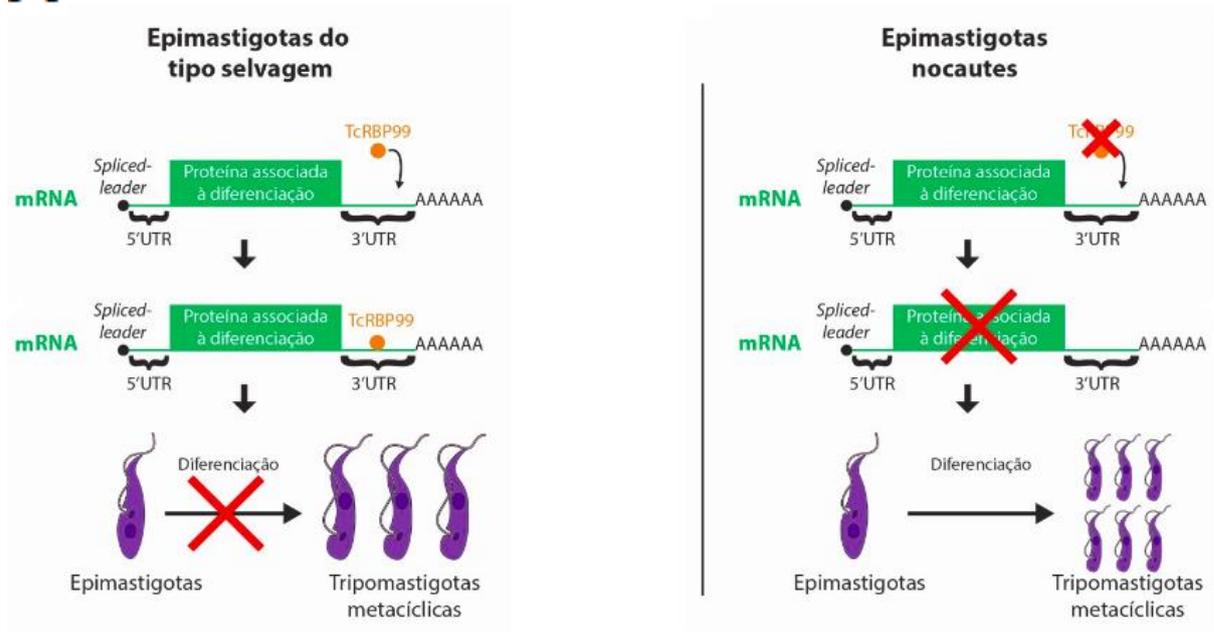
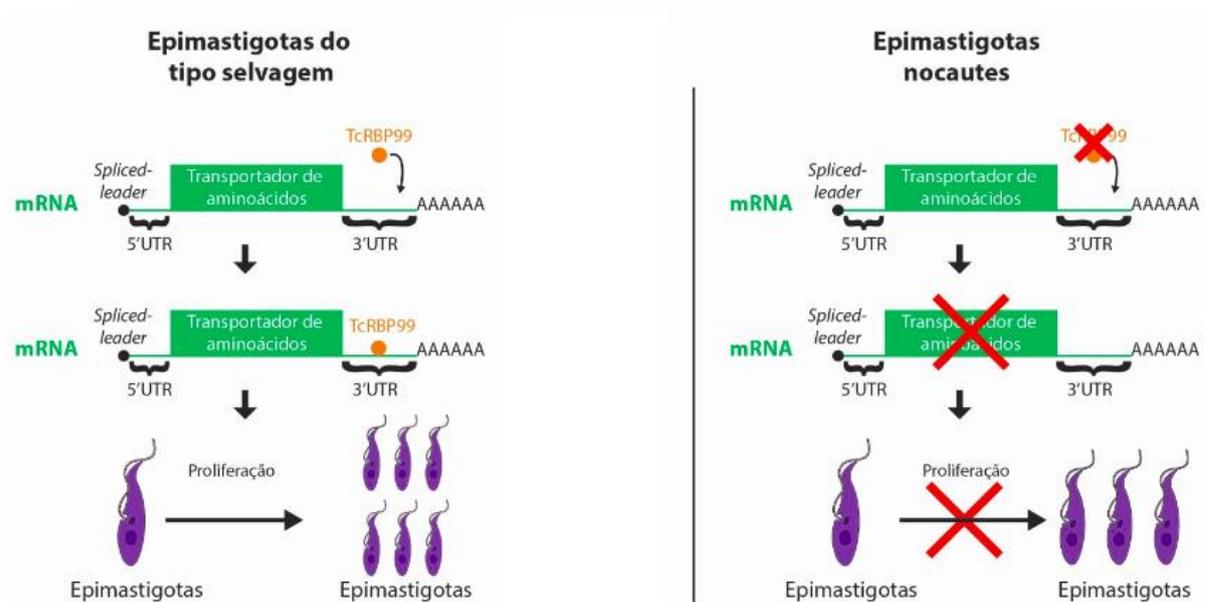
A**B**

Figura 38. Epimastigotas do tipo selvagem versus epimastigotas nocautes para a TcRBP99. (A) Em epimastigotas do tipo selvagem (lado esquerdo da figura), a TcRBP99 parece regular de forma positiva mRNAs que codificam para a proteína associada a diferenciação, estabilizando-os através de sua ligação à região 3' UTR. Dessa forma, a expressão deste gene funciona normalmente, o que parece resultar na inibição da diferenciação de epimastigotas em tripomastigotas metacíclicas. Diferentemente, em epimastigotas nocautes para a TcRBP99 (lado direito da figura), os níveis de expressão do gene TcCLB.506551.10 é reduzido, pois não ocorre a ligação da RBP à região 3' UTR. Sendo assim, na ausência da proteína associada à diferenciação, a taxa de diferenciação de epimastigotas em tripomastigotas metacíclicas aumenta. (B) Genes que codificam para transportadores de aminoácidos poderiam estar também sendo estabilizados pela TcRBP99, quando a mesma se liga à motivos presentes na 3' UTR, de forma a simular a situação anterior. A estabilização desses genes e consequente aumento da expressão levariam ao aumento da proliferação de epimastigotas. Em situação contrária

(lado direito da figura), ou seja, na ausência da RBP, os genes que codificam para transportadores de aminoácidos teriam sua expressão reduzida e conseqüente redução da taxa de proliferação.

A identificação de uma RBP que se encontra aumentada nas formas epimastigotas, assim como de seus potenciais alvos, que também mostraram estar aumentados nessa forma de vida, confirma a existência de mecanismos de regulação gênica estágio-específica, dependente de elementos presentes em regiões UTRs e da interação destes com proteínas como a TcRBP99, que é composta por um domínio capaz de se ligar aos motivos de sequência de mRNAs. Além disso, a ação indireta de RBPs foi colocada em questão, levantando a hipótese de que essas proteínas podem formar complexos com outras proteínas ou ainda desempenhar seu papel por meio da ligação a mRNAs primários com conseqüente efeito estabilizador em um mRNA secundário. Por isso, esse e outros estudos, desenvolvidos com o auxílio de ferramentas computacionais são importantes para a expansão do conhecimento acerca da biologia do *T. cruzi* e de outros tripanossomatídeos.

Outros estudos baseados no transcriptoma de *T. cruzi* estão sendo desenvolvidos em nosso laboratório, com o objetivo de avaliar o papel de outras RBPs na regulação da expressão gênica. Assim, pretendemos identificar outros elementos regulatórios e processos biológicos importantes nas diferentes formas de vida de *T. cruzi*, de modo a contribuir para o conhecimento acerca do parasita e fornecer bases sólidas para o desenvolvimento de novas terapias e/ou novos fármacos para o tratamento da doença de chagas.

7. PERSPECTIVAS

- Validar por qRT-PCR a expressão dos 10 genes identificados como tendo expressão alterada pela TcRBP99;
- Gerar linhagens nas quais o gene TcRBP99 é re-inserido no genoma de linhagens nocautes (*add-back*) e avaliar a expressão dos genes que codificam PAD e transportador de aminoácidos para verificar se ocorre reversão dos níveis de mRNAs;
- Realizar experimentos de metaciclogênese *in vitro* e *in vivo*, por meio da infecção de vetores com as formas epimastigotas WT e nocautes;
- Validar por co-imunoprecipitação, CLIP-seq e EMSA a ligação da TcRBP99 aos mRNAs correspondentes aos genes que codificam PAD, transportador de aminoácidos e proteínas RHS;
- Investigar o papel da proteína associada à diferenciação como possível transdutora de sinal para a diferenciação de tripomastigotas para epimastigotas por meio da geração de nocaute desse gene;
- Identificar por espectrometria de massa, SDS-PAGE e metodologias computacionais, possíveis interações proteína-proteína para a TcRBP99;
- Confirmar o tamanho das regiões 3' UTRs dos possíveis mRNAs alvos da TcRBP99 por meio da amplificação dessas regiões, utilizando uma metodologia que envolve oligos-dT como *primer*, e posterior sequenciamento;
- Verificar a interação da TcRBP99 com regiões 3' UTR ricas em C e regiões ricas em U.

REFERÊNCIAS BIBLIOGRÁFICAS

- AFROZ, T. et al. One, two, three, four! How multiple RRM's read the genome sequence. **Methods in Enzymology**, v. 558, n. 1, p. 235–278, 2015.
- ALARCÓN DE NOYA, B. et al. Large Urban Outbreak of Orally Acquired Acute Chagas Disease at a School in Caracas, Venezuela. **The Journal of Infectious Diseases**, v. 201, n. 9, p. 1308–1315, 2010.
- ALCANTARA, M. V. et al. Knockout of the CCCH zinc finger protein TcZC3H31 blocks *Trypanosoma cruzi* differentiation into the infective metacyclic form. **Molecular and Biochemical Parasitology**, v. 221, p. 1–9, abr. 2018.
- ALVES, L. R. et al. Proteomic analysis reveals the dynamic association of proteins with translated mRNAs in *Trypanosoma cruzi*. **Gene**, v. 452, n. 2, p. 72–78, 2010.
- ALVES, L. R. et al. The mRNAs associated to a zinc finger protein from *Trypanosoma cruzi* shift during stress conditions. **RNA Biology**, v. 11, n. 7, p. 921–933, 2014.
- ALVES, L. R.; GOLDENBERG, S. RNA-binding proteins related to stress response and differentiation in protozoa. **World journal of biological chemistry**, v. 7, n. 1, p. 78–87, 2016.
- ALY, R. et al. A regulatory role for the 5' and 3' untranslated regions in differential expression of hsp83 in Leishmania. **Nucleic acids research**, v. 22, n. 15, p. 2922–9, 11 ago. 1994.
- ANDERS, S.; HUBER, W. Differential expression analysis for sequence count data. **Genome biology**, v. 11, n. 10, p. R106, 2010.
- ANDERS, S. et al. HTSeq-A Python framework to work with high-throughput sequencing data. **Bioinformatics**, v. 31, n. 2, p. 166–169, 2015.
- ANDREWS, S. **FastQC: a quality control tool for high throughput sequence data**. Disponível em: <<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>>. Acesso em: 3 mar. 2017.
- ARAÚJO, P. R. et al. Development of a dual reporter system to identify regulatory cis-acting elements in untranslated regions of *Trypanosoma cruzi* mRNAs. **Parasitology International**, v. 60, n. 2, p. 161–169, 1 jun. 2011.

- ARAÚJO, P. R.; TEIXEIRA, S. M. Regulatory elements involved in the post-transcriptional control of stage-specific gene expression in *Trypanosoma cruzi* - A review. **Memorias do Instituto Oswaldo Cruz**, v. 106, n. 3, p. 257–266, 2011.
- ASLETT, M. et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. **Nucleic acids research**, v. 38, n. Database issue, p. D457-62, jan. 2010.
- ATWOOD, J. A. The *Trypanosoma cruzi* Proteome. **Science**, v. 309, n. 5733, p. 473–476, 2005.
- BARTHOLOMEU, D. C. et al. *Trypanosoma cruzi*: RNA structure and post-transcriptional control of tubulin gene expression. **Experimental Parasitology**, v. 102, n. 3–4, p. 123–133, 2002.
- BATISTA, J. A et al. Characterization of a *Trypanosoma cruzi* poly(A)-binding protein and its genes. **Molecular and biochemical parasitology**, v. 67, n. 2, p. 301–12, 1994.
- BAYER-SANTOS, E. et al. Regulatory elements in the 3' untranslated region of the GP82 glycoprotein are responsible for its stage-specific expression in *Trypanosoma cruzi* metacyclic trypomastigotes. **Acta Tropica**, v. 123, n. 3, p. 230–233, 2012.
- BAILEY, T. L. et al. The MEME Suite. **Nucleic acids research**, v. 43, n. W1, p. W39-49, 1 jul. 2015.
- BELEW, T. et al. Comparative transcriptome profiling of virulent and non-virulent *Trypanosoma cruzi* underlines the role of surface proteins during infection. **PLoS Pathogens**, v. 13, n. 12, p. 1–23, 2017.
- BENABDELLAH, K. et al. Alternative trans-splicing of the *Trypanosoma cruzi* LYT1 gene transcript results in compartmental and functional switch for the encoded protein. **Molecular Microbiology**, v. 65, n. 6, p. 1559–1567, 2007.
- BENCHIMOL BARBOSA, P. R. The oral transmission of Chagas' disease: An acute form of infection responsible for regional outbreaks. **International Journal of Cardiology**, v. 112, n. 1, p. 132–133, 2006.
- BENZ, C. et al. Messenger RNA processing sites in *Trypanosoma brucei*. **Molecular and Biochemical Parasitology**, v. 143, n. 2, p. 125–134, 2005.
- BERN, C. M. D. M. P. Chagas' Disease. **New England Journal of Medicine**, v. 373, n. 5, p. 456–466, 2015.

- BERNÁ, L. et al. Transcriptomic analysis reveals metabolic switches and surface remodeling as key processes for stage transition in *Trypanosoma cruzi*. **PeerJ**, v. 5, p. e3017, 2017.
- BOLGER, A. M. et al. Trimmomatic: A flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114–2120, 2014.
- BRANCHE, C. et al. Comparative karyotyping as a tool for genome structure analysis of *Trypanosoma cruzi*. **Molecular and Biochemical Parasitology**, v. 147, n. 1, p. 30–38, 2006.
- BRANDÃO, A. et al. Identification of Transcribed Sequences (ESTs) in the *Trypanosoma cruzi* Genome Project. **Memorias do Instituto Oswaldo Cruz**, v. 92, n. 6, p. 863–866, 1997.
- BRANDÃO, A.; JIANG, T. The composition of untranslated regions in *Trypanosoma cruzi* genes. **Parasitology International**, v. 58, n. 3, p. 215–219, 2009.
- BRENER, Z. Intraspecific variations in *Trypanosoma cruzi*: two types of parasite populations presenting distinct characteristics. **Revista do Instituto de Medicina Tropical de Sao Paulo**, v. 347, n. Sci Pub, p. 209–219, 1977.
- BRENER, Z.; CHIARI, E. Variações morfológicas observadas em diferentes amostras de *Trypanosoma cruzi*. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 5, n. 5, p. 220–224, 1963.
- BRINGAUD, F. et al. A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*. **Eukaryotic cell**, v. 1, n. 1, p. 137–51, 1 fev. 2002.
- BRINGAUD, F. et al. Energy metabolism of trypanosomatids: Adaptation to available carbon sources. **Molecular & Biochemical Parasitology**, v. 149, p. 1–9, 2006.
- BUCKNER, F. S. et al. Induction of resistance to azole drugs in *Trypanosoma cruzi*. **Antimicrobial Agents and Chemotherapy**, v. 42, n. 12, p. 3245–3250, 1998.
- BUEKENS, P. et al. Congenital transmission of *Trypanosoma cruzi* in Argentina, Honduras, and Mexico: Study protocol. **Reproductive Health**, v. 10, n. 1, p. 1–10, 2013.

- BULLARD, J. H. et al. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. **BMC Bioinformatics**, v. 11, n. 1, p. 94, 18 fev. 2010.
- CAMPOS, P. C. et al. Sequences involved in mRNA processing in *Trypanosoma cruzi*. **International Journal for Parasitology**, v. 38, n. 12, p. 1383–1389, 2008.
- CANEPA, G. et al. Biochemical characterization of a low-affinity arginine permease from the parasite *Trypanosoma cruzi*. **FEMS Microbiology Letters**, v. 236, n. 1, p. 79–84, 1 jul. 2004.
- CANO, M. I. et al. Molecular karyotype of clone CL Brener chosen for the *Trypanosoma cruzi* Genome Project. **Molecular and Biochemical Parasitology**, v. 71, n. 2, p. 273–278, 1995.
- CARO, F. et al. *Trypanosoma cruzi*: Analysis of the complete PUF RNA-binding protein family. **Experimental Parasitology**, v. 113, n. 2, p. 112–124, 2006.
- CENTERS FOR DISEASE CONTROL AND PREVENTION, C. **Life cycle**. Disponível em: <<http://www.cdc.gov/parasites/chagas/biology.html>>. Acesso em: 20 dez. 2017.
- CERQUEIRA, G. C. et al. Analysis of expressed sequence tags from *Trypanosoma cruzi* amastigotes. **Memorias do Instituto Oswaldo Cruz**, v. 100, n. 4, p. 385–389, 2005.
- CHAUDHURI, N.; CHAUDHURI, N. Interventions to improve children's health by improving the housing environment. [Review] [96 refs]. **Reviews on Environmental Health**, v. 19, n. 3–4, p. 197–222, 2004.
- CLAYTON, C.; SHAPIRA, M. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. **Mol Biochem Parasitol**, v. 156, n. 2, p. 93–101, 2007.
- CLÉRY, A.; ALLAIN, F. H. T. A Structural biology perspective of proteins involved in splicing regulation. In: **Alternative pre-mRNA Splicing: Theory and Protocols**. [s.l.: s.n.]. p. 33–48.
- CLOONAN, N. et al. Stem cell transcriptome profiling via massive-scale mRNA sequencing. **Nature Methods**, v. 5, n. 7, p. 613–619, 30 jul. 2008.

CONNERS, E. E. et al. Chagas disease infection among migrants at the Mexico/Guatemala border. **American Journal of Tropical Medicine and Hygiene**, v. 97, n. 4, p. 1134–1140, 2017.

CORDOVA, E. et al. Reactivation of Chagas disease with central nervous system involvement in HIV-infected patients in Argentina, 1992–2007. **International Journal of Infectious Diseases**, v. 12, n. 6, p. 587-592, 2008.

COUGHLIN, B. C. et al. Amastin mRNA abundance in *Trypanosoma cruzi* is controlled by a 3'-untranslated region position-dependent cis-element and an untranslated region-binding protein. **The Journal of biological chemistry**, v. 275, n. 16, p. 12051–60, 21 abr. 2000.

COURA, J. et al. Comparative controlled study on the use of benznidazole, nifurtimox and placebo, in the chronic form of Chagas' disease, in a field area with interrupted transmission. I. Preliminary evaluation. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 30, n. 2, p. 139–44, 1997.

COURA, J. R. **Evolutionary pattern in Chagas' disease and the life span of *Trypanosoma cruzi* in human infection.** International Symposium on New Approaches in American Trypanosomiasis. **Anais...**Belo Horizonte: PAHO Sci, 1975.

COURA, J. R. Transmission of chagasic infection by oral route in the natural history of Chagas disease. **Rev Soc Bras Med Trop**, v. 39 Suppl 3, p. 113–117, 2006.

CRIBB, P. et al. *Trypanosoma cruzi* TBP shows preference for C/G-rich DNA sequences in vitro. **Experimental Parasitology**, v. 124, n. 3, p. 346–349, 2010.

CRIBB, P.; SERRA, E. One- and two-hybrid analysis of the interactions between components of the *Trypanosoma cruzi* spliced leader RNA gene promoter binding complex. **International Journal for Parasitology**, v. 39, n. 5, p. 525–532, 2009.

D'ORSO, I.; FRASCH, A. C. C. TcUBP-1, a Developmentally Regulated U-rich RNA-binding Protein Involved in Selective mRNA Destabilization in Trypanosomes. **Journal of Biological Chemistry**, v. 276, n. 37, p. 34801–34809, 2001.

D'ORSO, I.; FRASCH, A. C. C. TcUBP-1, an mRNA destabilizing factor from trypanosomes, homodimerizes and interacts with novel AU-rich element- and poly(A)-

binding proteins forming a ribonucleoprotein complex. **Journal of Biological Chemistry**, v. 277, n. 52, p. 50520–50528, 2002.

D'ORSO et al. RNA-binding proteins and mRNA turnover in trypanosomes. **Trends in Parasitology**, v. 4, n. 4, p. 151-155, 2003.

DALLAGIOVANNA, B. et al. Trypanosoma cruzi: Molecular characterization of TcPUF6, a Pumilio protein. **Experimental Parasitology**, v. 109, n. 4, p. 260–264, 2005.

DALLAGIOVANNA, B. et al. Functional genomic characterization of mRNAs associated with TcPUF6, a pumilio-like protein from *Trypanosoma cruzi*. **Journal of Biological Chemistry**, v. 283, n. 13, p. 8266–8273, 2008.

DE AVALOS, S. V. et al. Immediate/Early Response to *Trypanosoma cruzi* Infection Involves Minimal Modulation of Host Cell Transcription. **Journal of Biological Chemistry**, v. 277, n. 1, p. 639–644, 4 jan. 2002.

DE GAUDENZI, J. G. et al. Insights into the Regulation of mRNA Processing of Polycistronic Transcripts Mediated by DRBD4/PTB2, a Trypanosome Homolog of the Polypyrimidine Tract-Binding Protein. **Journal of Eukaryotic Microbiology**, v. 63, n. 4, p. 440–452, 2016a.

DE GAUDENZI, J. G. et al. Insights into the Regulation of mRNA Processing of Polycistronic Transcripts Mediated by DRBD4/PTB2, a Trypanosome Homolog of the Polypyrimidine Tract-Binding Protein. **The Journal of eukaryotic microbiology**, v. 63, n. 4, p. 440–452, 2016b.

DE GAUDENZI, J. G. et al. RNA recognition motif-type RNA-binding proteins in *Trypanosoma cruzi* form a family involved in the interaction with specific transcripts in vivo. **Journal of Biological Chemistry**, v. 278, n. 21, p. 18884–18894, 2003.

DE GAUDENZI J. et al. RNA-binding domain proteins in Kinetoplastids: A comparative analysis. **Eukaryot Cell**, vol. 4, n. 12, p. 2106–2114, 2005.

DE OLIVEIRA, R. B. et al. Gastrointestinal manifestations of Chagas' disease. **The American journal of gastroenterology**, v. 93, n. 6, p. 884–9, 1998.

DEAN, S. et al. A surface transporter family conveys the trypanosome differentiation signal. **Nature**, v. 459, n. 7244, p. 213–7, 14 maio 2009.

DEGRAVE, W. et al. Parasite Genome Projects and the *Trypanosoma cruzi* Genome. **Memorias do Instituto Oswaldo Cruz**, v. 92, n. 6, p. 859–862, 1997.

DIAS, J. C. The indeterminate form of human chronic Chagas' disease: A clinical epidemiological review. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 22, n. 3, p. 147–56, 1989.

DIAS, J. P. et al. Acute Chagas disease outbreak associated with oral transmission. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 41, n. 3, p. 296–300, 2008.

DILLON, L. A. L. et al. Transcriptomic profiling of gene expression and RNA processing during *Leishmania major* differentiation. **Nucleic acids research**, v. 43, n. 14, p. 6799–813, 18 ago. 2015.

DNDI, D. FOR N. D. INITIATIVE. Securing access to current treatments while researching better options. 2015.

DUHAGON, M. A.; DALLAGIOVANNA, B.; GARAT, B. Unusual features of poly[dT-dG]·[dC-dA] stretches in CDS-flanking regions of *Trypanosoma cruzi* genome. **Biochemical and Biophysical Research Communications**, v. 287, n. 1, p. 98–103, 2001.

DVORAK, J. The natural heterogeneity of *Trypanosoma cruzi*: Biological and medical implications. **Journal of Cellular Biochemistry**, v. 24, n. 4, p. 357–371, 1984.

EDGER, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic acids research**, v. 32, n. 5, p. 1792-1797, 2004.

EDWARDS, T. A. Bespoke RNA recognition by Pumilios. **Biochemical Society Transactions**, v. 43, n. 5, p. 801–806, 2015.

EICKHOFF, C. S. et al. Comparison of the infectivity of *Trypanosoma cruzi* insect-derived metacyclic trypomastigotes after mucosal and cutaneous contaminative challenges. **Memorias do Instituto Oswaldo Cruz**, v. 108, n. 4, p. 508–511, 2013.

EINARSON, M. **Detection of Protein-Protein Interactions Using the GST Fusion Protein Pulldown Technique**. 3 ed ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 2001.

EL-SAYED, N. M. et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. **Science**, v. 309, n. July, p. 409–415, 2005a.

EL-SAYED, N. M. et al. Comparative genomics of trypanosomatid parasitic protozoa.

ENGMAN, D. M. et al. *Trypanosoma cruzi* exhibits inter- and intra-strain heterogeneity in molecular karyotype and chromosomal gene location. **Molecular and Biochemical Parasitology**, v. 22, n. 2–3, p. 115–123, 1987. **Science**, v. 309, n. 5733, p. 404–9, 2005.

FERREIRA, C. S. et al. Pasteurization of human milk to prevent transmission of Chagas disease. **Revista do Instituto de Medicina Tropical de Sao Paulo**, v. 43, n. 3, p. 161–162, 2001.

FILARDI, L. S.; BRENER, Z. Susceptibility and natural resistance of *Trypanosoma cruzi* strains to drugs used clinically in Chagas disease. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 81, n. 5, p. 755–9, 1987.

FINOTELLO, F.; DI CAMILLO, B. Measuring differential gene expression with RNA-seq: challenges and strategies for data analysis. **Briefings in Functional Genomics**, v. 14, n. 2, p. 130–142, 1 mar. 2015.

GALINDO, M. et al. Chromosomal size conservation through the cell cycle supports karyotype stability in *Trypanosoma cruzi*. **FEBS Letters**, v. 581, n. 10, p. 2022–2026, 2007.

GOMES, G. G. et al. TcRRMs and TcP28 genes are intercalated and differentially expressed in *Trypanosoma cruzi* life cycle. **Biochemical and Biophysical Research Communications**, v. 322, n. 3, p. 985–992, 2004.

GRANT, I. H. et al. Transfusion-associated acute Chagas disease acquired in the United States. **Annals of internal medicine**, v. 111, n. 10, p. 849–51, 15 nov. 1989.

GROOM, Z. C. et al. Tropical diseases of the myocardium: A review. **International Journal of General Medicine**, v. 10, p. 101–111, 2017.

GUERRA-SLOMPO, E. P. et al. Molecular characterization of the *Trypanosoma cruzi* specific RNA binding protein TcRBP40 and its associated mRNAs. **Biochemical and Biophysical Research Communications**, v. 420, n. 2, p. 302–307, 2012.

GUO et al. Evaluation of read count based RNAseq analysis methods. **BMC Genomics**, v. 14, n. Suppl 8, 2013.

GUPTA, S. K. et al. Basal splicing factors regulate the stability of mature mRNAs in trypanosomes. **The Journal of biological chemistry**, v. 288, n. 7, p. 4991–5006, 15 fev. 2013.

HAILE, S.; PAPADOPOULOU, B. Developmental regulation of gene expression in trypanosomatid parasitic protozoa. **Current Opinion in Microbiology**, v. 10, n. 6, p. 569–577, 2007.

HAN, Y. et al. Advanced Applications of RNA Sequencing and Challenges. **Bioinformatics and biology insights**, v. 9, n. Suppl 1, p. 29–46, 2015.

HELLMAN, K. et al. *Trypanosoma brucei* RNA binding proteins p34 and p37 mediate NOPP44/46 cellular localization via the exportin 1 nuclear export pathway. **Eukaryotic Cell**, v. 6, n. 12, p. 2206–2213, 2007.

HENDRIKS, E. F. et al. A novel CCCH protein which modulates differentiation of *Trypanosoma brucei* to its procyclic form. **EMBO Journal**, v. 20, n. 23, p. 6700–6711, 2001.

HENDRIKS, E. F.; MATTHEWS, K. R. Disruption of the developmental programme of *Trypanosoma brucei* by genetic ablation of TbZFP1, a differentiation-enriched CCCH protein. **Molecular microbiology**, v. 57, n. 3, p. 706–16, 30 ago. 2005.

HENDRIKS, E. F.; MATTHEWS, K. **Post-transcriptional control of gene expression in African trypanosomes**. In **Trypanosomes: After The Genome**. Norfolk: Orizon Bioscience, 2007.

HENRIKSSON, J. et al. Chromosomal localization of seven cloned antigen genes provides evidence of diploidy and further demonstration of karyotype variability in *Trypanosoma cruzi*. **Molecular and Biochemical Parasitology**, v. 42, n. 2, p. 213–223, 1990.

HENRIKSSON, J. et al. Karyotype variability in *Trypanosoma cruzi*. **Parasitology Today**, v. 12, n. 3, p. 108–114, 1996.

- HERNANDEZ-RIVAS, R.; SCHERF, A. Separation and Mapping of Chromosomes of Parasitic Protozoa. **Memorias do Instituto Oswaldo Cruz**, v. 92, n. 6, p. 815–819, 1997.
- HERWALDT, B. Laboratory-Acquired Parasitic Infections from Accidental Exposures. **Clin Microbiol Rev**, v. 14, n. 4, p. 659–88, 2001.
- HOFFLIN, J. M. et al. Laboratory-acquired Chagas disease. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 81, n. 3, p. 437–10, 1987.
- HORN, D.; BARRY, J. D. The central roles of telomeres and subtelomeres in antigenic variation in African trypanosomes. **Chromosome Research**, v. 13, n. 5, p. 525–533, 25 jul. 2005.
- HOWARD, E. J. et al. Frequency of the congenital transmission of *Trypanosoma cruzi*: A systematic review and meta-analysis. **BJOG: An International Journal of Obstetrics and Gynaecology**, v. 121, n. 1, p. 22–33, 2014.
- HUMMEL, H. S. et al. Mutational analysis of 3' splice site selection during trans-splicing. **The Journal of biological chemistry**, v. 275, n. 45, p. 35522–35531, 2000.
- ILLUMINA. NovaSeq 6000 Sequencing System. p. 1–4, 2017.
- ISAKOV, O.; SHOMRO, N. Deep Sequencing Data Analysis: Challenges and Solutions. In: **Bioinformatics - Trends and Methodologies**. [s.l.] InTech, 2011.
- JACKSON, A. P. The evolution of amastin surface glycoproteins in trypanosomatid parasites. **Molecular Biology and Evolution**, v. 27, n. 1, p. 33–45, 2010.
- JACKSON, Y. et al. Congenital Transmission of Chagas Disease in Latin American Immigrants in Switzerland. **Emerging Infectious Diseases**, v. 15, n. 4, p. 601–603, abr. 2009.
- JACKSON, Y. et al. Economic crisis and increased immigrant mobility: new challenges in managing Chagas disease in Europe. **Bulletin of the World Health Organization**, v. 92, n. 10, p. 771–772, 1 out. 2014.
- KANEHISA, M.; GOTO, S. KEGG: kyoto encyclopedia of genes and genomes. **Nucleic acids research**, v. 28, n. 1, p. 27–30, 1 jan. 2000.

KELLER, J. I. et al. Chagas disease vector blood meal sources identified by protein mass spectrometry. **Plos One**, v. 12, n.12, dezembro 2017.

KESSLER, D. A. et al. Results of lookback for Chagas disease since the inception of donor screening at New York Blood Center. **Transfusion**, v. 53, n. 5, p. 1083–1087, maio 2013.

KIM, D. et al. Telomere and subtelomere of *Trypanosoma cruzi* chromosomes are enriched in (pseudo)genes of retrotransposon hot spot and trans-sialidase-like gene families: the origins of *T. cruzi* telomeres. **Gene**, v. 346, p. 153–161, 14 fev. 2005.

KIM, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. **Genome biology**, v. 14, n. 4, p. R36, 25 abr. 2013.

KINOSHITA-YANAGA, A. T. et al. Accidental infection by *Trypanosoma cruzi* follow-up by the polymerase chain reaction: case report. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 51, n. 5, p. 295–298, 2009.

KISHORE, S. et al. Deciphering the role of RNA-binding proteins in the post-transcriptional control of gene expression. **Briefings in Functional Genomics**, v. 9, n. 5–6, p. 391–404, 2010.

KOBOLDT, D. C. et al. Challenges of sequencing human genomes. **Briefings in Bioinformatics**, v. 11, n. 5, p. 484–498, 1 set. 2010.

KOLEV, N. G. et al. The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. **PLoS Pathogens**, v. 6, n. 9, p. 1–15, 2010.

KRAMER, S. et al. Differential Localization of the Two *T. brucei* Poly(A) Binding Proteins to the Nucleus and RNP Granules Suggests Binding to Distinct mRNA Pools. **PLoS ONE**, v. 8, n. 1, 2013.

KRAMER, S.; CARRINGTON, M. Trans-acting proteins regulating mRNA maturation, stability and translation in trypanosomatids. **Trends in Parasitology**, v. 27, n. 1, p. 23–30, 2011.

LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nature methods**, v. 9, n. 4, p. 357–9, 4 mar. 2012.

- LARANJA, F. et al. Chagas' disease; a clinical, epidemiologic, and pathologic study. **Circulation**, v. 14, n. December, p. 1035–1060, 1956.
- LAW, C. W. et al. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. **Genome biology**, v. 15, n. 2, p. R29, 3 fev. 2014.
- LAWRIE, N. R. Studies in the metabolism of protozoa: The nitrogenous metabolism and respiration of *Bodo caudatus*. **The Biochemical journal**, v. 29, n. 3, p. 588–98, mar. 1935.
- LEBOWITZ et al. Coupling of poly(A) site selection and trans-splicing in *Leishmania*. **Genes & Development**, v. 7, n. 6, p. 996-1007, 1993.
- LESCURE, F. X. et al. Chagas disease: Changes in knowledge and management. **The Lancet Infectious Diseases**, v. 10, n. 8, p. 556–570, 2010.
- LI, H. et al. The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078–2079, 2009.
- LI, Y. et al. Transcriptome Remodeling in *Trypanosoma cruzi* and Human Cells during Intracellular Infection. **PLoS Pathogens**, v. 12, n. 4, p. 1–30, 2016.
- LIANG, X. H. et al. Trans and cis splicing in trypanosomatids: Mechanism, factors, and regulation. **Eukaryotic Cell**, v. 2, n. 5, p. 830–840, 2003.
- LIN, J. S.; LAI, E. M. Protein–Protein Interactions: Co-Immunoprecipitation. In: **Methods in molecular biology (Clifton, N.J.)**. [s.l: s.n.]. v. 1615p. 211–219.
- LIRA, C. B. B. et al. Telomere biology of trypanosomatids: beginning to answer some questions. **Trends in Parasitology**, v. 23, n. 8, p. 357–362, ago. 2007.
- LOVE, M. I. et al. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome biology**, v. 15, n. 12, p. 550, 2014.
- LUNDE, B. M. et al. RNA-binding proteins: Modular design for efficient function. **Nature Reviews Molecular Cell Biology**, v. 8, n. 6, p. 479–490, 2007.
- MACKEY, T. K. et al. Emerging and reemerging neglected tropical diseases: A review of key characteristics, Risk factors, and the Policy and innovation environment. **Clinical Microbiology Reviews**, v. 27, n. 4, p. 949–979, 2014.

- MACKEY, T. K.; LIANG, B. A. Threats from emerging and re-emerging neglected tropical diseases (NTDs). **Infection ecology & epidemiology**, v. 2, p. 1–7, 2012.
- MAIR, G. et al. A new twist in trypanosome RNA metabolism: Cis-splicing of pre-mRNA. **RNA**, v. 6, n. 2, p. 163–169, 2000.
- MANNING-CELA, R. et al. Alternative Splicing of LYT1 Transcripts in *Trypanosoma cruzi* Alternative Splicing of LYT1 Transcripts in *Trypanosoma cruzi*. v. 70, n. 8, p. 12–15, 2002.
- MARCILI, A. et al. *Trypanosoma cruzi* in Brazilian Amazonia: Lineages TCI and TCIIa in wild primates, *Rhodnius* spp. and in humans with Chagas disease associated with oral transmission. **International Journal for Parasitology**, v. 39, n. 5, p. 615–623, 2009a.
- MARCILI, A. et al. A new genotype of *Trypanosoma cruzi* associated with bats evidenced by phylogenetic analyses using SSU rDNA, cytochrome b and Histone H2B genes and genotyping based on ITS1 rDNA. **Parasitology**, v. 136, n. 6, p. 641–55, 16 maio 2009b.
- MARIONI, J. C. et al. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. **Genome Research**, v. 18, n. 9, p. 1509–1517, 30 jul. 2008.
- MARIS, C. et al. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. **FEBS Journal**, v. 272, n. 9, p. 2118–2131, 2005.
- MARTIN, M. **cutadapt Documentation**, 2015. Disponível em: <<https://media.readthedocs.org/pdf/cutadapt/stable/cutadapt.pdf>>. Acesso em: 2 jan. 2017
- MARTÍNEZ-CALVILLO, S. et al. Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. **Molecular Cell**, v. 11, n. 5, p. 1291–1299, 2003.
- MELO, R. C.; BRENER, Z. Tissue tropism of different *Trypanosoma cruzi* strains. **The Journal of parasitology**, v. 64, n. 3, p. 475–82, jun. 1978.
- MINNING, T. A. et al. The steady-state transcriptome of the four major life-cycle stages of *Trypanosoma cruzi*. **BMC genomics**, v. 10, p. 370, 7 ago. 2009.

- MÖRKING, P. A. et al. TcZFP1: A CCCH zinc finger protein of *Trypanosoma cruzi* that binds poly-C oligoribonucleotides in vitro. **Biochemical and Biophysical Research Communications**, v. 319, n. 1, p. 169–177, 2004.
- MÖRKING, P. A. et al. The zinc finger protein TcZFP2 binds target mRNAs enriched during *Trypanosoma cruzi* metacyclogenesis. **Memorias do Instituto Oswaldo Cruz**, v. 107, n. 6, p. 790–799, 2012.
- MOROZOVA, O. et al. A. Applications of New Sequencing Technologies for Transcriptome Analysis. **Annual Review of Genomics and Human Genetics**, v. 10, n. 1, p. 135–151, set. 2009.
- MORTAZAVI, A. et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. **Nature Methods**, v. 5, n. 7, p. 621–628, 30 jul. 2008.
- MUTZ, K.-O. et al. Transcriptome analysis using next-generation sequencing. **Current Opinion in Biotechnology**, v. 24, n. 1, p. 22–30, fev. 2013.
- NAGALAKSHMI, U. et al. The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. **Science**, v. 320, n. 5881, p. 1344–1349, 6 jun. 2008.
- NAGULESWARAN, A. et al. *Trypanosoma brucei* RRM1 is a nuclear RNA-binding protein and modulator of chromatin structure. **mBio**, v. 6, n. 2, p. 1–11, 2015a.
- NEVES, D. et al. **Parasitologia Humana**. 12^a ed ed. São Paulo: Atheneu, 2011.
- NGUYEN, T.; WASEEM, M. **Chagas Disease (American Trypanosomiasis) - StatPearls - NCBI Bookshelf**, 2017.
- NILSSON, D. et al. Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of *Trypanosoma brucei*. **PLoS Pathogens**, v. 6, n. 8, p. 21–22, 2010.
- NÓBREGA, A. A. et al. Oral transmission of chagas disease by consumption of Açaí palm fruit, Brazil. **Emerging Infectious Diseases**, v. 15, n. 4, p. 653–655, 2009.
- NOIREAU, F. et al. *Trypanosoma cruzi*: Adaptation to its vectors and its hosts. **Veterinary Research**, v. 40, n. 2, 2009.

NOZAKI, T.; CROSS, G. A. M. Effects of 3' untranslated and intergenic regions on gene expression in *Trypanosoma cruzi*. **Molecular and Biochemical Parasitology**, v. 75, n. 1, p. 55–67, 1995.

OLIVARES, M. et al. Genomic clustering of the *Trypanosoma cruzi* nonlong terminal L1Tc retrotransposon with defined interspersed repeated DNA elements. **Electrophoresis**, v. 21, n. 14, p. 2973–2982, 1 ago. 2000.

OSHLACK, A. et al. From RNA-seq reads to differential expression results. **Genome Biology**, v. 11, n. 12, p. 220, 22 dez. 2010.

OUBRIDGE, C. et al. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. **Nature**, v. 372, n. 6505, p. 432–8, 1994.

OUNA, B. A. et al. The *Trypanosoma brucei* CCCH zinc finger proteins ZC3H12 and ZC3H13. **Molecular and Biochemical Parasitology**, v. 183, n. 2, p. 184–188, 2012.

PAIS, F. S. et al. Molecular characterization of ribonucleoproteic antigens containing repeated amino acid sequences from *Trypanosoma cruzi*. **Microbes and Infection**, v. 10, n. 7, p. 716–725, 2008.

PALENCHAR, J. B.; BELLOFATTO, V. Gene transcription in trypanosomes. **Molecular and Biochemical Parasitology**, v. 146, n. 2, p. 135–141, 2006.

PAN AMERICAN HEALTH ORGANIZATION, P. **Chagas disease**. Disponível em: <http://www.paho.org/hq/index.php?option=com_topics&view=article&id=10&Itemid=40743&lang=pt>. Acesso em: 19 dez. 2017.

PARK, S. T.; KIM, J. Trends in Next-Generation Sequencing and a New Era for Whole Genome Sequencing. **International Neurology Journal**, v. 20, n. Suppl 2, p. S76–83, 30 nov. 2016.

PARODI-TALICE, A. et al. Proteome analysis of the causative agent of Chagas disease: *Trypanosoma cruzi*. **International Journal for Parasitology**, v. 34, n. 8, p. 881–886, 1 jul. 2004.

PARSONS, M. et al. Trypanosome mRNAs share a common 5' spliced leader sequence. **Cell**, v. 38, n. 1, p. 309–316, 1984.

- PASTRO, L. et al. Nuclear Compartmentalization Contributes to Stage-Specific Gene Expression Control in *Trypanosoma cruzi*. **Frontiers in cell and developmental biology**, v. 5, p. 8, 2017.
- PEREIRA, C. A. et al. L-arginine uptake and L-phosphoarginine synthesis in *Trypanosoma cruzi*. **The Journal of eukaryotic microbiology**, v. 46, n. 6, p. 566–570, 1999.
- PEREIRA, K. S. et al. Chagas' Disease as a Foodborne Illness. **Journal of Food Protection**, v. 72, n. 2, p. 441–446, 2009.
- PÉREZ-CAÑADILLAS, J. M.; VARANI, G. Recent advances in RNA-protein recognition. **Current Opinion in Structural Biology**, v. 11, n. 1, p. 53–58, 2001.
- PÉREZ-DÍAZ, L. et al. *Trypanosoma cruzi*: Molecular characterization of an RNA binding protein differentially expressed in the parasite life cycle. **Experimental Parasitology**, v. 117, n. 1, p. 99–105, 2007.
- PÉREZ-DÍAZ, L. et al. The overexpression of the trypanosomatid-exclusive TcRBP19 RNA- binding protein affects cellular infection by *Trypanosoma cruzi*. **Memorias do Instituto Oswaldo Cruz**, v. 107, n. 8, p. 1076–1079, 2012.
- PÉREZ-DÍAZ, L. et al. Evidence for a negative feedback control mediated by the 3' untranslated region assuring the low expression level of the RNA binding protein TcRBP19 in *T. cruzi* epimastigotes. **Biochemical and Biophysical Research Communications**, v. 436, n. 2, p. 295–299, 2013.
- PÉREZ-MOLINA, J. A. et al. EuroTravNet: imported Chagas disease in nine European countries, 2008 to 2009. **Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin**, v. 16, n. 37, 15 set. 2011.
- PÉREZ-MOLINA, J. A.; MOLINA, I. Chagas disease. **The Lancet**, v. 6736, n. 17, p. 1–13, 2017.
- PINTO, A. Y. DAS N. et al. Clinical Follow-Up of Responses to Treatment with Benznidazol in Amazon: A Cohort Study of Acute Chagas Disease. **PLoS ONE**, v. 8, n. 5, p. 1–9, 2013.

- PORCEL, B. M. et al. Gene survey of the pathogenic protozoan *Trypanosoma cruzi*. **Genome Research**, v. 10, n. 8, p. 1103–1107, 2000.
- RAO, V. S. et al. Protein-protein interaction detection: methods and analysis. **International journal of proteomics**, v. 2014, p. 147648, 2014.
- RAPAPORT, F. et al. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. **Genome Biology**, v. 14, n. 9, p. R95, 2013.
- RASSI JR, A. et al. American Trypanosomiasis (Chagas Disease). **Infectious Disease Clinics of North America**, v. 26, n. 2, p. 275–291, 2012.
- RASSI JR, A. et al. Chagas disease. **The Lancet**, v. 375, n. 9723, p. 1388–1402, 2010.
- ROBINSON, M. D.; OSHLACK, A. A scaling normalization method for differential expression analysis of RNA-seq data. **Genome Biology**, v. 11, n. 3, p. R25, 2010.
- ROBINSON, M. D.; SMYTH, G. K. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. **Biostatistics**, v. 9, n. 2, p. 321–332, 11 jul. 2007a.
- ROBINSON, M. D.; SMYTH, G. K. Moderated statistical tests for assessing differences in tag abundance. **Bioinformatics**, v. 23, n. 21, p. 2881–2887, 1 nov. 2007b.
- RODRIGUEZ-GUERINEAU, L. et al. Pediatric Chagas Disease in Europe. **The Pediatric Infectious Disease Journal**, v. 33, n. 5, p. 458–462, maio 2014.
- RUIZ-ECHEVARRÍA, M. J.; PELTZ, S. W. The RNA binding protein Pub1 modulates the stability of transcripts containing upstream open reading frames. **Cell**, v. 101, n. 7, p. 741–51, 23 jun. 2000.
- SÁNCHEZ, L. V.; RAMÍREZ, J. D. Congenital and oral transmission of American trypanosomiasis: An overview of physiopathogenic aspects. **Parasitology**, v. 140, n. 2, p. 147–159, 2013.
- SANTIAGO, C. et al. Transcriptome-wide analysis of the *Trypanosoma cruzi* proliferative cycle identifies the periodically expressed mRNAs and their multiple levels of control Santiago. **PLoS ONE**, v. 12, n. 11, p. 1–27, 2017.

SANTOS, M. R. M. et al. The *Trypanosoma cruzi* Genome Project: Nuclear Karyotype and Gene Mapping of Clone CL Brener. **Memorias do Instituto Oswaldo Cruz**, v. 92, n. 6, p. 821–828, 1997.

SCHURCH, N. J. et al. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use?. **RNA (New York, N.Y.)**, v. 22, n. 6, p. 839–51, 2016.

SHIKANAI-YASUDA, M. A.; CARVALHO, N. B. Oral transmission of chagas disease. **Clinical Infectious Diseases**, v. 54, n. 6, p. 845–852, 2012.

SILBER, A. M. et al. Amino acid metabolic routes in *Trypanosoma cruzi*: possible therapeutic targets against Chagas' disease. **Current drug targets. Infectious disorders**, v. 5, n. 1, p. 53–64, mar. 2005.

SIEGEL, T. N. et al. Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites. **Nucleic Acids Research**, v. 38, n. 15, p. 4946–4957, 2010.

SIEGEL, T. N. et al. Systematic study of sequence motifs for RNA trans splicing in *Trypanosoma brucei*. **Molecular and cellular biology**, v. 25, n. 21, p. 9586–94, 2005.

SILVA-DOS-SANTOS, D. et al. Unraveling Chagas disease transmission through the oral route: Gateways to *Trypanosoma cruzi* infection and target tissues. **PLoS Neglected Tropical Diseases**, v. 11, n. 4, p. 1–26, 2017.

SMIRCICH, P. et al. Ribosome profiling reveals translation control as a key mechanism generating differential gene expression in *Trypanosoma cruzi*. **BMC Genomics**, v. 16, n. 1, p. 443, 2015.

SMITH, M. et al. Improving the prediction of mRNA extremities in the parasitic protozoan *Leishmania*. **BMC Bioinformatics**, v. 9, 2008.

SMYTH, G. K. limma: Linear Models for Microarray Data. In: **Bioinformatics and Computational Biology Solutions Using R and Bioconductor**. New York: Springer-Verlag, 2005. p. 397–420.

SONESON, C.; DELORENZI, M. A comparison of methods for differential expression analysis of RNA-seq data. **BMC Bioinformatics**, v. 14, n. 1, p. 1, 2013.

- SOUTO, R. P.; ZINGALES, B. Sensitive detection and strain classification of *Trypanosoma cruzi* by amplification of a ribosomal RNA sequence. **Molecular and Biochemical Parasitology**, v. 62, n. 1, p. 45–52, 1993.
- SOUZA, R. T. et al. Genome Size, Karyotype Polymorphism and Chromosomal Evolution in *Trypanosoma cruzi*. **PLoS ONE**, v. 6, n. 8, p. e23042, 12 ago. 2011.
- STANAWAY, J. D.; ROTH, G. The Burden of Chagas Disease Estimates and Challenges. **Global Heart**, v. 10, n. 3, p. 139–144, 2015.
- TEIXEIRA, A. R. L. et al. Evolution and pathology in Chagas disease - A Review. **Memorias do Instituto Oswaldo Cruz**, v. 101, n. 5, p. 463–491, 2006.
- TEIXEIRA, S. M. R.; DAROCHA, W. D. Control of gene expression and genetic manipulation in the Trypanosomatidae. **Genetics and molecular research : GMR**, v. 2, n. 1, p. 148–158, 2003.
- TEIXEIRA, S. M. R. et al. *Trypanosoma cruzi*: Suppression of tuzin gene expression by its 5'-UTR and spliced leader addition site. **Experimental Parasitology**, v. 93, n. 3, p. 143–151, 1999.
- THAKUR, V.; VARSHNEY, R. Challenges and Strategies for Next Generation Sequencing (NGS) Data Analysis. **J Comput Sci Syst Biol**, v. 33, n. 2, 2010.
- TIBAYRENC, M.; AYALA, F. J. Isozyme variability in *Trypanosoma cruzi*, the agent of Chagas' disease: genetical, taxonomical, and epidemiological significance. **Evolution**, v. 42, n. 2, p. 277–292, 1988.
- TIBAYRENC, M. Genetic epidemiology of parasitic protozoa and other infectious agents: the need for an integrated approach. **International journal for parasitology**, v. 28, n. 1, p. 85–104, jan. 1998.
- TOSO, A. et al. Transmisión de la enfermedad de Chagas por vía oral Oral transmission of Chagas' disease: Oral transmission of Chagas' disease. **Revista médica de Chile**, v. 139, n. 2, p. 258–66, 2011.
- URBINA, J. A.; AZAVACHE, V. Regulation of energy metabolism in *Trypanosoma* (schizotrypanum) *cruzi* epimastigotes. II. NAD⁺-dependent glutamate dehydrogenase. **Molecular and Biochemical Parasitology**, v. 11, p. 241–255, 1 abr. 1984.

- VANHAMME, L.; PAYS, E. Control of gene expression in trypanosomes. **Microbiological reviews**, v. 59, n. 2, p. 223–240, 1995.
- VAZ, N. P. Can the Cure for Chagas' Disease be Found in Nature? In: **Natural Remedies in the Fight Against Parasites**. [s.l.] InTech, 2017.
- WANG, X. et al. Modular recognition of RNA by a human Pumilio-homology domain. **Cell**, v. 110, n. 4, p. 501–512, 2002.
- WANG, Z. et al. RNA-Seq: a revolutionary tool for transcriptomics. **Nature reviews. Genetics**, v. 10, n. 1, p. 57–63, jan. 2009.
- WEATHERLY, D. B. et al. Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. **BMC Genomics**, v. 10, p. 1–13, 2009.
- WEGNER, D. H. G.; ROHWEDDER, R. W. **The effect of nifurtimox in acute Chagas' infection** *Arzneimittel-Forschung*, 1972.
- WHO. Investing to overcome the global impact of neglected tropical diseases: third WHO report on neglected diseases 2015. **Investing to overcome the global impact of neglected tropical diseases: third WHO report on neglected diseases**, p. 191, 2015a.
- WHO, W. H. O. **Weekly epidemiological record**, 2015b. Disponível em: <<http://www.who.int/wer/2015/wer9006.pdf?ua=1>>. Acesso em: 22 dez. 2017.
- WHO, W. H. O. **Chagas in the Americas for the General Public**. Disponível em: <http://www.paho.org/hq/index.php?option=com_content&view=article&id=13566%3Achagas-in-the-americas-for-the-general-public&catid=6648%3Afact-sheets&Itemid=40721&lang=en>. Acesso em: 19 dez. 2017.
- WORLD HEALTH ORGANIZATION. **Chagas disease (American trypanosomiasis)**. Disponível em: <<http://www.who.int/mediacentre/factsheets/fs340/en/>>. Acesso em: 19 dez. 2017.
- WU, L. et al. MicroRNAs direct rapid deadenylation of mRNA. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 11, p. 4034–4039, 2006.

XAVIER, S. C. DAS C. et al. Distantiae Transmission of *Trypanosoma cruzi*: A New Epidemiological Feature of Acute Chagas Disease in Brazil. **PLoS Neglected Tropical Diseases**, v. 8, n. 5, 2014.

ZAHIRI, J. et al. Computational Prediction of Protein-Protein Interaction Networks: Algorithms and Resources. **Current genomics**, v. 14, n. 6, p. 397–414, set. 2013.

ZINGALES, B. et al. Biological Parameters and Molecular Markers of Clone CL Brener - The Reference Organism of the *Trypanosoma cruzi* Genome Project. **Memorias do Instituto Oswaldo Cruz**, v. 92, n. 6, p. 811–814, 1997.

ZINGALES, B. et al. A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: Second revision meeting recommends TcI to TcVI. **Memorias do Instituto Oswaldo Cruz**, v. 104, n. 7, p. 1051–1054, 2009.

ZINGALES, B.; COLLI, W. *Trypanosoma cruzi*: interaction with host cells. **Current topics in microbiology and immunology**, v. 117, p. 129–52, 1985.

ZWIERZYNSKI, T. A. et al. A. In vitro 3' end processing and poly (A) tailing of RNA in *Trypanosoma cruzi*. **Nucleic Acids Research**, v. 17, n. 12, p. 4647–4660, 1989.

APÊNDICE

Tabela 8. Script utilizado para a realização de análise de expressão diferencial.

```
library(DESeq2)

#Lendo as contagens
count_table <- read.table("count_table.csv", header=T, row.names=1)

#Criando replicatas
replicas <- rep(c("cl", "WT"), c(4,3))

#Criando colData
colData <- data.frame (condition=as.factor (replicas))

#Pré-filtro (Filtrando baixas contagens)
A <- rowSums(dados)
isexpr <- A>50
dados.filt <- dados[isexpr,]

#Análise de expressão diferencial
#Criando objeto dds
ddsData <- DESeqDataSetFromMatrix(dados.filt, colData, formula (~condition))
dds <- DESeq(ddsData)

#Transformação rlog dos dados # Passo necessário para a construção do
HeatMap
rld <- rlog(dds, blind = FALSE) # Transforma os dados de contagens para a escala
log2
dds <- estimateSizeFactors(dds)

#Obtendo os resultados da análise de expressão diferencial
res <- results(dds, alpha = 0.1, pAdjustMethod = "BH",
              lfcThreshold = 0, contrast = c("condition", "cl", "WT"))

#Produzindo as tabelas DEGs
resOrde <- res[order(res$padj),]
res.tab <- data.frame(resOrde)

res.padj <- # Cria tabela com todos os genes estatisticamente
res.tab[res.tab$padj<0.05, ] diferencialmente expressos, independente do
res.padj <- Log2FoldChange
res.padj[!is.na(res.padj$padj),]

de_up <- #Cria tabela com genes que apresentaram
log2FoldChange >= 1
res.padj[res.padj$log2FoldChange >= 1, ]
```

```

de_down <-
res.padj[res.padj$log2FoldChange <= -1, ] # Cria tabela com genes que apresentaram
log2FoldChange <= -1

# Criando MA Plot
plotMA(res, ylim=c(-2,2))
#Barplot DE (logFC > 2 ou logFC < -2)

resumo <- read.csv("genes_de.csv", header = T)
rownames(resumo) <- c("package","Up1", "Up2", "Up3", "Down1", "Down2",
"Down3")
resumo <- as.data.frame(t(resumo))

plot_ly(resumo, x = ~package, y = ~Up2, type = "bar", name = "FC 2 a 4", color =
l("cyan3")) %>%
  add_trace(y = ~Up3, name = "FC > 4", color = l("cyan4")) %>%
  add_trace(y = ~Down2, name = "FC -2 a -4", color =l("coral3")) %>%
  add_trace(y = ~Down3, name = "FC < -4", color =l("coral4")) %>%
  layout(title = "Genes DE",
         xaxis = list (title = "Comparação (Nocaute vs WT)",
                       yaxis = list (title = "Número de genes"),
                       bargroup = "relative")

#Criando uma lista de tabelas
de_list <- list(deseq_up= ids_deseq2_up, edgeR_up = ids_edgeR_up, limma_up =
ids_limma_up, deseq_down = ids_deseq2_down, edgeR_down = ids_edgeR_down,
limma_down = ids_limma_down, cex = 2)

#Criando VennDiagram
vp <- venn.diagram(de_list[1:3], fill = 1:3, alpha = 0.3, filename = NULL, main =
"Venn Diagram Up", main.cex = 1.7)

vp <- venn.diagram(de_list[4:6], fill = 4:6, alpha = 0.3, filename = NULL, main =
"Venn Diagram Down",main.cex = 1.7)

```